UNIVERSITÉ DE PROVENCE
U.F.R. M.I.M.
ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE E.D. 184

THÈSE

présentée pour obtenir le grade de

Docteur de l'Université de Provence

Spécialité : Informatique

par

Laurent BIENVENU

sous la direction de Bruno DURAND et Alexander SHEN

Titre :

CARACTÉRISATIONS DE L'ALÉATOIRE PAR LES JEUX : IMPREDICTIBILITÉ ET STOCHASTICITÉ

soutenue publiquement le 4 avril 2008

JURY

M. Eugène ASARIN	Université de Paris 7	*Rapporteur*
M. Bruno DURAND	Université de Provence	*Directeur de thèse*
M. Peter GACS	Boston University	*Rapporteur*
M. Serge GRIGORIEFF	Université de Paris 7	*Examinateur*
M. Alexander SHEN	CNRS & LIF, Marseille	*Directeur de thèse*
M. Vladimir VOVK	Royal Holloway College, London	*Examinateur*

à Melanie, Rodrigo, Cusi, Karol, Carmen, Leydi, et les autres,

pour tout ce qu'ils m'ont donné

Contents

Remerciements

Je tiens avant toute chose à remercier mes directeurs de thèse, Bruno Durand et Alexander (Sasha) Shen.

Bruno le premier m'a fait confiance en acceptant de me prendre comme stagiaire de Master avant même de m'avoir rencontré. De lui j'ai appris énormément de choses, et même si nos domaines d'intérêt ont un peu divergé au cours de ma thèse, il a toujours veillé à ce que cette dernière se passe dans les meilleures conditions; de cela je lui suis infiniment reconnaissant.

Sasha, par son enthousiasme, sa gentillesse, sa patience et son intelligence, a fait de mes années de thèse une période captivante, en partageant sans cesse ses nombreuses idées et questions. Je souhaite de tout coeur que notre collaboration et surtout notre amitié se prolonge bien au-delà de ces trois ans passés ensemble à Marseille.

Je tiens ensuite à remercier Serge Grigorieff. Le remercier d'abord pour ses cours passionnés et passionnants (pour lesquels j'ai traversé la moitié de Paris en courant un jour de grève de la RATP !) qui m'ont fait découvrir les splendeurs de la complexité de Kolmogorov. Le remercier également pour son oreille attentive et son soutien indéfectible lors des périodes de doute que j'ai pu rencontrer au cours de ma thèse. Le retrouver trois ans après mon Master dans mon jury de soutenance est pour moi un honneur et une grande joie.

Depuis que j'ai eu la chance de le rencontrer à la conférence STACS 2006, Wolfgang Merkle a été un collaborateur exceptionnel, tant par ses qualités humaines que scientifiques. De nombreux résultats de cette thèse sont issus de cette collaboration qui j'espère n'en est qu'à ses débuts.

Je remercie vivement les autres membres de mon jury: Eugène Asarin et Peter Gacs, qui ont généreusement accepté d'être les rapporteurs du manuscrit, ainsi que Vladimir Vovk. Merci à eux d'avoir pu se libérer pour assister à la soutenance malgré un emploi du temps que je sais chargé.

En espérant n'oublier personne, je remercie aussi: les excellents professeurs que j'ai pu avoir tout au long de mon cursus et qui m'ont donné le goût de la science et de la recherche (un grand merci notamment à Jacques Mazoyer et Marianne Delorme); mes co-auteurs: David Doty, Andrei A. Muchnik, Mathieu Sablik, Frank Stephan, Nicolay Vereshchagin; tous les membres de l'équipe Escape pour les discussions

animées du coin café; Martine, Sylvie et Nathalie, qui m'ont tant de fois sorti de situations délicates, toujours avec patience et bonne humeur; Audrey Romano et tous les participants de Vitascience pour le projet que nous avons mené, adoucissant grandement cette pénible expérience que constitue le CIES; Rod Downey et Denis Hirschfeldt pour avoir mis leur excellent livre en libre accès, ce dont j'ai grandement profité; merci également à l'Association for Symbolic Logic pour son soutien financier, qui m'a permis de me rendre à la conférence "Logic Complexity and Randomness" à Buenos Aires en janvier 2007.

Enfin, je remercie par dessus tout mes amis, ma famille et ma belle-famille, pour ces innombrables bons moments que nous avons partagés et qui font que la vie vaut la peine d'être vécue. J'ai aujourd'hui une pensée toute particulière pour ma femme Meghyn et mon frère Thomas, qui ont eux aussi débuté une thèse peu après moi. Je leur souhaite tout le bonheur possible dans cette belle aventure.

Résumé de la thèse

Théorie effective de l'aléatoire: motivations

Cette thèse est une contribution à la théorie effective de l'aléatoire, également connue sous le nom de théorie algorithmique de l'aléatoire. Le but de cette théorie est de donner un sens à l'idée intuitive d'objet aléatoire.

Commençons par illustrer les motivations de cette théorie par un exemple tiré de la "vie courante". Imaginons une entreprise commercialisant des suites de bits aléatoires (par exemple des suites de 10^{10} bits gravées sur DVD). Cette entreprise promeut ses DVD comme "contenant des suites de bits aléatoires, générés par un processus véritablement aléatoire" (comme par exemple la désintégration d'éléments radioactifs). Nous sommes intéressés par de telles suites, que nous souhaitons utiliser comme source de hasard pour un algorithme probabiliste (par exemple pour un test de primalité). Nous commandons donc à cette entreprise une suite de 10^{10}, et nous reçevons un DVD contenant la suite

0101010101010101010101.... (01 répété 5 milliards de fois)

Le moins que l'on puisse dire est que nous ne sommes pas satisfaits par cette suite. Nous écrivons donc une lettre de réclamation, dans laquelle nous nous plaignons que la suite que nous avons reçue n'est absolument pas aléatoire, et demandons à être remboursés. Nous reçevons alors la réponse suivante:

"Madame, Monsieur. Nous avons bien reçu votre lettre au sujet de la suite de bits aléatoires que vous nous avez achetée. Nous sommes navrés qu'elle ne vous ait pas donné satisfaction. Cependant, nous ne comprenons pas votre affirmation que cette suite n'est pas aléatoire. En effet, cette suite a la même probabilité d'occurence que n'importe quelle autre suite de longueur 10^{10}. En conséquence, toute propriété de cette suite qui vous fait douter de sa nature aléatoire n'est que pure coïncidence."

On en conviendra, cette réponse n'est pas satisfaisante. L'argument employé comme quoi cette suite a la même probabilité d'occurence que toute autre suite est tout-à-fait correct mais malgré cela, notre intuition persiste à dire que cette suite n'est pas aléatoire. Comment pouvons-nous l'expliquer rigoureusement ? Et cette question en amène une autre: quel type de garantie l'entreprise peut-elle offrir sur les suites de bits qu'elle vend afin de gagner la confiance de ses clients ?

Ces questions sont de nature philosophique, et n'admettent donc pas de réponse unique. La théorie effective de l'aléatoire ne constitue qu'une tentative parmi d'autres de répondre à ces questions. En fait, au sein même de cette théorie, diverses réponses sont envisagées. Cependant, elles partagent toutes le même leitmotiv:

Un objet individuel est aléatoire s'il n'existe aucune façon calculable de prouver qu'il n'est pas aléatoire.

Ceci veut dire que toutes les définitions effectives d'objet aléatoire sont des définitions négatives: on donne d'abord une définition effective (calculable) d'objet non-aléatoire, puis on définit un objet aléatoire comme étant un objet qui n'est pas non-aléatoire.

Il reste à préciser ce que l'on entend par "calculable". La thèse de Church-Turing, presque universellement acceptée aujourd'hui, affirme que "calculable" doit être compris comme "calculable par machine de Turing" (ou tout autre modèle de calcul équivalent). Ainsi, toute la théorie effective de l'aléatoire se base sur la théorie de la calculabilité.

Structure de cette thèse

Dans cette thèse on étudie la théorie effective de l'aléatoire pour les suites binaires, finies ou infinies. Le premier chapitre présente diverses défintions de suite binaire aléatoire. On peut classer ces dernières en deux grandes catégories: les notions de typicalité et les notions d'imprédictibilité.

Les notions de typicalité – basées sur la théorie de la mesure – formalisent l'idée intuitive qu'une suite binaire infinie satisfait toutes les propriétés de probabilité 1 pouvant être testées de façon "effective". Les deux principales notions de cette catégorie sont l'aléatoire au sens de Martin-Löf et l'aléatoire au sens de Schnorr. On présente également le concept de dimension de Hausdorff effective. Bien qu'un peu trop faible pour définir le concept de suite aléatoire, la dimension de Hausdorff effective permet d'associer à toute suite binaire infinie un réel entre 0 et 1 mesurant son degré d'aléatoire (0 signifiant "pas du tout aléatoire", et 1 signifiant "assez aléatoire").

Comme on peut s'en douter, les notions d'imprédictibilité expriment le fait qu'une suite binaire infinie est aléatoire s'il n'existe aucune façon "effective" d'en prédire les bits. Historiquement, le premier modèle formalisant cette intuition fut le modèle de règle de sélection, introduit par von Mises. Une règle de sélection est un procédé qui, étant donnée une suite binaire infinie, en sélectionne une sous-suite (qui peut être finie ou infinie). Un point important est qu'une règle de sélection ne doit pas pouvoir connaître la valeur d'un bit avant de prendre la décision de le sélectionner ou non. Von Mises souscrivait à l'idée que la probabilité d'un évènement est la fréquence asymptotique d'occurence de cet évènement lorsque l'on repète une expérience une infinité de fois. Il définit donc une suite binaire in-

finie aléatoire comme étant une suites dont toutes les sous-suites infinies obtenues par une règle de sélection effective (la notion de calculabilité n'étant pas connue du temps des travaux de von Mises) satisfont la Loi des Grands Nombres (i.e. comprennent asymptotiquement autant de zéros que de uns). Ville montra par la suite que cette définition est en réalité trop faible, certaines suites aléatoires au sens de von Mises ayant des propriétés clairement non-aléatoires (comme par exemple des suites dont tous les préfixes contiennent plus de zéros que de uns). Plutôt qu'aléatoire, nous utilisons un autre terme, celui de suites "stochastiques", pour désigner de telles suites. Nous présenterons deux classes de suites stochastiques: les suites Church-stochastiques (pour lesquelles on considère des règles de sélection monotones, sélectionnant les bits de gauche à droite), et les suites Kolmogorov-Loveland-stochastiques (pour lesquelles on considère des règles de sélection qui peuvent sélectionner les bits dans un ordre quelconque). Afin d'affiner le modèle de von Mises et le concept d'imprédictibilité, Schnorr (dans la continuité des idées de Ville) proposa un modèle plus général, dans lequel un joueur tente de prédire les bits d'une suite binaire infinie en pariant de l'argent sur leurs valeurs, doublant sa mise quand il a raison, la perdant quand il a tort. Le joueur gagne si son capital tend vers l'infini au cours de la partie. La suite binaire infinie est dite imprédictible si aucune stratégie calculable ne permet au joueur de gagner contre cette suite. Comme pour la stochasticité, deux classes de suites imprédictibles sont présentées: les suites récursivement aléatoires (en anglais "computably random") pour lesquelles le joueur parie sur les bits de gauche à droite, et les suites Kolmogorov-Loveland aléatoires, pour lesquelles le joueur peut parier sur les bits dans un ordre quelconque.

Bien que toutes les notions d'aléatoires mentionnées ci-dessus soient distinctes (ceci est déjà connu et sera expliqué au fur et à mesure de cette thèse), elles peuvent toutes être caractérisées en termes de jeux, sur le modèle des suites récursivement aléatoires. Ceci montre que typicalité et imprédictibilité sont intrinsèquement liées, et permet de donner une classification précise des différentes notions d'aléatoire. En particulier, on montrera qu'une suite infinie est non-stochastique s'il existe une stratégie calculable permettant de gagner de l'argent exponentiellement vite (par rapport au nombre de coups où le joueur mise un montant d'argent strictement positif). Ce résultat avait été obtenu auparavant par Schnorr, mais nous donnerons un résultat quantitatif, reliant précisément la vitesse des stratégies gagnantes au biais des sous-suites sélectionnées.

Un autre concept relié à la Loi des Grands Nombres est celui de normalité. Un nombre réel est dit normal dans une certaine base si la suites de ses décimales dans cette base a la propriété que tous les motifs (i.e. suites finies de chiffres) apparaissent avec la même fréquence (asymptotiquement). Un nombre est dit absolument normal s'il est normal dans toute base entière. Becher et Figueira ont prouvé qu'il existait un nombre absolument normal calculable. Nous donnons une preuve alternative de ce résultat comme illustration des concepts et résultats présentés dans ce premier chapitre.

Le chapitre se termine par une brève discussion portant sur les liens entre aléatoire et théorie des catégories de Baire. Bien qu'elles aient pour mesure 1, toutes les classes "raisonnables" de suites aléatoires sont maigres au sens de Baire.

Dans le second chapitre, nous revenons à la question de l'aléatoire pour les suites binaire finies. Nous commençons par introduire la notion fondamentale de complexité de Kolmogorov, introduite indépendamment et presque simultanément par Solomonoff, Kolmogorov et Chaitin. La complexité de Kolmogorov d'une suite binaire finie est définie comme étant la taille du plus court programme qui la génère. Il est facile de voir que la complexité de Kolmogorov d'une suite est comprise entre 0 et (en gros) sa longueur. Intuitivement, une suite finie aléatoire doit être difficile à décrire, tandis qu'une suite non-aléatoire présente une certaine régularité (sinon, sur quel critère pourrait-on la juger non-aléatoire ?) et donc admet une description plus courte qu'elle-même. Cette intuition montre que la complexité de Kolmogorov est une bonne mesure du caractère aléatoire d'une suite binaire finie.

On présente deux variantes de la complexité de Kolmogorov: la complexité de Kolmogorov dite "pleine" et la complexité de Kolmogorov "préfixe". Nous en donnons les principales propriétés, notamment le fait que la majorité des suites finies ont une complexité quasi-maximale (i.e. proche de leur longueur), ce qui est bien sûr le moins que l'on puisse attendre d'une mesure d'aléatoire. Nous expliquons également pourquoi aucune de ces deux complexités n'est calculable.

La seconde partie du chapitre étudie les liens entre la complexité de Kolmogorov et les notions d'aléatoire introduites au chapitre précédent. Nous commençons par les suites aléatoires de Martin-Löf, pour lesquelles la situation est bien comprise. En effet, le théorème fondamental de Levin et Schnorr fournit un critère précis (c'est-à-dire une condition nécessaire et suffisante) pour cette notion: une suite binaire infinie est aléatoire au sens de Martin-Löf si tous ses segments initiaux ont une complexité préfixe plus grande que leur longueur, à une constante additive près. Nous présentons également le théorème de Miller et Yu, qui donne une caractérisation similaire en termes de complexité de Kolmogorov pleine. Le rapport à la complexité de Kolmogorov est moins limpide pour les autres notions d'aléatoire. En particulier, les notions de suite récursivement aléatoire, de suite Schnorr aléatoire, et de suite Church stochastique sont en quelque sorte "orthogonales" à la complexité de Kolmogorov. En effet, il existe des suites qui sont Schnorr aléatoires, récursivement aléatoires et Church stochastiques et dont les segments initiaux ont une complexité de Kolmogorov très faible; et inversement, ils existe des suites dont les segments initiaux ont une complexité élevée et qui ne sont ni Schnorr aléatoires, ni récursivement aléatoires, ni Church stochastiques. Nous donnons cependant une condition suffisante pour les suites Schnorr aléatoires en termes de complexité préfixe, et nous montrons que cette condition n'est suffisante ni pour les suites récursivement aléatoires ni pour les suites Church stochastiques. Le cas des suites Kolmogorov-Loveland stochastiques est également intéressant. Un résultat récent de Merkle, Miller, Nies, Reimann et Stephan montre qu'une suite Kolmogorov-Loveland stochastique doit avoir une forte complexité de Kolmogorov. Plus précisément, le rapport "complexité sur longueur" de ses segments initiaux doit tendre vers 1. Nous étudions la stochasticité de Kolmogorov-Loveland du point de vue inverse: si le rapport "complexité sur longueur" d'une suite ne tend pas vers 1, il est possible d'en sélectionner (de façon non-monotone) une sous-suite biaisée; quel lien y-a-t'il entre le biais des sous-suites sélectionnées et la limite

inférieure du rapport "complexité sur longueur" ? Cette question fut étudiée par Asarin, Durand et Vereshchagin dans le cas des suites binaires finies, mais leurs résultats ne sont pas directement applicables aux suites infinies. En combinant les techniques de Merkle et al. avec la caractérisation de la stochasticité par les jeux (prouvée au Chapitre 1), nous donnons une borne inférieure précise pour le biais maximal des sous-suites sélectionnées (borne dont nous prouvons l'optimalité au Chapitre 3). Nous discutons également les interprétations de ces résultats en termes de dimension de Hausdorff effective.

La complexité de Kolmogorov n'étant pas calculable, on est en droit de se demander en quoi il est légitime d'en faire la notion centrale de la théorie "effective" de l'aléatoire. De plus, ceci donne peu d'espoir de lui trouver une quelconque application pratique. Une façon de contourner le problème est de considérer des approximations calculables de la complexité de Kolmogorov. La complexité d'une suite finie est définie comme étant la longueur de sa plus courte "description", qui peut également être vue comme étant sa plus courte forme compressée. Il n'y a pas en général de méthode effective pour trouver cette meilleure compression, mais il est possible d'en trouver une raisonnable en compressant la suite par un algorithme classique de compression (par exemple Lempel-Ziv ou Burrows-Wheeler). La longueur de la suite compressée ainsi obtenue est alors une approximation de la complexité de Kolmogorov de la suite initiale. Plus précisément, elle en est une borne supérieure. Cette approche fut utilisée par Cilibrasi et Vitanyi pour des applications pratiques (classification de données notamment). La dernière partie de ce chapitre reprend cette approche, d'un point de vue assez théorique toutefois. Nous réexaminons les liens entre l'aléatoire et la complexité lorsque l'on remplace la complexité de Kolmogorov par ses bornes supérieures calculables. De façon surprenante, toutes les caractérisations des suites aléatoires de Martin-Löf restent vraies dans ce contexte. En particulier, le théorème de Levin-Schnorr reste vrai si l'on remplace la complexité de Kolmogorov par ine borne supérieure calculable bien choisie. Ceci permet même de donner une preuve simple du théorème de Miller et Yu. Mieux encore, certaines classes de suites d'aléatoire, comme par exemple les suites Schnorr aléatoires, qui semblaient présenter peu de liens avec la complexité de Kolmogorov (comme expliqué plus haut), se caractérisent de façon fort naturelle en termes de bornes supérieures calculables.

Les deux premiers chapitres traitent de la théorie effective de l'aléatoire pour la mesure uniforme (de Lebesgue), pour lesquels les bits sont choisis indépendamment les uns des autres, et où chaque bit a exactement une chance sur deux d'être 0. Le troisième et dernier chapitre généralise l'étude de la théorie effective de l'aléatoire à toutes les mesures de probabilité calculables. Hormis la stochasticité, toutes les notions d'aléatoire discutées plus haut peuvent être adaptées à des mesures calculables quelconques. Informellement, la question principale étudiée dans ce chapitre est la suivante: "Jusqu'à quel point peut-on changer la mesure de probabilité sans affecter les notions d'aléatoire ?" Par exemple, quelles sont les mesures de probabilité dont les suites Martin-Löf aléatoires sont les mêmes que les suites Martin-Löf aléatoires pour la mesure uniforme ? Cette question est liée à la notion d'équivalence entre mesures. En théorie classique des probabilités, deux mesures sont équivalentes si

elles ont les mêmes ensembles de mesure 0 (intuitivement, tout évènement improbable pour l'une est improbable pour l'autre). On peut donner une version effective de cette notion, en disant que deux mesures sont "effectivement équivalentes" si elles admettent les mêmes suites aléatoires. Bien sûr, avec ce schéma, différentes notions d'aléatoires induisent (a priori) différentes relations d'équivalence.

La première partie du chapitre est consacrée aux relations d'équivalence pour une classe de mesures particulière: les mesures de Bernoulli généralisées. Elles correspondent à la situation où les bits d'une suite binaire infinie sont choisis indépendamment les uns des autres, mais la distribution de probabilité entre 0 et 1 dépend de la position du bit dans la suite. En théorie classique des probabilités, il existe un critère bien connu pour déterminer si deux mesures de Bernoulli gńéralisées sont équivalentes: le théorème de Kakutani. Nous étendons les travaux de Muchnik, Semenov, Uspenski, et Vovk, pour montrer que le critère de Kakutani est également valide pour toutes les relations d'équivalence effectives que nous considérons. Pour prouver ce résultat, nous utilisons un argument basé sur la théorie des jeux: si deux mesures de Bernoulli généralisées satisfont le critère de Kakutani, il est possible de transformer de façon effective une stratégie gagnante par rapport à la première mesure en une stratégie gagnante pour la seconde mesure. Des applications de cette classe de mesures sont également présentées: nous montrons comment elles peuvent être utilisées pour prouver que la classe des suites Kolmogorov-Loveland stochastiques est strictement plus grande que la classe des suites Martin-Löf aléatoires (un résultat de van Lambalgen et Shen). Nous les utilisons aussi pour prouver l'optimalité des bornes (présentées au Chapitre 2) reliant le défaut d'léatoire d'une suite au biais de ses sous-suites sélectionnées.

Bien que toutes les relations d'équivalences effectives que nous étudions coïncident sur la classe des mesures de Bernoulli généralisées calculables, ce n'est plus vrai dans le cas général (avec des mesures calculables quelconques). Dans le reste du chapitre, nous donnons une classification complète des relations d'équivalence effectives. Pour cela, nous utilisons diverses notions de calculabilité "pure", comme les degrés de Turing hauts et l'hyperimmunité. Un résultat important de cette classification est que deux mesures calculables ayant les mêmes suites Martin-Löf aléatoires sont nécessairement équivalentes, mais la réciproque n'est pas vraie. Autre résultat remarquable: alors que deux mesures sur l'espace de Cantor sont classiquement équivalentes si elles ont les mêmes fermés de mesure nulle, ceci n'est plus vrai quand on passe au point de vue constructif. Précisément, il n'est pas vrai que deux mesures calculables ayant les mêmes fermés effectifs de mesure nulle sont nécessairement équivalentes. Enfin, on remarquera le fait que les implications entre les différentes relations d'équivalence effectives ne sont aucunement liées aux notions d'aléatoire sous-jacentes.

Principales contributions

Pour résumer ce qui précède, les principaux résultats originaux de cette thèse sont les suivants:

- Nous donnons une analyse quantitative du résultat de Schnorr affirmant

qu'une stratégie gagnant exponentiellement vite peut être transformée en règle de sélection qui sélectionne une sous-suite infinie biaisée (Théorème 1.4.16). Nous utilisons ce résultat pour donner des bornes précises reliant le défaut d'aléatoire des suites binaires infinies au biais maximal des sous-suites sélectionnées (Théorème 2.2.31). Ce travail a été publié dans [8].

- Nous donnons une preuve originale du fait que les suites aléatoires de Schnorr ne sont pas toutes Church-stochastiques: nous donnons pour cela une condition sur la complexité de Kolmogorov des segments initiaux qui est suffisante pour les suites Schnorr aléatoires, et insuffisante pour les suites Church stochastiques (Proposition 2.2.20 et Théorème 2.2.21). Le cas particulier des suites approchables par le bas est également traité (Proposition 2.2.24).

- Nous présentons une nouvelle approche de la théorie effective de l'aléatoire basée sur les bornes supérieures calculables de la complexité de Kolmogorov (Section 2.3), et on montre qu'elles fournissent un cadre unifié dans lequel on peut exprimer de nombreuses notions d'aléatoire. Ce travail a fait l'objet de la publication [10].

- Nous prouvons une version constructive du théorème de Kakutani (Théorème3.2.5) en utilisant des arguments de théorie des jeux. Ce théorème généralise des résultats antérieurs, et peut être utilisé pour prouver le théorème de Kakutani classique. Ce résultat figure dans l'article [7].

- Nous donnons une classification complète des relations d'équivalence effectives induites par les différentes notions d'aléatoire sur l'espace des mesures de probabilité calculables (Sous-section 3.3.2). Ce travail a été publié dans [9].

Introduction

Effective randomness

This thesis is a contribution to the field of effective randomness, also known as algorithmic randomness. The purpose of this theory is to give a mathematical meaning to the idea of a "random object".

To understand the motivation of this theory, let us start with a "real-life" example. Imagine a company whose business is to sell sequences of random bits (say, sequences of 10^{10} bits burnt on a DVD), which they advertise as "Genuine random bits, generated by some truly random process" (radioactive decay for example). We are interested in buying such a sequence, which we want to use as a source of randomness for a probabilistic algorithm (e.g. primality testing). We order a sequence from this company, and we receive a DVD containing the sequence

0101010101010101010101.... (repeated five billion times)

Needless to say that we are particularly unsatisfied by this sequence, so we write the company a letter in which we complain that the sequence they sent is not random at all, asking for our money back. Their answer is the following:

"Dear sir or madam. We received your letter about the sequence of random bits you have purchased. We are sorry that it did not give you full satisfaction. However, we do not understand your claim that this sequence is "not random". Indeed, this sequence has the same probability of occurrence as any other sequence of length 10^{10}, so whatever property this sequence has that makes you question its random nature is just a coincidence."

Somehow, we are still not satisfied. Their argument that this sequence has the same probability of occurence as any other one is correct, but we still think that our sequence is not random. How can we explain this formally? A related question is: if a company wants to sell sequences of random bits, what kind of guarantee can they offer to earn trust from their potential customers?

These questions are somewhat philosophical, hence there is no unique way to answer them. Effective randomness is only *one* possible attempt to get a decent answer. Actually, even within effective randomness, there are different answers to

these questions. However, they all have the same leitmotiv:

An individual object is random if there is no computable way to prove its non-randomness.

This means that all the effective definitions of randomness are negative definitions: one first gives a definition of (computable) non-randomness for individual objects, and then defines a random object to be an object that is not non-random.

It remains to explain what we mean by "computable". The Church-Turing thesis, which is almost universally accepted nowadays, tells us that "computable" should be understood as "computable by a Turing machine" (or any equivalent model of computation). Thus, the whole field of effective randomness relies on classical computability theory.

Structure of this thesis

This thesis is concerned with the algorithmic randomness of binary sequences, finite or infinite. The first chapter presents several possible definitions of "random binary sequences". They can be classified into two categories: notions of typicalness and notions of unpredictability.

Typicalness notions of randomness rely on the intuition – based on measure theory – that a random infinite sequence should satisfy all the properties of probability 1 that can be "effectively tested". The two main definitions in this category are Martin-Löf randomness and Schnorr randomness. We also present the concept of effective Hausdorff dimension. Although it is a little too weak to fully define what a random sequence is, effective Hausdorff dimension allows us to assign to every infinite sequence a real number between 0 and 1 measuring the degree of randomness of the sequence (0 meaning "not random at all" and 1 meaning "quite random").

As one can guess, unpredictability notions express the fact that an infinite binary sequence is random if there is no "effective way" to predict its bits. The first model to support this intuition is the model of selection rules introduced by von Mises. A selection rule is a process which, given an infinite binary sequence, selects from it a subsequence (which can be finite or infinite). An important restriction is that the selection rule should not be able to look at a bit before making the decision to select it or not. Von Mises, who was motivated by the frequency approach to randomness, defined an infinite sequence to be "random" if all the infinite subsequences selected from it by an effective selection rule satisfy the Law of Large Numbers. As shown by Ville, this model is too weak; some sequences have this property and yet can hardly be called random. Hence, instead of "random", we will use another term, "stochastic", for sequences having this property. Two variations of stochasticity will be presented: Church stochasticity, where the selection rules select bits in order (from left to right), and Kolomogorov-Loveland stochasticity, where the selection rules can select the bits in any order. To refine the concept

of unpredictability, Schnorr (elaborating on the ideas of Ville) proposed a more general model. He considered an infinite gambling game, where a player bets money on the bits of an infinite sequence, losing his stake when he is wrong, doubling its stake when he is right. The player wins if his capital grows unboundedly during the game. A sequence is unpredictable if no computable betting strategy allows the player to win. Similarly to stochasticity, two notions of unpredictability are defined: computable randomness, for which we consider the betting strategies which bets on all the bits in order, and Kolmogorov-Loveland randomness, for which we consider betting strategies that can be on the bits in any order.

Although all the notions of randomness we just mentioned are different (which will be explained throughout this thesis), they can all be characterized in terms of betting strategies. This proves that typicalness and unpredictability are closely related, and yields a classification of randomness notions. In particular, we show that a random sequence is not stochastic if and only if there exists a computable betting strategy which, betting on this sequence, has a capital which increases exponentially in the number of non-zero bets made. This fact was proven earlier by Schnorr, but we provide a careful analysis (based on a compactness argument) of the relation between the maximal rate of success of betting strategies and the maximal bias of the infinite selected subsequences.

Another concept related to the Law of Large Numbers is normality. A real number is said to be normal in a certain base if its expansion in this base forms a sequence of digits in which all patterns of a given length appear with the same frequency. A number is absolutely normal if it is normal in all bases. Becher and Figueira proved that there exists a computable normal number. We provide an alternative proof of this fact as an illustration of the concepts presented earlier.

The first chapter ends with a brief discussion of effective randomness from the point of view of Baire category. Although they have measure 1, all reasonable classes of random sequences turn out to be meager.

In the second chapter, we return to effective randomness for finite binary sequence. We start by introducing the fundamental notion of Kolmogorov complexity, defined independently and almost simultaneously by Solomonoff, Kolmogorov and Chaitin. The Kolmogorov complexity of a finite binary sequence is defined as the length of the shortest program which outputs this sequence. It is easy to see that the Kolmogorov complexity of a finite sequence is roughly between 0 and its length. Intuitively, a random finite sequence should be hard to describe while a non-random one should contain some kind of pattern or regularity (otherwise, what could possibly make us think that it is not random?) hence should have a description shorter than itself. This makes Kolmogorov a powerful tool to measure the degree of randomness of a finite binary sequence.

We present two versions of Kolmogorov complexity: the plain version and the prefix version. We discuss their basic properties, among which the fact that most strings have maximal complexity (i.e. close to their length), which is the least we can ask from a measure of randomness. We also explain why neither version of Kolmogorov complexity is computable.

The second part of the chapter studies the interaction between Kolmogorov

complexity and the notions of randomness for infinite sequences. We start with Martin-Löf randomness, for which the situation is well understood. Indeed, the celebrated Levin-Schnorr theorem provides a precise criterion (i.e. a necessary and sufficient condition) for this notion: an infinite binary sequence is Martin-Löf random if and only if all its initial segments have prefix complexity greater than their length, up to an additive constant. We also present the recent Miller-Yu theorem, which provides a similar characterization in terms of plain complexity (stonger versions of Martin-Löf randomness are also discussed). The situation regarding Kolmogorov complexity is not so simple for other notions of randomness. In particular, computable randomness, Schnorr randomness and Church stochasticity are in some sense orthogonal to Kolmogorov complexity. Indeed, there exist Schnorr random, computably random, and Church stochastic sequences of very low Kolmogorov complexity, and there exist sequences of very high complexity which are not computably random, not Schnorr random and not Church stochastic. We can nonetheless give a sufficient condition for Schnorr randomness in terms of prefix complexity, and we show that this condition is not sufficient for computable randomness nor for Church stochasticity, obtaining a new proof of the separation of these concepts (originally obtained by Wang). The case of Kolmogorov-Loveland stochasticity is also interesting. A recent result of Merkle, Miller, Nies, Reimann and Stephan states that a Kolmogorov-Loveland sequence must have high Kolmogorov complexity. More precisely, the ratio "complexity over length" of its initial segments must tend to 1. We investigate the reverse direction of this fact: if the ratio does not tend to 1, i.e. its limit inferior is smaller than 1, then it is possible to select (non-monotonically) an infinite biased subsequence. How does the bias of the selected subsequences relate to the inferior limit of the ratio "complexity over length"? The same question was previously studied by Asarin, Durand and Vereshchagin for finite strings, but their work could not be directly applied to infinite ones. Combining the techinques of Merkle et al. with the characterization of stochasticity via betting strategies proven in Chapter 1, we provide a precise lower bound for this maximal bias (which we prove to be optimal in Chapter 3). An interpretation of this result in terms of effective Hausdorff dimension is also given.

As Kolmogorov complexity is not computable, one can question its legitimity as the central notion of "effective randomness". Also, this non-computability a priori rules out any hope of practical applications. A way to avoid this problem is to consider computable approximations of Kolmogorov complexity. The complexity of a finite string is the length of its shortest "description", which can also be seen as the "shortest compressed form" of the string. In general, there is no way to find this best compression, but we can try to get a reasonable one by running our favorite lossless compression algorithm (e.g. Lempel-Ziv, Burrows-Wheeler...) on the sequence. The length of the compressed string is then an approximation of the Kolmogorov complexity of the original string. More precisely, it is an upper bound, because a better way to compress it might very well exist. This approach was for example followed by Cilibrasi and Vitanyi for practical purposes. The last part of the chapter studies this approach but still from a rather theoretical viewpoint. We reexamine the interaction between randomness and Kolmogorov complexity when Kolmogorov complexity is replaced by its computable upper bounds. Perhaps sur-

prisingly, all the characterizations of Martin-Löf randomness remain true in this setting. In particular, the Levin-Schnorr criterion for Martin-Löf randomness remains true if one replaces the prefix Kolmogorov complexity by a well-chosen upper bound. This even allows us to give a simple proof of the Miller-Yu criterion. Even better, some notions of randomness, like Schnorr randomness, that seem unrelated to Kolmogorov complexity (as explained above) admit very natural characterizations in terms of computable upper bounds.

The first two chapters focus on randomness with respect to the uniform measure for which the bits are chosen independently and both values 0 and 1 have probability 1/2. The last chapter extends the study of effective randomness to all computable probability measures. Stochasticity aside, all notions of randomness can be adapted to any such measure. Informally, the main question studied in the chapter is "how much can we modify the measure without changing the notions of randomness?" For example, what are the measures whose Martin-Löf random sequences are the same as the Martin-Löf sequences for the uniform measure? This question is related to that of equivalence of measures. In classical probability theory, two measures are said to be equivalent if they have the same sets of measure 0 (informally, every unlikely event for one measure is unlikely for the other). We can define effective versions of this equivalence relation, by saying that two measures are "effectively equivalent" if they have the same random elements. Of course, different notions of randomness induce different equivalence relations.

The first part of the chapter studies equivalence relations for a particular class of measures: generalized Bernoulli measures. They correspond to the situation where the bits are chosen independently but where the probability distribution of a bit depends on its position in the sequence. In classical probability theory, there exists a well-known criterion for the equivalence of two generalized Bernoulli measures: Kakutani's theorem. Extending the work of Muchnik, Semenov and Uspenski, and of Vovk, we prove that Kakutani's criterion holds for all the effective equivalence relations we consider. To prove this, we use a game-theoretic argument: if two measures satisfy Kakutani's criterion, we show that every winning betting strategy for a measure can be transformed into a winning betting strategy for the second measure. Applications of generalized Bernoulli measures to stochasticity are presented. We show how they can be used to prove that Kolmogorov-Loveland stochasticity is weaker than Martin-Löf randomness (a result of van Lambalgen and Shen). We also use them to prove the optimality of the bounds (given in Chapter 2) relating randomness deficiency to the biases of selected subsequences.

While all the effective equivalence relations we consider coincide when restricted to generalized Bernoulli measures, this is no longer the case for arbitrary computable measures. The rest of the chapter provides a complete classification of the equivalence relations, whose proof involves various notions of computability theory, like hyperimmunity or highness. An interesting result in this classification is that two computable measures which have the same class of Martin-Löf random elements are equivalent in the classical sense, but the converse is not true. Also, two measures on the Cantor space which have the same closed nullsets are equivalent in the classical sense, but this fact cannot be effectivized: we prove that two

computable measures which have the same effectively closed sets need not be equivalent. Another very interesting fact is that the implications between the different equivalence relations are completely unrelated to the implications of the underlying notions of randomness.

Main contributions

To sum up, the main original contributions of this thesis are the following.

- We give a quantitative analysis of how an exponentially winning betting strategy can be transformed into a selection rule which selects an infinite biased subsequence (Theorem 1.4.16). We use this result to give precise bounds expressing how biased the (non-monotonically selected) subsequences can be given the degree of non-randomness of the sequence (Theorem 2.2.31). This work was published in [8]

- We provide an original proof for the separation of Schnorr randomness and Church stochasticity: we state a condition on the Kolmogorov complexity of the initial segments which is sufficient for Schnorr randomness but not for Church stochasticity (Proposition 2.2.20 and Theorem 2.2.21). The particular case of the left-c.e. sequences is studied (Proposition 2.2.24).

- We present a new approach of effective randomness based on computable approximations of Kolmogorov complexity (Section 2.3), and we show that they provide a unified framework for many notions of randomness. This work was published in [10].

- We prove a constructive version of Kakutani's theorem (Theorem 3.2.5) using game-theoretic arguments. This generalizes earlier results and can even be used as an alternative proof of Kakutani's theorem. This result appears in [7].

- We give a complete classification of constructive equivalence relations induced by randomness notions on computable probability measures (Subsection 3.3.2). This work was published in [9].

Chapter 1

Randomness notions

This chapter presents a survey of the most popular notions of algorithmic randomness. They arise from two principles: typicalness and unpredictability. Although their formulations are very different, we will see that these two principles are in fact closely related. In particular, the game-theoretic notion of martingale, which we use to support the unpredictability principle, turns out to be sufficient to characterize *all* the notions of randomness that we will introduce, and this allows us to classify them hierarchically. An interesting application of effective randomness to the construction of absolutely normal real numbers is also presented.

1.1 Notation and basic definitions

We assume that the reader is familiar with basic computability notions, like computable functions, computable sets, computably enumerable sets, Turing reduction, oracle etc. We write $\alpha \leq_T \beta$ if α is Turing-reducible to β. We denote by $\mathbf{0}'$ the oracle for the halting problem. If A is a computably enumerable set, for some fixed enumeration of A, we denote by $A[t]$ the elements of A which are enumerated during the first t steps of the enumeration. If f is a computable function, given an input x, we write $f(x)\downarrow$ to express that $f(x)$ is defined. Moreover, for $t \in \mathbb{N}$, we write $f(x)[t]\downarrow$ to express that $f(x)$ is defined and the computation of $f(x)$ requires at most t steps of computation.

We denote the set of STRINGS (i.e. finite sequences of zeros and ones, which we also call WORDS) by $2^{<\omega}$. We denote by $w_{(i)}$ the i-th bit of w (by convention there is a 0-th bit), by $w\restriction_n$ the string made of the first n bits of w (with the convention $w\restriction_n = w$ if $n > |w|$), by ww' the concatenation of two strings w and w', and by ϵ the empty word. The length of a word w is denoted by $|w|$. The PREFIX ORDER on strings is denoted by $\sqsubseteq$ where $w \sqsubseteq w'$ if $|w| \leq |w'|$ and for all $i < |w|$, $w_{(i)} = w'_{(i)}$ (w is then said to be a PREFIX of w'). The corresponding strict order is denoted by $\sqsubset$. A subset of A of $2^{<\omega}$ is said to be PREFIX-FREE if any two elements of A are uncomparable for the prefix-order. Given a property $\mathcal{P}$ on strings, the *set of*

minimal strings satisfying this property is the set of strings satisfying $\mathcal{P}$ such that none of their prefixes satisfy $\mathcal{P}$. The lexicographic order is denoted by $\leq_{lex}$. We will frequently need to identify strings with integers. To do so, we will use the function Bin, where $\text{Bin}(0) = \epsilon$, $\text{Bin}(1) = 0$, $\text{Bin}(2) = 1$, $\text{Bin}(3) = 00$, $\text{Bin}(4) = 01$, $\text{Bin}(5) = 10$, $\text{Bin}(6) = 11$, $\text{Bin}(7) = 000$, etc. Bin is a computable bijection, we denote by Bin^{-1} its inverse. Notice that $|\text{Bin}(n)| = \lfloor \log_2(n+1) \rfloor$ for all n. We abbreviate $\lfloor \log_2 \rfloor$ by log. The number of zeros (resp. of ones) in a string w is denoted by $\#0(w)$ (resp. $\#1(w)$).

We call CANTOR SPACE, and denote it by 2^ω, the set of infinite binary sequences. We will use greek letters $\alpha, \beta, \ldots$ to name elements of this set. We denote by $\alpha_{(i)}$ the i-th bit of a sequence α, and by $\alpha\restriction_n$ the prefix of α of length n (i.e. $\alpha\restriction_n = \alpha_{(0)} \ldots \alpha_{(n-1)}$). We also extend the prefix order to express that a string is a prefix of an infinite sequence, i.e. we have $w \sqsubset \alpha$ for some $w \in 2^{<\omega}$ and $\alpha \in 2^\omega$ if for all $i < |w|$, $w_{(i)} = \alpha_{(i)}$. We also extend the lexicographic order to $2^{<\omega} \cup 2^\omega$, where for all $x, y \in 2^{<\omega} \cup 2^\omega$, $x \leq_{lex} y$ if and only if for every finite prefix x' of x, there exists a finite prefix y' of y such that $x' \leq_{lex} y'$. We usually denote with calligraphic letters $\mathcal{U}$, $\mathcal{V}$, $\mathcal{X}$, etc. the subsets of 2^ω. For $\mathcal{X} \subseteq 2^\omega$, we denote by $\bar{\mathcal{X}}$ the complement of $\mathcal{X}$ in 2^ω. The sequence 0^ω (resp. 1^ω) is the infinite sequence α such that $\alpha_{(i)} = 0$ (resp. $\alpha_{(i)} = 1$) for all i.

We extend the notion of computable function from $\mathbb{N}$ to $\mathbb{N}$ to functions from D to D' whenever D and D' can be identified in a computable way to $\mathbb{N}$ or a subset of $\mathbb{N}$. This includes for example $\mathbb{N} \times \mathbb{N}$, $2^{<\omega}$, $\mathbb{Q}$, $\{0,1\}$ etc. An infinite binary sequence $\alpha \in 2^\omega$ is computable if it is computable when seen as a function from $\mathbb{N}$ to $\{0,1\}$. A function $g : D \to \mathbb{R}$ is computable if there exists a computable function $h : D \times \mathbb{N} \to \mathbb{Q}$ such that for all $(x,t) \in D \times \mathbb{N}$, $\left|h(x,t) - g(x)\right| \leq 2^{-t}$ (that is, g can be effectively approximated by a rational-valued function with any givem precision). A function $g : D \to \mathbb{R}$ is LEFT-COMPUTABLY ENUMERABLE (LEFT-C.E. for short) if there exists a computable function $h : D \times \mathbb{N} \to \mathbb{Q}$ such that for all $x \in D$, the sequence $(h(x,t))_{t \in \mathbb{N}}$ is non-decreasing and converges to $g(x)$. Note that if $f : D \to \mathbb{R}$ is left-c.e. , the set $\{(x,q) \in D \times \mathbb{Q} : f(x) > q\}$ is c.e. A real number r is left-c.e. if the constant function r is left-c.e. A sequence $\alpha \in 2^\omega$ is left-c.e. if it is the binary expansion of a left-c.e. real $r \in [0,1)$. Equivalently, it is left-c.e. if it is the pointwise limit of a computable sequence $(w_n)_{n \in \mathbb{N}}$ of strings that is non-decreasing for the lexicographic order.

For functions from $\mathbb{N}$ to $\mathbb{N}$ (or $\mathbb{R}$), we will extensively use the O and o notation. Given two functions $f, g : \mathbb{N} \to \mathbb{N}$, we say that $f = O(g)$ if there exists a constant $c > 0$ such that $f(n) \leq c\, g(n)$ for all n. We say that $f = o(g)$ if for all $c > 0$, $f(n) \leq c\, g(n)$ for almost all n. When we write a formula like $f(n) \leq g(n) + O(h(n))$ without quantifiers, we mean that there exist a constant c such that for all n: $f(n) \leq g(n) + c\, h(n)$. We say that a function f DOMINATES a function g if $f(n) \geq g(n)$ for almost all n. An ORDER is a non-decreasing unbounded function $h : \mathbb{N} \to \mathbb{N}$. Given

a order f, we define the function f^{-1} by

$$f^{-1}(k) = \min\{n \in \mathbb{N} : \ f(n) \geq k\}$$

which itself is an order. Notice that if f is computable, f^{-1} is computable as well.

1.2 The Cantor space: probability, topology and computability

As we want to approach randomness from a computability point of view, we naturally study randomness on the central space of computability theory: the Cantor space. It is possible to define randomness for many other spaces (see Levin [36], Gacs [21], Hoyrup and Rojas [24]), but this will not be discussed in this thesis. We begin our discussion by a quick review of the topological properties of the Cantor space.

1.2.1 The topology

The canonical topology on the Cantor space is the product topology (the Cantor space can be viewed as the product of countably many copies of $\{0, 1\}$ each of them endowed with the discrete topology). The product topology is generated by the CYLINDERS $[w]$ (where $w \in 2^{<\omega}$) defined by:

$$[w] = \{\alpha \in 2^\omega : \ w \sqsubset \alpha\}$$

As the cylinders form a basis for the product topology, every open set can be written as a union of cylinders, i.e. as $\bigcup_{w\in A}[w]$ for some $A \subseteq 2^{<\omega}$. For any $A \subseteq 2^\omega$, we abbreviate $\bigcup_{w\in A}[w]$ by $[A]$, which we call the open set GENERATED BY A.

Remark 1.2.1. *It should be noticed that for any two strings w and w', if $w \sqsubseteq w'$, then $[w'] \subseteq [w]$ (and if w and w' are incomparable for the prefix order, then $[w]$ and $[w']$ are disjoint). Hence, when writing an open set as a union of cylinders $\bigcup_{w\in A}[w]$, one can remove from A all the strings w' such that there exist $w \sqsubset w'$ in A. We obtain a subset A' of A such that $[A'] = [A]$ and A' is prefix-free (according to our terminoloy, A' is the set of minimal strings in A). This proves that every open set is generated by a prefix-free set of strings.*

Remark 1.2.2. *Another important fact is that every cylinder $[w]$ is also a closed set. Indeed, the complement of $[w]$ is the set of sequences α such that w* is not *a prefix of α. But this is equivalent to say that one of the strings of length $|w|$ that are different from w is a prefix of α. Hence $2^\omega \setminus [w] = \bigcup\{[u] : u \neq w \wedge |u| = |w|\}$ is open.*

Recall also that the Cantor space is metrizable, via the Cantor distance

$$\partial_C(\alpha, \beta) = 2^{-\min\{i:\ \alpha_{(i)} \neq \beta_{(i)}\}}$$

1.2.2 Effectivizing the topology

In order to define *effective* randomness, we will need to restrict our attention to topological objects in the Cantor space that can be effectively (i.e. computably) described. The most fundamental objects of that category are *effectively open sets*:

Definition 1.2.3. *An open set $\mathcal{U} \subseteq 2^\omega$ is call an* EFFECTIVELY OPEN SET *or a* COMPUTABLY ENUMERABLE OPEN SET *(which we abbreviate by* C.E. OPEN SET*) if there exists a computably enumerable subset $\{u_i : i \in \mathbb{N}\}$ of $2^{<\omega}$ such that*

$$\mathcal{U} = \bigcup_{i \in \mathbb{N}} [u_i]$$

Remark 1.2.4. *We saw above that any open set $\mathcal{U}$ of the Cantor set can be written as $\mathcal{U} = [A]$, where A is a prefix-free subset of $2^{<\omega}$. The effective version of this remark holds true: in the above definition, we can assume that the u_i's form a prefix-free set. Indeed, whenever some u_i is enumerated, if there exists $j < i$ such that $u_j \sqsubseteq u_i$, then u_i can be omitted (as $[u_i] \subseteq [u_j]$), and if there exists $j < i$ such that $u_i \sqsubseteq u_j$, then instead of u_i, one can enumerate all the extensions of u_i of length $|u_j|$ that are different from u_j.*

We also define:

Definition 1.2.5. *A sequence $(\mathcal{U}_n)_{n \in \mathbb{N}}$ of open sets is called a* COMPUTABLE SEQUENCE OF C.E. OPEN SETS *if there exists a sequence of subsets $A_n = \{u_i^{(n)} : i \in \mathbb{N}\}$ of $2^{<\omega}$ which are computably enumerable uniformly in n and such that*

$$\mathcal{U}_n = \bigcup_{i \in \mathbb{N}} [u_i^{(n)}]$$

for all n.

Remark 1.2.6. *Similarly to Remark 1.2.6, we can assume in the above definition that for all n, the set $\{u_i^{(n)} : i \in \mathbb{N}\}$ is prefix-free.*

1.2.3 Lebesgue measure

Of course, if we want to talk about randomness, we will have to talk about probability measures. Most of the work in algorithmic randomness is focused on randomness with respect to the uniform measure, also known as LEBESGUE MEASURE. The Lebesgue measure of a subset $\mathcal{X}$ of 2^ω (denoted by $\lambda(\mathcal{X})$) is the probability that $\alpha \in \mathcal{X}$ when α is generated by independent tosses of a balanced coin. Hence, the measure $\lambda([w])$ of every cylinder $[w]$ must be $2^{-|w|}$. We can then extend λ to a larger class of subsets of 2^ω as follows. Let $\mathcal{X} \subseteq 2^\omega$. The outer measure λ^* of $\mathcal{X}$ is the quantity:

$$\lambda^*(\mathcal{X}) = \inf \left\{ \sum_{w \in A} 2^{-|w|} : A \subseteq 2^{<\omega} \wedge \mathcal{X} \subseteq \bigcup_{w \in A} [w] \right\}$$

The inner measure of $\mathcal{X}$ is then defined as $\lambda_*(\mathcal{X}) = 1 - \lambda^*(\bar{\mathcal{X}})$.

Definition 1.2.7. *Let $\mathcal{X} \subseteq 2^\omega$. If $\lambda^*(\mathcal{X}) = \lambda_*(\mathcal{X})$, then $\mathcal{X}$ is said to be* MEASURABLE*, and its measure $\lambda(\mathcal{X})$ is defined to be $\lambda^*(\mathcal{X})$. If $\mathcal{X}$ has measure 0, it is called a* NULLSET*.*

All Borel subsets of 2^ω are Lebesgue measurable, which is all we need for the rest of this thesis as the subsets of 2^ω we will consider are all Borel sets. In Chapter 3, we will discuss effective randomness with respect to other probability measures but until then, each time we talk about "measure", we implicitely mean "Lebesgue measure".

1.3 The typicalness paradigm

The first satisfactory definition of effective randomness for an infinite sequence (still considered to be the best nowadays) was given by P. Martin-Löf [42]. Martin-Löf's intuition was the following: a random infinite binary sequence α should belong to no subset of 2^ω that has probability 0. In other words, it should satisfy all the laws that have probability 1, like the Law of Large Numbers or the Law of Iterated Logarithm. Of course, stated like that, this is impossible because α belongs to the singleton $\{\alpha\}$ which has probability 0. Hence, instead of requiring α to avoid all nullsets, Martin-Löf called "random" the sequences that avoid all *effective* nullsets.

Typicalness paradigm.
An infinite binary sequence is random if it belongs to no effective nullset.

1.3.1 Martin-Löf tests

Of course, there are many ways to interpret this paradigm, as there is no unique formalization of the concept of "effective nullset" (see the discussion below on Schnorr's critique of Martin-Löf randomness). Here is how Martin-Löf understood it:

Definition 1.3.1. *A* MARTIN-LÖF TEST *is a computable sequence $(\mathcal{U}_n)_{n\in\mathbb{N}}$ of c.e. open sets such that for all n, $\lambda(\mathcal{U}_n) \leq 2^{-n}$.*
For every Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$, every subset of $\bigcap_{n\in\mathbb{N}} \mathcal{U}_n$ is called a MARTIN-LÖF NULLSET*.*
A sequence $\alpha \in 2^\omega$ is said to be MARTIN-LÖF RANDOM *if it belongs to no Martin-Löf nullset. We denote by* **MLR** *the set of Martin-Löf random sequences.*

The term 2^{-n} in this definition is arbitrary. Any other computable function tending to 0 would yield the same notion of randomness:

Lemma 1.3.2. *Let $f : \mathbb{N} \to \mathbb{R}$ be a computable function with $\lim_n f(n) = 0$. A sequence α is Martin-Löf random if and only if for every computable sequence of c.e. open sets $(\mathcal{U}_n)_{n\in\mathbb{N}}$ such that $\lambda(\mathcal{U}_n) \leq f(n)$ for all n, $\alpha \notin \bigcap_n \mathcal{U}_n$*

Proof. Let $f : \mathbb{N} \to \mathbb{R}$. Suppose a sequence α is not Martin-Löf random i.e. it belongs to the intersection of a computable sequence of c.e. open sets $\{\mathcal{U}_n : n \in \mathbb{N}\}$ such that $\lambda(\mathcal{U}_n) \leq 2^{-n}$. Since f is computable, for all n, one can effectively find an integer $k(n)$ such that $2^{-k(n)} \leq f(n)$. Then, $(\mathcal{U}_{k(n)})_{n\in\mathbb{N}}$ is a computable sequence of c.e. open sets such that $\lambda(\mathcal{U}_{k(n)}) \leq f(n)$ for all n, and $\alpha \in \bigcap_n \mathcal{U}_{k(n)}$. Conversely, let $\{\mathcal{V}_n : n \in \mathbb{N}\}$ be a computable sequence of c.e. open sets such that $\lambda(\mathcal{V}_n) \leq f(n)$ for all n and $\alpha \in \bigcap_n \mathcal{V}_n$. Since f tends to 0, for all n, one can effectively find some integer $k'(n)$ such that $f(k'(n)) \leq 2^{-n}$. Then, the sequence $(\mathcal{V}_{k'(n)})_{n\in\mathbb{N}}$ is a Martin-Löf test, whose intersection contains α, hence α is not Martin-Löf random. □

Informally, a sequence $\alpha \in 2^\omega$ is not Martin-Löf random if for any given precision ε, one can, uniformly in ε, provide an open subset $\mathcal{U} \subseteq 2^\omega$ of measure smaller than ε which contains α. Of course, for any Martin-Löf test $\{\mathcal{U}_n : n \in \mathbb{N}\}$, the set of sequences that belong to $\bigcap_{n\in\mathbb{N}} \mathcal{U}_n$ (we say that these sequences are COVERED by the test) has measure 0. And since there are only countably many Martin-Löf tests (by the computability requirement), the class **MLR** has measure 1.

There are several reasons why Martin-Löf's definition of randomness is considered to be the best one. The main reason is that, similarly to the notion of computability which came from many definitions (Recursive functions, Turing machines, lambda-calculus, Markov model, etc.) later proven equivalent, Martin-Löf randomness arises naturally from different intuitions one can have on randomness (see later the "unpredictability paradigm" and "incompressibility paradigm"). Another reason for the popularity of this notion is the existence of a *universal* element in the objects used to characterize it, a property that none of the other notions of randomness we will consider have.

Proposition 1.3.3. *There exists a universal Martin-Löf test, that is a Martin-Löf test $(\mathcal{V}_n)_{n\in\mathbb{N}}$ such that*

$$\alpha \in \mathbf{MLR} \Leftrightarrow \alpha \notin \bigcap_{n\in\mathbb{N}} \mathcal{V}_n$$

This means that among Martin-Löf nullsets, there is actually a largest one, which is exactly the set of sequences that are not Martin-Löf random.

Proof. First, notice that one can effectively enumerate Martin-Löf's tests: given an index for a computable sequence of c.e. open sets $(\mathcal{W})_{i\in\mathbb{N}}$, one can enumerate the $\mathcal{W}_i$ until a stage t such that $\lambda(\mathcal{W}_i[t]) > 2^{-i}$ for some $i \in \mathbb{N}$. In that case, the enumeration is stopped *before* stage t. Now, given an enumeration of Martin-Löf tests $\mathcal{U}^{(n)}$ (by this we mean that for all n, $(\mathcal{U}_i^{(n)})_{i\in\mathbb{N}}$ is an Martin-Löf test and all Martin-Löf tests appear in the sequence), we set

$$\mathcal{V}_n = \bigcup_{k\in\mathbb{N}} \mathcal{U}^{(k)}_{n+1+k}$$

The $\mathcal{V}_n$ form a computable sequence of c.e. open sets (a computable union of c.e. open sets is obviously an open set as a computable union of c.e. subsets of $2^{<\omega}$ is

a c.e. subset of $2^{<\omega}$) and for all n:

$$\lambda\left(\mathcal{V}_n\right) \leq \sum_{k\in\mathbb{N}} \lambda(\mathcal{U}^{(k)}_{n+1+k}) \leq \sum_{k\in\mathbb{N}} 2^{-n-k-1} \leq 2^{-n}$$

Finally, if $\alpha \notin \mathbf{MLR}$, α belongs to a Martin-Löf nullset, i.e. for some m, $\alpha \in \bigcap_{k\in\mathbb{N}} \mathcal{U}^{(m)}_k$. But then for all $k > m$, $\mathcal{U}^{(m)}_k \subseteq \mathcal{V}_{k-m-1}$, hence $\alpha \in \bigcap_{n\in\mathbb{N}} \mathcal{V}_n$. ■

In classical probability theory, the Borel-Cantelli lemma asserts that, given a sequence of sets $(\mathcal{A}_n)_{n\in\mathbb{N}}$ such that $\sum_n \lambda(\mathcal{A}_n) < +\infty$, the set of sequences α that belong to infinitely many $\mathcal{A}_n$ has measure 0. Martin-Löf randomness can be characterized by an effectivization of this result:

Theorem 1.3.4 (Solovay [55], Shen [54])**.** *A sequence $\alpha \in 2^\omega$ is Martin-Löf random if and only if for every computable sequence of c.e. open sets $(\mathcal{U}_n)_{n\in\mathbb{N}}$ such that $\sum_n \lambda(\mathcal{U}_n) < +\infty$, α belongs only to finitely many $\mathcal{U}_n$.*

Proof. First, it is easy to see that a sequence α which has this property is Martin-Löf random. Indeed, a Martin-Löf test is a computable sequence of c.e. open sets $(\mathcal{U}_n)_{n\in\mathbb{N}}$ such that $\lambda(\mathcal{U}_n) \leq 2^{-n}$ for all n. In particular, this implies $\sum_n \lambda(\mathcal{U}_n) < +\infty$. Hence, if α has the above effective Borel-Cantelli property, it cannot belong to the intersection of the $\mathcal{U}_n$.

Conversely, let $(\mathcal{U}_n)_{n\in\mathbb{N}}$ be a computable sequence of c.e. open sets such that $\sum_n \lambda(\mathcal{U}_n) < C$ for some $C \in \mathbb{N}$. Given $k \in \mathbb{N}$, the set

$$\mathcal{V}_k = \{\alpha \in 2^\omega : \ \alpha \text{ belongs to at least } k \text{ sets } \mathcal{U}_n\}$$

is a c.e. open set (uniformly in $\mathbb{N}$) and has measure at most C/k. By Lemma 1.3.2, the intersection of the $\mathcal{V}_k$ is a Martin-Löf nullset. By definition of the $\mathcal{V}_k$, this intersection is exactly the set of sequences that belong to infinitely many $\mathcal{U}_n$. Hence, any Martin-Löf random sequence belongs to finitely many $\mathcal{U}_n$. ■

1.3.2 Schnorr randomness

In [52], Schnorr raised a criticism against Martin-Löf randomness, arguing that it is not *effective* enough. Schnorr was motivated by a game/prediction point of view, and he noticed that the knowledge that $\alpha \in \mathcal{U}$ for a c.e. open set of small measure is not enough to predict effectively the bits of α. We will come back to this later, when we will discuss the *unpredictability paradigm*. For now, let us just give the definition of Schnorr randomness. It is similar to the definition of Martin-Löf randomness, but in the underlying notion of test, we require the open sets to have *exactly* measure 2^{-n}.

Definition 1.3.5. *A* SCHNORR TEST *is a computable sequence* $\{\mathcal{U}_n : n \in \mathbb{N}\}$ *of c.e. open sets such that for all* n, $\lambda(\mathcal{U}_n) = 2^{-n}$.
For every Schnorr test $\{\mathcal{U}_n : n \in \mathbb{N}\}$, *any set contained in intersection* $\bigcap_{n\in\mathbb{N}} \mathcal{U}_n$ *is called a* SCHNORR NULLSET.
A sequence $\alpha \in 2^\omega$ *is said to be* SCHNORR RANDOM *if it belongs to no Schnorr nullset. We denote by* **SR** *the set of Schnorr random sequences.*

Here again, 2^{-n} can be replaced by any positive computable function tending to 0.

Lemma 1.3.6. *A sequence* $\alpha \in 2^\omega$ *is not Schnorr random if and only if it belongs to the intersection of a computable sequence of c.e. open sets* $(\mathcal{U}_n)_{n\in\mathbb{N}}$ *such that the measure of* $\mathcal{U}_n$ *is computable uniformly in* n *and tends to* 0.

Proof. The "only if" part is immediate as 2^{-n} is computable uniformly in n and tends to 0 as n tends to $+\infty$. For the "if" part, suppose that there exists a computable sequence of c.e. open sets $(\mathcal{U}_n)_{n\in\mathbb{N}}$ whose intersection contains α and such that $f(n) = \lambda(\mathcal{U}_n)$ is computable uniformly in n and tends to 0. For all n, one can effectively find some integer $k(n)$ such that $f(k(n)) < 2^{-n}$. Set $f' = f(k(n))$, and $\mathcal{V}_n = \mathcal{U}_{k(n)}$. Then, the $\mathcal{V}_n$ is a c.e. open set uniformly in n with $\lambda(\mathcal{V}_n) = f'(n) < 2^{-n}$ (and f' is computable), and $\alpha \in \bigcap_n \mathcal{V}_n$. All we need to do is to make each $\mathcal{V}_n$ bigger so its measure attains 2^{-n}. Fix an n. We construct by induction an increasing sequence $(A_i)_{i\in\mathbb{N}}$ of finite subsets of $2^{<\omega}$ such that for all i:

$$\lambda(\mathcal{V}_n \cup [A_i]) < 2^{-n} < \lambda(\mathcal{V}_n \cup [A_i]) + 2^{-i} \tag{1.1}$$

Set $A_0 = \emptyset$. If A_i has been constructed, since $\lambda(\mathcal{V}_n)$, so is $\lambda(\mathcal{V}_n \cup [A_i])$, and so is $\lambda(\mathcal{V}_n \cup [A_i] \cup [B])$ for any given finite subset B of $2^{<\omega}$. Hence, one can effectively find a B satisfying

$$\lambda(\mathcal{V}_n \cup [A_i] \cup [B]) < 2^{-n} < \lambda(\mathcal{V}_n \cup [A_i] \cup [B]) + 2^{-i-1}$$

Set $A_{i+1} = A_i \cup B$, completing the induction. Since all the steps of the induction are effective, the union A of the A_i is a c.e. subset of $2^{<\omega}$, hence $\mathcal{V}'_n = \mathcal{V}_n \cup [A]$ is a c.e. open set containing $\mathcal{V}_n$ (hence containing α) and

$$\lambda(\mathcal{V}_n \cup [A]) = \lim_{i\to\infty} \lambda(\mathcal{V}_n \cup [A_i]) = 2^{-n}$$

by equation (1.1). Hence, the $\mathcal{V}'_n$ form a Schnorr test covering α, i.e. α is not Schnorr random.

□

1.3.3 Weak randomness

Weak randomness, as it was introduced by Kurtz [32] (it is sometimes called KURTZ RANDOMNESS) is also based on the typicalness paradigm, but understood in the opposite way: instead of avoiding all effectively null sets, a sequence should be random if it *belongs* to all effective sets of measure 1:

Definition 1.3.7. *A sequence $\alpha \in 2^\omega$ is* WEAKLY RANDOM *if it belongs to all c.e. open sets of measure* 1. *We denote by* **WR** *the set of weakly random sequences.*

It turns out that, although it is interesting to study, this notion of randomness cannot be completely satisfactory as a weakly random sequence need not satisfy the Law of Large Numbers. We will prove this in the sequel.

1.3.4 Effective Hausdorff dimension

General definition

Hausdorff dimension was introduced by F. Hausdorff [22] as an attempt to assign a dimension to every subset of a metric space. For example, in the space $\mathbb{R}^3$ (endowed with the Euclidean distance), it is our intuition that a point, a circle, a sphere and a ball have dimension respectively 0, 1, 2 and 3. The notion of measure is too coarse to make this distinction, as the first three objects are nullsets (for Lebesgue measure). Hausdorff dimension on the other hand allows us to formalize our intuition.

In great generality, in a metric space $(\mathcal{X}, d)$, we define for all $\mathcal{E} \subseteq \mathcal{X}$, $\delta > 0$ and $s > 0$:

$$H^s_\delta(\mathcal{E}) = \inf \left\{ \sum_{k \in \mathbb{N}} \operatorname{diam}(\mathcal{A}_k)^s \right\}$$

where the infimum is taken over the sequences $(\mathcal{A}_k)_{k\in\mathbb{N}}$ of subsets of $\mathcal{X}$ that cover $\mathcal{E}$ (i.e. $\mathcal{E} \subseteq \bigcup_k \mathcal{A}_k$) and such that all $\mathcal{A}_k$ all have a diameter of at most δ. For a fixed s, this quantity is non-decreasing as δ tends to 0, and we set

$$H^s(\mathcal{E}) = \lim_{\delta \to 0} H^s_\delta(\mathcal{E})$$

Clearly $H^s(\mathcal{E})$ is non-decreasing as s tends to 0. In fact, it can be shown that for every space $\mathcal{E}$, there exists a threshold s_0 such that $H^s(\mathcal{E}) = 0$ for all $s > s_0$ and $H^s(\mathcal{E}) = +\infty$ for all $s < s_0$. We call HAUSDORFF DIMENSION of $\mathcal{E}$, and we denote by $\dim(\mathcal{E})$ this threshold s_0. This definition of dimension works for the objects we mentionned above (point, circle etc). More interestingly, the dimension of a set needs not be an integer. For example, the Hausdorff dimension of von Koch's snowflake is $\log(4)/\log(3)$. For more on Hausdorff dimension on general metric spaces, see the standard reference Falconer [17].

Hausdorff dimension in the Cantor space

The particular topology of the Cantor space makes the study of Hausdorff dimension simple. First, notice that the definition of the Cantor distance ∂_C (see page 3), every subset of 2^ω has a diameter of type 2^{-k} with $k \in \mathbb{N} \cup \{+\infty\}$. Moreover, if $\mathcal{A}$ has diameter 2^{-k}, all its elements have a common prefix of length k. Calling w this prefix, we have $\mathcal{A} \subseteq [w]$. But $[w]$ itself has diameter 2^{-k}. Hence, in the above definition of H^s_δ, we can restrict our attention to the particular case where δ is a negative power of 2 and the $\mathcal{A}_k$ are cylinders. This implies:

Proposition 1.3.8. *Let $\mathcal{E}$ be a subset of 2^ω. We have $H^s(\mathcal{E}) = 0$ if and only if for all $\varepsilon > 0$, there exists a family $([w_k])_{k\in\mathbb{N}}$ of cylinders such that $\mathcal{E} \subseteq \bigcup_{k\in\mathbb{N}}[w_k]$ and $\sum_k 2^{-s|w_k|} < \varepsilon$ The dimension $\dim(\mathcal{E})$ is then the infimum over those s for which $H^s(\mathcal{E}) = 0$.*

We now make things effective:

Definition 1.3.9. *Let $s > 0$. A* CONSTRUCTIVE s-TEST *is a computable sequence $(A_n)_{n\in\mathbb{N}}$ of c.e. subsets of $2^{<\omega}$ such that for all n:*

$$\sum_{w\in A_n} 2^{-s|w|} \leq 2^{-n}$$

A set $\mathcal{E} \subseteq 2^\omega$ is CONSTRUCTIVELY s-NULL *if there exist a constructive s-test $(A_n)_{n\in\mathbb{N}}$ such that $\mathcal{E} \subseteq [A_n]$ for all n.*
The CONSTRUCTIVE DIMENSION *of a set $\mathcal{E} \subseteq 2^\omega$ is defined by*

$$\mathrm{cdim}(\mathcal{E}) = \inf\{s > 0 :\ \mathcal{E} \text{ is constructively } s\text{-null}\}$$

For any 1-test $(A_n)_{n\in\mathbb{N}}$, the computable sequence of c.e. open sets $([A_n])_{n\in\mathbb{N}}$ form a Martin-Löf test. This means that if $\alpha \in 2^\omega$ is a Martin-Löf random sequence, the singleton $\{\alpha\}$ is not constructively 1-null, hence has constructive Hausdorff dimension of at least 1. This can seem a little suprising at first (since the classical dimension of a singleton is always 0), but this is exactly the same phenomenon as for Martin-Löf randomess: a singleton $\{\alpha\}$ always has measure 0 but if α is Martin-Löf random, it cannot be covered effectively by smaller and smaller open sets. From now on, for a given sequence $\alpha \in 2^\omega$, we abbreviate $\mathrm{cdim}(\{\alpha\})$ by $\mathrm{cdim}(\alpha)$.

One can also define an even stronger notion of s-tests, computable s-tests, which leads to computable Hausdorff dimension.

Definition 1.3.10. *Let $s > 0$. A* COMPUTABLE s-TEST *is a computable sequence $(A_n)_{n\in\mathbb{N}}$ of computable subsets of $2^{<\omega}$ such that for all n:*

$$\sum_{w\in A_n} 2^{-s|w|} \leq 2^{-n}$$

A set $\mathcal{E} \subseteq 2^\omega$ is COMPUTABLY s-NULL *if there exist a computable s-test $(A_n)_{n\in\mathbb{N}}$ such that $\mathcal{E} \subseteq [A_n]$ for all n.*
The COMPUTABLE DIMENSION *of a set $\mathcal{E} \subseteq 2^\omega$ is defined by*

$$\dim_{\mathrm{comp}}(\mathcal{E}) = \inf\{s > 0 :\ \mathcal{E} \text{ is computably } s\text{-null}\}$$

Let us make a few remarks about dimension. First, one should notice that, from the very definition of computable, constructive and classical Hausdorff dimension, for all $\mathcal{E} \subseteq 2^\omega$, one has:

$$\dim(\mathcal{E}) \leq \mathrm{cdim}(\mathcal{E}) \leq \dim_{\mathrm{comp}}(\mathcal{E})$$

Also, it is clear that all three notions of dimension are monotonic in the sense that if $\mathcal{E} \subseteq \mathcal{E}'$ then $\dim(\mathcal{E}) \leq \dim(\mathcal{E}')$ (and the same holds true for constructive dimension). This allows us to prove:

Remark 1.3.11. *Every subset $\mathcal{E}$ of 2^ω has a dimension (classical, constructive, or computable) lying between 0 and 1.*

The fact that the dimension is non-negative comes from the definition. To see that the dimension of any $\mathcal{E}$ is no greater than one it suffices to prove that $\mathrm{cdim}(2^\omega) = 1$ as $\dim(\mathcal{E}) \leq \mathrm{cdim}(\mathcal{E}) \leq \mathrm{cdim}(2^\omega)$. As 2^ω contains Martin-Löf random sequences, 2^ω cannot be covered by any constructive 1-test, hence $\mathrm{cdim}(2^\omega) \geq 1$. Let now s be a rational number greater than 1. For all n, we can uniformly compute an integer k_n such that $2^{(1-s)k_n} \leq 2^{-n}$ (as $s > 1$). For all n, let A_n be the set of strings of length k_n. Clearly, $[A_n]$ covers 2^ω for all n and

$$\sum_{w \in A_k} 2^{-s|w|} = 2^{k_n}\, 2^{-s\, k_n} = 2^{(1-s)k_n} \leq 2^{-n}$$

Hence $(A_n)_{n\in\mathbb{N}}$ is a coomputable s-test covering 2^ω, meaning that 2^ω is constructively s-null. Being true for all $s > 1$, this proves that $\dim_{\mathrm{comp}}(2^\omega) \leq 1$ (hence $\mathrm{cdim}(2^\omega) \leq 1$ and $\dim(2^\omega) \leq 1$).

1.4 The unpredictability paradigm, selection rules, martingales and strategies

Another way of apprehending randomness is via the so-called *unpredictability paradigm*. According to this paradigm, the essence of a random experiment is that no reasonable prediction can be made on its outcomes.

Unpredictability paradigm.
An infinite binary sequence is random if it is impossible to effectively predict its bits with good accuracy.

This point of view especially makes sense from a game-theoretic perspective. In a gambling game (head-tail, roulette...), players make bets on the outcome of a random trial. The sequence is random (at least, sufficiently random from the bank's point of view) if one cannot predict the future outcomes (given the past outcomes) with a good success rate. We shall present different notions of randomness that are based on this intuition.

1.4.1 Stochasticity

To modelize the notion of prediction, imagine that the bits of an infinite sequence α are written on cards (one bit per card) which lie on a table, face down. We try to guess the value of the first card, then the first card is revealed. We then try to predict the value of the second card, which is then revealed and so on (of course,

at each stage we can use the previously revealed bits to make our prediction). The Law of Large Numbers tells us that if the sequence is actually random, we should be right as often as we are wrong, asymptotically (that is, the limit of the success rate should tend to 50%). Thus, we could define a notion of randomness by saying that a sequence is random if there is no computable way to make predictions with an upper limit of success that exceeds 50%. But we need to correct this intuition a little. Indeed, suppose the bits of α are actually chosen at random, except for the $\alpha_{(i)}$ where i is a power of 2, which are all equal to 0. Clearly the sequence should not be considered random as we can predict these bits with certainty. However, the density of such bits is negligeable in the whole sequence. Hence, in this case, our limit success rate will still be of 50%. This can be corrected by allowing the player to make a guess only if he wants to. That is, at each move, the player can either make a guess or simply ask to see the value of the card without making any guess. In the above example, this solves the problem, as we will the player will then only make a prediction the values of the $\alpha_{(i)}$'s when i is a power of 2. This leads to a notion of randomness called Church stochasticity (the idea of selection rule was proposed by von Mises [59], but it was Church who suggested to consider only computable ones) which we now formalize.

Definition 1.4.1. *A* SELECTION RULE *is a total function* $\sigma : 2^{<\omega} \to \{\texttt{select},\texttt{scan}\}$

For any $w \in 2^{<\omega}$, $\sigma(w)$ represents the choice made by the player at stage $n = |w|$ when the first n bits he has seen are $w_{(0)}, \dots, w_{(n-1)}$. For all $w \in 2^{<\omega}$, we denote by $\sigma[w]$ the sequences of bits selected after having read w. We formally define it by induction on $|w|$: $\sigma[\epsilon] = \epsilon$ and if $\sigma[w]$ is already defined, for all $\iota \in \{0, 1\}$:

- if $\sigma(w) = \texttt{scan}$, set $\sigma[w\iota] = \sigma[w]$
- if $\sigma(w) = \texttt{select}$, set $\sigma[w\iota] = \sigma[w]\iota$

For a fixed $\alpha \in 2^\omega$, the sequence of strings $\left(\sigma[\alpha\restriction_n]\right)_{n\in\mathbb{N}}$ is non-decreasing for the prefix order $\sqsubseteq$. Hence, either it is stationnary, in which case we set $\sigma[\alpha]$ to be the limit of the sequence. Or the sequence is not stationnary, in which case the $\sigma[\alpha\restriction_n]$ are all prefixes of an infinite binary sequence, which we call $\sigma[\alpha]$. In both cases, $\sigma[\alpha]$ is called the SUBSEQUENCE OF α SELECTED BY σ.

We are ready to define Church stochasticity.

Definition 1.4.2. *A sequence* α *is* CHURCH STOCHASTIC *if for every total computable selection rule, either* $\beta = \sigma[\alpha]$ *is finite, or it satisfies*

$$\lim_{n\to+\infty} \frac{\#0(\beta\restriction_n)}{n} = \frac{1}{2}$$

We denote by **ChStoch** *the set of Church stochastic sequences.*

Although this notion is interesting, it can hardly be considered a suitable notion of effective randomness. Indeed, Ville [58] was able to show that there exists a sequence α that is Church stochastic and such that for all n, $\#0(\alpha\restriction_n) > \#1(\alpha\restriction_n)$. The set of sequences having this property is highly "effective", and has measure 0 by the Law of Iterated Logarithm (which states that the quantity $\#0(\alpha\restriction_n) - \#1(\alpha\restriction_n)$ should more or less oscillate between $-\sqrt{2n\log\log n}$ and $\sqrt{2n\log\log n}$ with probability 1). We will now see how Church stochasticity can be improved by considering betting strategies instead of selection rules. This leads us to the notion of computable randomness.

1.4.2 Computable randomness

In order to define computable randomness, we consider again the situation where we are trying to predict the values of the bits of a sequence α all the bits of which are initially hidden. But this time, instead of the binary choice `select`/`scan`, we will allow the player to bet some amount of money on the values of the bits. The game goes as follows. Initially, the player starts with a positive capital. During the n-th move, the player (based on his knowledge of the first $n-1$ bits) bets an amount of money – which can be anything between 0 and his current capital – on the value of the n-th bit. The bit is then revealed. If his guess was correct, the player doubles his stake; otherwise he loses his stake. The player is said to SUCCEED against α if his capital takes on arbitrarily large values throughout the infinite game. In order to make things completely formal, we introduce the fundamental notion of martingale.

Definition 1.4.3. *A* MARTINGALE *is a function* $d : 2^{<\omega} \to \mathbb{R}_+$ *such that for all* $w \in 2^{<\omega}$*:*

$$d(w) = \frac{d(w0) + d(w1)}{2}$$

we say that d *is* NORMED *if* $d(\epsilon) = 1$.

Intuitively, a martingale represents the capital of a player during the game, the condtion $d(w0) + d(w1) = 2d(w)$ ensuring that the game is fair. More precisely, $d(w)$ represents the capital of the player after betting on the first $n = |w|$ bits of a sequence whose prefix of length n is w. For example, if $d(\epsilon) = 1$, $d(0) = 3/2$ and $d(1) = 1/2$, this means that the player bets that the value of the first bit is 0 with a stake of $1/2$. Accordingly we set:

$$\mathrm{Stake}(d, u) = d(u0) - d(u) \text{ and } \mathrm{Bet}(d, u) = \frac{\mathrm{Stake}(d, u)}{d(u)}$$

where we let $\mathrm{Bet}(d, u) = 0$ when $d(u) = 0$.

In this notation, $\mathrm{Stake}(d, u)$ represents the amount of money that d, having read u, bets on the value of the next bit to be 0. This amount can be negative: if d bets an amount a on the value 1, this can be seen as a bet on the value 0 with stake $-a$. $\mathrm{Bet}(d, u)$ represents the *fraction* of the capital d bets after having

read u, with the same convention that $\mathrm{Bet}(d,u)$ can be negative. Notice that for all $u \in 2^{<\omega}$:

$$d(u0) = d(u) + \mathrm{Stake}(d,u) = (1 + \mathrm{Bet}(d,u))d(u)$$

and

$$d(u1) = d(u) - \mathrm{Stake}(d,u) = (1 - \mathrm{Bet}(d,u))d(u)$$

Remark 1.4.4. *The notion of "martingale" is fundamental in probability theory, where it can be defined with a very high level of generality (using filtrations of sigma-algebras; see for example Jacod and Protter [25]). The above definition of "martingale" is a very particular case of the classical notion.*

Remark 1.4.5. *In the sequel, we will very often use the following simple fact: any finite sum of martingales is a martingale; and for every sequence $(d_n)_{n\in\mathbb{N}}$ of martingales and every sequence $(r_n)_{n\in\mathbb{N}}$ of nonnegative real numbers, if $\sum_n r_n\, d_n(\epsilon) < +\infty$, then $\sum_n r_n\, d_n$ is a martingale.*

Definition 1.4.6. *For a martingale d and a sequence $\alpha \in 2^\omega$ we define*

$$\bar{d}(\alpha) = \limsup_{n\to+\infty} d(\alpha\restriction_n)$$

A martingale SUCCEEDS *on a sequence α if $\bar{d}(\alpha) = +\infty$. We call* SUCCESS SET OF *d, and denote by* $\mathrm{Succ}(d)$, *the set of sequences on which d succeeds.*

This allows us to give the following definition:

Definition 1.4.7. *A sequence $\alpha \in 2^\omega$ is* COMPUTABLY RANDOM *if there is no computable martingale that succeeds on it. We denote by* **CR** *the set of computably random sequences.*

It is not immediately clear why this is a notion of randomness i.e. that the set of computably random sequences has measure 1. We will discuss this in a moment (see page 23). For now, we make an important remark:

Remark 1.4.8. *The notion of computable randomness remains the same if we define the success of a martingale by* $\lim\ d(\alpha\restriction_n) = +\infty$ *instead of* $\limsup\ d(\alpha\restriction_n) = +\infty$. *To see this, imagine that a player has a strategy ensuring his capital reaches arbitrarily large values. Each time his capital reaches a certain threshold (say 2), he puts half of his capital on a "bank account" and keep playing with the rest, making the same bets in terms of fraction of the capital. Then the capital will grow unboundedly again, reach the threshold etc. The capital on the bank account is then non-decreasing and unbounded, ensuring that the* limit *of the capital is $+\infty$.*

Remark 1.4.9. *The notion of computable randomness remains the same if we replace "computable martingale" by "computable normed martingale" in the definition. This is because if d is a computable martingale that succeeds on α, the initial capital $d(\epsilon)$ is computable, hence the martingale $d' = d/d(\epsilon)$ is computable, normed, and also succeeds against α as $d'(\alpha\restriction_n)$ is equal to $d(\alpha\restriction_n)$ up to a multiplicative constant.*

Another important fact about computable randomness is that, unlike Martin-Löf tests for Martin-Löf randomness, there is no universal element:

Proposition 1.4.10. *There exists no computable martingale d such that* $\mathrm{Succ}(d') \subseteq \mathrm{Succ}(d)$ *for every computable martingale d'. In other words, there is no computable martingale d that succeeds on every sequence α that is not computably random. In fact, for every computable martingale d, there exists $\alpha \in 2^\omega$ that is computable and that is not in* $\mathrm{Succ}(d)$.

Proof. We prove the second part of the proposition, which implies the first one (because a computable α is obviously not computably random!). This is done by diagonalizing against d, by induction. Let c be an integer such that $d(\epsilon) < c$. Set $w^{(0)} = \epsilon$. For all n, suppose that $w^{(n)}$ is already defined and $d(w^{(n)}) < c$. Then, by the fairness condition, there exists $\iota \in \{0,1\}$ such that $d(w^{(n)}\iota) < c$ and such a ι can be found effectively since d is computable. Then, set $w^{(n+1)} = w^{(n)}\iota$. The $w^{(n)}$ form an increasing sequence for the prefix order, hence by calling α the infinite sequence whose prefixes are the $w^{(n)}$ (α is computable by construction), we have $d(\alpha\restriction_n) < c$ for all n. ■

1.4.3 Stochasticity via martingales

We presented computable randomness as an improvement of the notion of Church stochasticity, hence implicitely stating that selection rules were in a sense a particular kind of martingales. We now explain why this is the case. A sequence is non-stochastic if we can extract from it a subsequence that does not satisfy the Law of Large Numbers. As one can easily guess, we will make money by betting on the bits of this selected subsequence. The question is: how to use the information that a sequence does not satisfy the Law of Large Numbers to make money by betting on its bits? Let us first introduce some piece of notation. The BIAS of a sequence $\alpha \in 2^\omega$ is defined by

$$\mathrm{Bias}(\alpha) = \limsup_{n\to+\infty} \left| \frac{\#0(\alpha\restriction_n)}{n} - \frac{1}{2} \right|$$

A sequence α does *not* satisfy the Law of Large Numbers if $\mathrm{Bias}(\alpha) > 0$.

We will also need the following definition:

Definition 1.4.11. *Let $s \in [0,1]$. We say that a martingale s-*SUCCEEDS *against a sequence α if*

$$\limsup_{n\to+\infty} \frac{d(\alpha\restriction_n)}{2^{(1-s)n}} = +\infty$$

A martingale is s-succesful if it succeeds exponentially fast. Moreover, the smaller s is, the faster the success of d. How is this related to stochasticity? Let us start with a simple example. Suppose we are betting (in an infinite game) on the outcomes of a coin that our opponent thinks is balanced, but we happen to know that the coin is in fact biased, namely the probability of each outcome to be 0 is $\frac{1}{2} + \delta$ for some $0 < \delta < \frac{1}{2}$ that we know. There should obviously be a way

to take advantage of that extra information, but what strategy should we use? To maximize our expectancy, it seems like a good idea to bet all the money we have on the value 0. But this is in fact a very bad strategy since if we do this on each coin toss, the outcome 1 will eventually happen and we will lose everything. Let us think more about what we should do. First, the best bet to make should be proportional to the amount we have: if the best move is to bet 0.3 when our capital is 1, it certainly is to bet 3 if our capital is 10 (this is just a change of scale). Also, since the coin tosses are independent, the best decision should be independent of the stage of the game. Hence we should always bet the same fraction ρ of our capital. In that case, after n bets on a sequence α, a simple computation shows that our capital will be

$$(1+\rho)^{\#0(\alpha\restriction_n)}\,(1-\rho)^{\#1(\alpha\restriction_n)}$$

If α is truly picked "at random" where each bit is chosen independently and has probability $(\frac{1}{2}+\delta)$ to be 0, the Law of Large Numbers tells us that for n large enough $\#0(\alpha\restriction_n)$ (resp. $\#1(\alpha\restriction_n)$) will be close to $(\frac{1}{2}+\delta)n$ (resp. to $(\frac{1}{2}-\delta)n$) hence our capital should be close to

$$(1+\rho)^{(\frac{1}{2}+\delta)n}\,(1-\rho)^{(\frac{1}{2}-\delta)n} = \Big[(1+\rho)^{(\frac{1}{2}+\delta)}\,(1-\rho)^{(\frac{1}{2}-\delta)}\Big]^n$$

Lemma 1.4.12. *The maximum of the function $\rho \mapsto (1+\rho)^{(\frac{1}{2}+\delta)}\,(1-\rho)^{(\frac{1}{2}-\delta)}$ (resp. its logarithm $\rho \mapsto (\frac{1}{2}+\delta)\log(1+\rho) + (\frac{1}{2}-\delta)\log(1-\rho)$) is achieved for $\rho = 2\delta$.*

(this is proved by simply taking the derivative).

Hence, if we apply the strategy consisting in betting the fraction 2δ of our current capital at each stage of the game, and still under the hypothesis that each bit of α has a probability $\frac{1}{2}+\delta$ to be 0, by the Law of Large Numbers, we get with probability 1:

$$\begin{aligned} d(\alpha\restriction_n) &= (1+2\delta)^{(\frac{1}{2}+\delta)n+o(n)}\,(1-2\delta)^{(\frac{1}{2}-\delta)n+o(n)} \\ &= 2^{(1-s)n+o(n)} \quad \text{with} \quad s = 1-\Big(\frac{1}{2}+\delta\Big)\log(1+2\delta) - \Big(\frac{1}{2}-\delta\Big)\log(1-2\delta) \end{aligned}$$

The value of s in this formula is related to Shannon's entropy. Recall that the Shannon entropy function is defined, for $x \in [0,1]$ by

$$\mathcal{H}(x) = -x\log(x) - (1-x)\log(1-x)$$

In the above formula, we can write $s = \mathcal{H}(1/2+\delta)$. Note that the function $\delta \mapsto \mathcal{H}(1/2+\delta)$ is a decreasing bijection of $[0,1/2]$ onto $[0,1]$.

The above argument works just as well if we only know the sequence we are about to bet against does not satisfy the Law of Large Numbers:

Proposition 1.4.13. *Let α be a sequence whose bias is greater than or equal to $\delta > 0$. There exists a martingale d, computable with oracle δ, such that for all $s > \mathcal{H}\left(\frac{1}{2}+\delta\right)$, d s-succeeds against α.*

Proof. Let α be a sequence of bias greater or equal to $\delta > 0$. Without loss of generality suppose that α is biased towards 0, that is $\limsup \frac{\#0(\alpha\restriction_n)}{n} \geq (\frac{1}{2}+\delta)$. Let d be the martingale such that at each move bets a fraction 2δ of its current capital on the value 0 (that is, for all $u \in 2^{<\omega}$, $d(u0) = (1+2\delta)d(u)$ and $d(u1) = (1-2\delta)d(u)$). Clearly, the martingale d is computable with oracle δ, and for every n:

$$d(\alpha\restriction_n) = (1+2\delta)^{\#0(\alpha\restriction_n)}(1-2\delta)^{\#1(\alpha\restriction_n)}$$

Thus

$$\frac{\log d(\alpha\restriction_n)}{n} = \frac{\#0(\alpha\restriction_n)}{n}\log(1+2\delta) + \frac{\#1(\alpha\restriction_n)}{n}\log(1-2\delta)$$

and thus

$$\begin{aligned}\limsup_{n\to+\infty} \frac{\log d(\alpha\restriction_n)}{n} &\geq \left(\frac{1}{2}+\delta\right)\log(1+2\delta) - \left(\frac{1}{2}-\delta\right)\log(1-2\delta)\\ &\geq 1 - \mathcal{H}\left(\frac{1}{2}+\delta\right)\end{aligned}$$

This completes the proof. ■

As a first corollary, we get the following unrelativized proposition.

Corollary 1.4.14. *Let α be a sequence of bias $\delta > 0$. For all $s > \mathcal{H}\left(\frac{1}{2}+\delta\right)$, there exists a computable martingale d that s-succeeds against α.*

Proof. Let $s > \mathcal{H}\left(\frac{1}{2}+\delta\right)$ and let s' be rational such that $s > s' > \mathcal{H}\left(\frac{1}{2}+\delta\right)$. Let δ' be such that $s' = \mathcal{H}\left(\frac{1}{2}+\delta'\right)$ (notice that δ' is computable since s' is, $\mathcal{H}$ being a computable function with computable inverse). Applying Proposition 1.4.13 to α and δ', we get the existence of a martingale d', computable with oracle δ' (hence computable since δ' itself is computable) such that d' t-succeeds on α for all $t > s'$. In particular, d' s-succeeds on α. ■

And as a corollary of this corollary, we get:

Corollary 1.4.15. *If α is computably random, it is Church stochastic.*

Proof. Let α be a sequence that is not Church stochastic, that is, there exists a total computable selection rule σ such that $\sigma[\alpha]$ is biased. Then, by the above Corollary 1.4.14, there exists a martingale d that succeeds on $\sigma[\alpha]$. Let d' be the martingale that bets nothing on bits that σ scans, and bets what d bets on bits that σ selects. Precisely, for all $w \in 2^{<\omega}$, if $\sigma(w) = \texttt{scan}$, $d'(w0) = d'(w1) = d'(w)$ and if $\sigma(w) = \texttt{select}$, $d'(w0) = (1+\mathrm{Bet}(d,\sigma[w]))d'(w)$ and $d'(w1) = (1-\mathrm{Bet}(d,\sigma[w]))d'(w)$. Then, d' is computable and

$$\limsup_n d'(\alpha\restriction_n) = \limsup_n d\big(\sigma[\alpha]\restriction_n\big) = +\infty$$

■

Schnorr [52] proved a result in the other direction: if there is a martingale that wins exponentially fast against a sequence α, then there exists a computable selection rule that selects from α a biased sequence β (hence α is not Church stochastic). Schnorr's result is purely qualitative, that is it does not say how the speed of success of the martingale and the bias of the extracted subsequence relate to each other. We will prove the following quantitative version of Schnorr's theorem:

Theorem 1.4.16. *Let d be a martingale that s-succeeds against a sequence $\alpha \in 2^\omega$. Then, there exists a selection rule σ, computable with oracle s, such that* $\mathrm{Bias}(\sigma[\alpha]) \geq \delta$ *where δ is such that $\mathcal{H}\left(\frac{1}{2}+\delta\right) = s$.*

Proof. The basic idea of the proof is the following: by an argument of Ambos-Spies et al. [2], the above theorem would be easier to prove if d always bet the same fraction of its capital on 0 i.e. for all w: $\mathrm{Bet}(d, w) = q$, where q is a fixed constant in $[-1, 1]$. Indeed, in this case, we have for all n:

$$d(\alpha\restriction_n) = (1+q)^{\#0(\alpha\restriction_n)}(1-q)^{\#1(\alpha\restriction_n)}$$

i.e.

$$\frac{\log d(\alpha\restriction_n)}{n} = \frac{\#0(\alpha\restriction_n)}{n}\log(1+q) + \frac{\#1(\alpha\restriction_n)}{n}\log(1-q)$$

Setting $\delta = \mathrm{Bias}(\alpha)$, we conclude that

$$\limsup_{n\to+\infty} \frac{\log d(\alpha\restriction_n)}{n} \leq \left(\frac{1}{2}+\delta\right)\log(1+q) + \left(\frac{1}{2}-\delta\right)\log(1-q)$$

By definition of s:

$$\limsup_{n\to+\infty} \frac{\log d(\alpha\restriction_n)}{n} \geq 1-s$$

It follows that

$$1-s \leq \left(\frac{1}{2}+\delta\right)\log(1+q) + \left(\frac{1}{2}-\delta\right)\log(1-q)$$

The function $x \mapsto \left(\frac{1}{2}+\delta\right)\log(1+x) + \left(\frac{1}{2}-\delta\right)\log(1-x)$ achieving its maximum for $x = 2\delta$, we then have

$$1-s \leq \left(\frac{1}{2}+\delta\right)\log(1+2\delta) + \left(\frac{1}{2}-\delta\right)\log(1-2\delta)$$

i.e.

$$s \geq \mathcal{H}\left(\frac{1}{2}+\delta\right)$$

Thus, the sequence α is biased with bias at least δ. Therefore, it suffices to use the selection rule that selects every bit to get the desired result.

Unfortunately, our martingale d is not restricted as above, hence we cannot directly apply this argument. However, since the values of the bets lie in the

interval $[-1,1]$, which is compact, we argue by a dichotomy technique that there must be some some condensation point $\bar{\rho}$ in the neighbourhood of which bets are often successfull. Applying Ambos-Spies et al.'s technique to this condensation point, we get the desired result.

We start with a slight generalization of Ambos-Spies et al.'s argument:

Lemma 1.4.17. *Suppose d is a martingale that s-succeeds on a sequence $\alpha \in 2^\omega$, and such that for all n, $\mathrm{Bet}(d, \alpha\restriction_n) \in [q_1, q_2]$ where q_1 and q_2 are nonnegative constants. Then, $\mathrm{Bias}(\alpha) \geq \eta$ for any η such that*

$$\left(\frac{1}{2}+\eta\right)\log(1+q_2) + \left(\frac{1}{2}-\eta\right)\log(1-q_1) \leq 1-s$$

Subproof. Let $\eta \in [0,1]$ satisfying the condition of the lemma. Suppose for the sake of contradiction that $\mathrm{Bias}(\alpha) = \eta' < \eta$. When a bit of α is 0, the capital of α is multiplied by at most $(1+q_2)$ (maximal gain, by assumption on the bets) and when a bit of α is 1, the capital of α is multiplied by at most $(1-q_1)$ (minimal loss, by assumption on the bets). Thus

$$d(\alpha\restriction_n) \leq (1+q_2)^{\#0(\alpha\restriction_n)}(1-q_1)^{\#1(\alpha\restriction_n)}$$

which implies

$$\limsup_{n\to+\infty} \frac{\log d(\alpha\restriction_n)}{n} \leq \left(\frac{1}{2}+\eta'\right)\log(1+q_2) + \left(\frac{1}{2}-\eta'\right)\log(1-q_1) \tag{1.2}$$

by definition of the bias.

But since d s-succeeds against α, we derive from (1.2):

$$\begin{aligned} 1-s &\leq \left(\frac{1}{2}+\eta'\right)\log(1+q_2) + \left(\frac{1}{2}-\eta'\right)\log(1-q_1) \\ &< \left(\frac{1}{2}+\eta\right)\log(1+q_2) + \left(\frac{1}{2}-\eta\right)\log(1-q_1) \quad \text{since } \eta' < \eta \end{aligned}$$

and this contradicts the assumption on η.

□

Let us denote by ρ_i the i-th bet made by d while playing against α (i.e. $\rho_i = \mathrm{Bet}(d, \alpha\restriction_i)$) and we let $\tilde{\rho}_i$ be ρ_i if $\alpha_i = 0$ and $\tilde{\rho}_i = -\rho_i$ if $\alpha_i = 1$. With this notation, we have for all n

$$d(\alpha\restriction_n) = \prod_{i=0}^{n-1}(1+\tilde{\rho}_i) \tag{1.3}$$

By definition of s-success, we know that:

$$\limsup_{n\to+\infty}\ d(\alpha\restriction_n)\, 2^{(s-1)n} = +\infty \tag{1.4}$$

From (1.3) and (1.4), we get:

$$\limsup_{n\to+\infty} \prod_{i=0}^{n-1} 2^{(s-1)}(1+\tilde{\rho}_i) = +\infty \tag{1.5}$$

Recall that for all i, $\rho_i \in I_0 = [-1,1]$. Thus, at least one of the following holds:

$$\limsup_{n\to+\infty} \prod_{\substack{0\leq i<n\\ \rho_i\in[-1,0]}} 2^{(s-1)}(1+\tilde{\rho}_i) = +\infty \quad \text{or} \quad \limsup_{n\to+\infty} \prod_{\substack{0\leq i<n\\ \rho_i\in[0,1]}} 2^{(s-1)}(1+\tilde{\rho}_i) = +\infty$$

Indeed, if both quantities above are bounded, then $\prod_{i=0}^{n-1} 2^{(s-1)}(1+\tilde{\rho}_i)$ is necessarily bounded, contradicting (1.5). Suppose for example that the first one is unbounded. We then set $I_1 = [-1,0]$, and we see that, again, at least one of the following holds:

$$\limsup_{n\to+\infty} \prod_{\substack{0\leq i<n\\ \rho_i\in[-1,-1/2]}} 2^{(s-1)}(1+\tilde{\rho}_i) = +\infty \quad \text{or} \quad \limsup_{n\to+\infty} \prod_{\substack{0\leq i<n\\ \rho_i\in[-1/2,0]}} 2^{(s-1)}(1+\tilde{\rho}_i) = +\infty$$

We repeat this argument infinitely many times, by induction, dividing each time the interval I_m into two halves, and choosing I_{m+1} to be one of these halves that satisfies

$$\limsup_{n\to+\infty} \prod_{\substack{0\leq i<n\\ \rho_i\in I_{m+1}}} 2^{(s-1)}(1+\tilde{\rho}_i) = +\infty$$

The sequence of intervals $(I_m)_{m\in\mathbb{N}}$ is decreasing and for all m, the length of I_m is 2^{-m+1}. Hence, by compactness, their intersection is a singleton $\{\bar{\rho}\}$.

We distinguish two cases:

Case 1: $\bar{\rho} = 2\delta$. In this case, since we can compute δ with oracle s, we can also compute $\bar{\rho}$, and also all the sequence intervals $(I_m)_{m\in\mathbb{N}}$ (since we know $\bar{\rho}$, we know for each m in which half of I_m it is located).

Let σ be the selection rule that proceeds by stages. At the beginning of stage N, we compute m such that $I_m = [a_m, b_m]$ (by symmetry, we can suppose without loss of generality that $0 \leq a_m \leq b_m$) satisfies

$$\left(\frac{1}{2}+\delta-\frac{1}{N}\right)\log(1+b_m) - \left(\frac{1}{2}-\delta+\frac{1}{N}\right)\log(1-a_m) \leq 1-s \tag{1.6}$$

Such an m can be found since

$$\begin{aligned}
&\lim_{n\to+\infty} \left(\tfrac{1}{2}+\delta-\tfrac{1}{N}\right)\log(1+b_m) - \left(\tfrac{1}{2}-\delta+\tfrac{1}{N}\right)\log(1-a_m)\\
&= \left(\tfrac{1}{2}+\delta-\tfrac{1}{N}\right)\log(1+2\delta) - \left(\tfrac{1}{2}-\delta+\tfrac{1}{N}\right)\log(1-2\delta)\\
&< \left(\tfrac{1}{2}+\delta\right)\log(1+2\delta) - \left(\tfrac{1}{2}-\delta\right)\log(1-2\delta) = 1-s
\end{aligned}$$

Knowing this m, σ selects during stage N all the bits α_i for which $\rho_i \in I_m = [a_m, b_m]$. By definition of a_m, b_m, knowing that d s-succeeds on the selected bits, we can apply Lemma 1.4.17 (with $\eta = \delta - 1/N$, $q_1 = a_m$ and $q_2 = b_m$), which asserts that if we stayed in stage N forever, the bias would be at least $\delta - 1/N$. Hence, at some point, we will have selected a finite sequence with a proportion of at least $\frac{1}{2} + \delta - 2/N$. As soon as this happens, we move on to stage $N+1$. Hence, at each stage, we will get closer to having bias δ, and we achieve it in the limit.

Case 2: $\bar{\rho} \neq 2\delta$. In this case, let m be such that $I_m = [a_m, b_m]$ satisifes

$$\left(\frac{1}{2} + \delta\right)\log(1 + b_m) - \left(\frac{1}{2} - \delta\right)\log(1 - a_m) \leq 1 - s$$

Such an m can be found since

$$\begin{aligned}
&\lim_{n\to+\infty} \left(\tfrac{1}{2} + \delta\right)\log(1 + b_m) - \left(\tfrac{1}{2} - \delta\right)\log(1 - a_m) \\
&= \left(\tfrac{1}{2} + \delta\right)\log(1 + \bar{\rho}) - \left(\tfrac{1}{2} - \delta\right)\log(1 - \bar{\rho}) \\
&< 1 - \mathcal{H}\left(\tfrac{1}{2} + \delta\right) \quad \text{by Lemma 1.4.12}
\end{aligned}$$

Let σ be the selection rule that selects a bit α_i whenever $\mathrm{Bet}(d, \alpha\restriction_n) \in I_m$. We can apply Lemma 1.4.17 with $\eta = \delta$, $q_1 = a_m$ and $q_2 = b_m$, remembering that d s-succeeds when restricted to the bits on which its bet belongs to I_m (this by definition of I_m), and the desired result follows.

■

As we did with Propostion 1.4.13, we can state an unrelativized version of the theorem we just proved:

Corollary 1.4.18. *If there exists a computable martingale d that s-succeeds against a sequence $\alpha \in 2^\omega$ then for all δ' such that $\mathcal{H}\left(\frac{1}{2} + \delta'\right) > s$ there exists a computable selection rule σ, computable with oracle s, such that* $\mathrm{Bias}(\sigma[\alpha]) \geq \delta'$.

Proof. It suffices to proves this for δ' computable. If δ' is computable, let s' be such that $\mathcal{H}\left(\frac{1}{2} + \delta'\right) = s'$. One has $s' > s$, and s' is computable since δ' is. Since d s-succeeds, it s'-succeeds and then one can apply Theorem 1.4.16 to (δ', s') to get the result. ■

The careful reader will notice that we claimed a stronger result in the introduction: we said that a sequence is Church stochastic if and only if there is no computable martingale which succeeds exponentially fast in the number of non-zero bets. In Theorem 1.4.16 and its corollaries, we used the stronger assumption of s-success (meaning that the martingale succeeds exponentially simply in the number of bets). Let us now prove the full result. Let $\alpha \in 2^\omega$ and d a computable martingale which s-succeeds "in the number of non-zero bets", meaning that:

$$\limsup_{n\to+\infty} \prod_{\substack{0\leq i<n \\ \rho_i\in[-1,0)\cup(0,1]}} 2^{(s-1)}(1 + \tilde{\rho}_i) = +\infty$$

(using the same notation as before). Then,

$$\limsup_{n\to+\infty} \prod_{\substack{0\leq i<n\\ \rho_i\in[-1,0)}} 2^{(s-1)}(1+\tilde{\rho}_i) = +\infty \quad \text{or} \quad \limsup_{n\to+\infty} \prod_{\substack{0\leq i<n\\ \rho_i\in(0,1]}} 2^{(s-1)}(1+\tilde{\rho}_i) = +\infty$$

And we proceed by dichotomy like in the proof of Theorem 1.4.16, ensuring that we never select a bit on which d makes a zero bet. The rest of the proof is identical.

1.5 Typicalness vs unpredictability

So far, we have discussed two philosophical approaches to randomness: typicalness and unpredictability. The obvious question is: how are these two approaches related? Is it atypical to be predictable? Conversely, to what extent can a player use atypicalness to make accurate predictions? This is what we discuss in this section.

1.5.1 When typicalness implies unpredictability

As one might expect, it is atypical for a sequence $\alpha \in 2^\omega$ to be predictable: the probability to make large amounts of money by betting on a sequence is very small; a fact that is quantified by the fundamental *Ville's inequality* (also known as Kolmogorov's inequality): for any fixed $r > 0$, the player has at most a probability $1/r$ to ever multiply his initial capital by r.

Theorem 1.5.1 (Ville [58])**.** *Let d be a martingale. For all $r > 0$, one has:*

$$\lambda\big\{\alpha \in 2^\omega : \exists n\ d(\alpha\restriction_n) \geq r\, d(\epsilon)\big\} \leq 1/r$$

To prove this theorem, we first show:

Lemma 1.5.2. *Let d be a martingale and let A be a prefix-free set of strings. Then:*

$$\sum_{w\in A} 2^{-|w|} d(w) \leq d(\epsilon)$$

Subproof. Since $\sum_{w\in A} 2^{-|w|}d(w)$ is the supremum of $\sum_{w\in A'} 2^{-|w|}d(w)$ over finite subsets $A' \subseteq A$, it suffices to prove this for A finite. We proceed by induction on $l = \max\{|w| : w \in A\}$. For $l = 0$ this is trivial. Assume this holds for l and let A be a finite prefix-free set of strings whose length is bounded by $l+1$. Let us set $A_0 = \{v \in 2^{<\omega} : 0v \in A\}$ and $A_1 = \{v \in 2^{<\omega} : 1v \in A\}$. Define also the martingales d_0 and d_1 by $d_0(v) = d(0v)$ and $d_1(v) = d(1v)$ for all $v \in 2^{<\omega}$ (it is easy to check that these are indeed martingales). Now:

$$\begin{aligned}\sum_{w\in A} 2^{-|w|}d(w) &= \sum_{v\in A_0} 2^{-|0v|}d(0v) + \sum_{v\in A_1} 2^{-|1v|}d(1v)\\ &\leq \frac{1}{2}\sum_{v\in A_0} 2^{-|v|}d_0(v) + \frac{1}{2}\sum_{v\in A_1} 2^{-|v|}d_1(v)\end{aligned}$$

Since the maximal length of strings in A_0 (resp. A_1) is at most l by definition, we can apply this induction hypothesis, and we get:

$$\sum_{w\in A} 2^{-|w|}d(w) \leq \frac{1}{2}d_0(\epsilon) + \frac{1}{2}d_1(\epsilon) = \frac{1}{2}d(0) + \frac{1}{2}d(1) = d(\epsilon)$$

□

Proof (of Theorem 1.5.1). Let $r > 0$ and

$$\mathcal{U} = \{\alpha \in 2^\omega : \exists n\ d(\alpha\restriction_n) \geq r\, d(\epsilon)\}$$

It is clear from its definition that $\mathcal{U}$ is an open set. We now let A be the set of minimal words w having the property $d(w) \geq r\, d(\epsilon)$. A is prefix-free and $\mathcal{U} = [A]$. Thus:

$$\lambda(\mathcal{U}) = \sum_{w\in A} 2^{-|w|} \leq \frac{1}{r\, d(\epsilon)} \sum_{w\in A} 2^{-|w|}d(w)$$

(the inequality holds since by definition of A, $d(w) \geq r\, d(\epsilon)$ for all $w \in A$). Applying Lemma 1.5.2, we get that $\sum_{w\in A} 2^{-|w|}d(w) \leq d(\epsilon)$ which, together with the above inequality, implies the desired result. ■

From Ville's inequality, we immediately get the following theorem, which illustrates how (in some sense) typicalness implies unpredictability.

Theorem 1.5.3. *For all martingales d, $\lambda\big(\mathrm{Succ}(d)\big) = 0$.*

Proof. This is because $\mathrm{Succ}(d) = \bigcap_{n\in\mathbb{N}} \mathcal{U}_k$ where

$$\mathcal{U}_k = \{\alpha \in 2^\omega : \exists n\ d(\alpha\restriction_n) \geq k\, d(\epsilon)\}$$

and we know from Ville's theorem that for all k, $\lambda(\mathcal{U}_k) \leq 1/k$. ■

Since there are only countably many computable martingales, the union of their success sets has measure 0. Hence, the complement of this union, which by definition is the set of computably random sequences, has measure 1. One can effectivize this by proving that for every computable martingale, the set $\mathrm{Succ}(d)$ is in fact a Martin-Löf nullset (which implies **MLR** $\subseteq$ **CR**). In fact, one can prove this for a larger class of martingales:

Theorem 1.5.4. *Let d be a left-c.e. martingale. Then $\mathrm{Succ}(d)$ is a Martin-Löf nullset.*

Proof. Let d be a left-c.e. martingale. Without loss of generality, we can assume that the initial capital $d(\epsilon)$ is smaller than 1 (up to dividing d by a large enough integer, which does not change its left-c.e. property). For all k, the set

$$\mathcal{V}_k = \{\alpha \in 2^\omega : \exists n\ d(\alpha\restriction_n) > 2^k\}$$

is a c.e. open set, uniformly in k. This is because the set

$$D = \{(w, q) : \ w \in 2^{<\omega} \wedge q \in \mathbb{Q} \wedge d(w) > q\}$$

is c.e. Since $d(\epsilon) \leq 1$, it follows from Ville's theorem that for all k, $\lambda(\mathcal{V}_k) \leq 2^{-k}$. Hence $(\mathcal{V}_k)_{k\in\mathbb{N}}$ is a Martin-Löf test, and $\text{Succ}(d) \subseteq \bigcap_k \mathcal{V}_k$. ■

Like we said, this immediately implies:

Corollary 1.5.5. *If $\alpha \in 2^\omega$ is Martin-Löf random, it is computably random.*

Proof. Let α be a sequence that is not computably random. There exists a martingale d such that $\alpha \in \text{Succ}(d)$ which is a Martin-Löf nullset by the above theorem. ■

1.5.2 When unpredictability implies typicalness

We now investigate the relation between unpredictability and typicalness in the reverse direction. The central question is: how can we turn the information that a sequence α belongs to a set of small measure into a martingale that will win money on α? First, we will not care about effectivity, and we will show that unpredictability implies typicalness in the sense that any nullset can be covered by the success set of some martingale:

Theorem 1.5.6 (Ville [58])**.** *For every $\mathcal{X} \subseteq 2^\omega$ such that $\lambda(\mathcal{X}) = 0$, there exists a martingale d such that $\mathcal{X} \subseteq \text{Succ}(d)$.*

In order to prove this, we first show how to make money if we know that the sequence we are playing against belongs to an open set of small mesure.

Proposition 1.5.7. *Let $\mathcal{U}$ be an open subset of 2^ω. There exists a martingale $d_\mathcal{U}$ with initial capital $d(\epsilon) = \lambda(\mathcal{U})$ such that for every $\alpha \in \mathcal{U}$, the sequence $(d(\alpha\restriction_n))_{n\in\mathbb{N}}$ becomes eventually equal to 1.*

Hence, the smaller $\lambda(\mathcal{U})$ is, the more profit (relatively to its initial capital) d makes.

Proof. First suppose that $\mathcal{U}$ is a cylinder, i.e. $\mathcal{U} = [u]$ for some $u \in 2^{<\omega}$. Let d_u be the martingale that starts with a capital $\lambda(\mathcal{U}) = 2^{-|u|}$ and at step k, for all k, bets all its money on the value u_k (and stops betting for $k \geq |u|$). Formally:

$$d_u(w) = \begin{cases} 2^{-|u|+|w|} \text{ if } w \sqsubseteq u \\ 1 \text{ if } u \sqsubseteq w \\ 0 \text{ otherwise} \end{cases}$$

It is easy to check that d_u works for $\mathcal{U} = [u]$. Now, in the general case, $\mathcal{U}$ is a union of cylinders, i.e. can be written as $\mathcal{U} = [A]$ where A is a prefix-free set of strings (hence $\lambda(\mathcal{U}) = \sum_{u\in A} 2^{-|u|}$). Let $d_\mathcal{U}$ be the martingale defined by

$$d_\mathcal{U} = \sum_{u\in A} d_u$$

with d_u defined in the particular case above. We have $d_{\mathcal{U}}(\epsilon) = \sum_{u\in A} 2^{-|u|} = \lambda(\mathcal{U})$ and d is a martingale as a weighted sum of martingales. Moreover, if $\alpha \in \mathcal{U}$, then α as a prefix v in A, hence for all $n \geq |v|$, $d_{\mathcal{U}}(\alpha\upharpoonright_n) = d_v(\alpha\upharpoonright_n) = 1$ by definition of d_v. This completes the proof. ■

In the sequel, we will keep using the notations introduced in this proof: d_u the doubling strategy for a word u with initial capital $2^{-|u|}$ and $d_{\mathcal{U}}$ the martingale whose existence is asserted by Proposition 1.5.7. Moreover, for a set of words D, we sometimes abbreviate $d_{[D]}$ by d_D.

We can now prove Theorem 1.5.6:

Proof (of Theorem 1.5.6). Let $\mathcal{X} \subseteq 2^\omega$ be a nullset. By definition of Lebesgue measure, for all n, there exists an open set $\mathcal{U}_n$ of measure at most 2^{-n} containing $\mathcal{X}$. Consider the martingale

$$d = \sum_{n\in\mathbb{N}} 2^n\, d_{\mathcal{U}_{2n}}$$

This is a martingale as a weighted sum of martingales with initial capital

$$d(\epsilon) = \sum_{n\in\mathbb{N}} 2^n\, \lambda(\mathcal{U}_{2n}) \leq \sum_{n\in\mathbb{N}} 2^{-n} \leq 2$$

Now, for all $\alpha \in \mathcal{X}$, α belongs to all $\mathcal{U}_{2n}$, hence $\sup_k d_{\mathcal{U}_{2n}}(\alpha\upharpoonright_k) = 1$, hence $\sup_k d(\alpha\upharpoonright_k) \geq 2^n$ for all n. This implies $\alpha \in \mathrm{Succ}(d)$. ■

Martin-Löf randomness via martingales

We have already seen with Theorem 1.5.4 that the success set of a left-c.e. martingale is a Martin-Löf nullset. We now prove the converse of this result hence proving:

Theorem 1.5.8. *A sequence α is* not *Martin-Löf if and only if some left-c.e. martingale succeeds on it.*

Proof. By Theorem 1.5.4, it only remains to prove that if α is not Martin-Löf random, some left-c.e. martingale succeeds on it. This is in fact simply the effective version of the proof of Theorem 1.5.6. If $(\mathcal{U}_n)_{n\in\mathbb{N}}$ is a Martin-Löf test covering α, set

$$d = \sum_{n\in\mathbb{N}} 2^n\, d_{\mathcal{U}_{2n}}$$

It only remains to show that d is left-c.e. This is because for a string u the martingale d_u is computable uniformly in u, hence for an effectively open set $\mathcal{U}$ generated by a prefix-free set of strings A, $d_{\mathcal{U}} = \sum_{u\in A} d_u$ is left-c.e. as it is the limit over t of $\sum_{u\in A[t]} d_u$. Since the $\mathcal{U}_n$ are uniformly c.e., the martingale d above is left-c.e. The rest of the proof is identical to the proof of Theorem 1.5.4. ■

Schnorr randomness via martingales

In order to give a characterization of Schnorr randomness in terms of computable martingale, we first prove the following characterization *à la* Borel-Cantelli.

Lemma 1.5.9. *The following are equivalent for every $\alpha \in 2^\omega$:*
(a) α is not Schnorr random.
(b) There exists a uniformly computable sequence $(D_n)_{n\in\mathbb{N}}$ of finite sets of strings such that $\lambda([D_n]) \leq 2^{-n}$ for all n, and such that α belongs to infinitely many D_n.

Proof. $(a) \Rightarrow (b)$. If α is not Schnorr random, then there exists by definition a computable sequence of c.e. prefix-free sets $(A_n)_{n\in\mathbb{N}}$ of strings such that $\lambda(A_n) = 2^{-n}$ for all n. Given n, we partition A_n into finite sets $(A_n^i)_{i\in\mathbb{N}}$ in such a way that for all i, $\lambda([A_n^i]) \leq 2^{-n-2i}$. This can be done as follows: in a standard enumeration of A_n, find stages $t_0 < t_1 < t_2 < ...$ such that $A_n \setminus A_n[t_i]$ (A_n minus its enumeration up to stage t_i) has measure at most $2^{-n-2i-1}$. Set $A_n^0 = A_n[t_0]$ and inductively $A_n^i = A_n[t_i] \setminus (A_n^0 \cup ... \cup A_n^{i-1})$. The A_n^i are as wanted. Moreover, the partition can be done effectively (and uniformly in n) since we know precisely the measure of A_n. Now, set $D_n = \bigcup_{i+j=n+1} A_i^j$. Clearly α is in infinitely many D_n, and for all n:

$$\lambda([D_n]) \leq \sum_{i=0}^{n+1} \lambda([A_i^{n+1-i}]) \leq \sum_{i=0}^{n+1} 2^{-i-2(n+1-i)} \leq 2^{-n}$$

$(b) \Rightarrow (a)$. If α is in infinitely many such D_n, then α belongs to

$$\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} [D_n]$$

For all N, $A_N = \bigcup_{n\geq N} D_n$ is a c.e. subset of $2^{<\omega}$ (uniformly in N), hence $[A_n]$ is a c.e. open set. Moreover, $\lambda([A_N])$ is computable (uniformly in n) as the measure of the $[D_n]$ is computable and exponentially decreasing. Finally, $\lambda([A_N])$ tends to 0 as n tends to infinity, since $\lambda([A_N]) \leq \sum_{n\geq N} 2^{-n}$. Taking $\mathcal{V}_n = [A_n]$, for all n, we get a Schnorr test that covers α, hence α is not Schnorr random. □

Like we said above, if a sequence α is covered by a Schnorr test, we can use this Schnorr test to predict the bits of α, i.e. α is not computably random. We can prove a little more than that: if a sequence is not Schnorr random, then there exists a martingale that succeeds on it with computable speed.

Theorem 1.5.10 (Schnorr [52])**.** *The following are equivalent for every $\alpha \in 2^\omega$:*
(a) α is not Schnorr random.
(b) There exists a computable martingale d and a computable order h such that $d(\alpha\restriction_n) \geq h(n)$ for infinitely many n.

Proof. $(a) \Rightarrow (b)$. If α is not random, by Lemma 1.5.9, there exists a computable sequence of finite sets of strings $(D_n)_{n\in\mathbb{N}}$ such that $\lambda([D_n]) \leq 2^{-n}$ for all n and α is in infinitely many $[D_n]$. Up to replacing some elements of D_n by all their

extension, we can assume that for all n, all elements of D_n have the same length $f(n)$, with f a computable function which we can assume to be increasing. Let d be the martingale defined by

$$d(w) = \sum_{n\in\mathbb{N}} n\, d_{D_n}(w)$$

Since $\lambda([D_n]) \leq 2^{-n}$, we have $d_{D_n}(w) \leq 2^{-n+|w|}$ for all w, hence d is indeed a martingale and is computable. Moreover, if n is such that $\alpha\restriction_{f(n)} \in D_n$, then by definition of d_{D_n}: $d_{D_n}(\alpha\restriction_{f(n)}) = 1$ and thus $d(\alpha\restriction_{f(n)}) \geq n\, d_{D_n}(\alpha\restriction_{f(n)}) \geq n$. This happen for infinitely many n. And if we set $k = f^{-1}(n)$, we have $d(\alpha\restriction_k) \geq f^{-1}(k)$ for infinitely many k. Since f^{-1} is an order, we are done.
$(b) \Rightarrow (a)$. Suppose there exists a computable martingale d and a computable order h such that $d(\alpha\restriction_n) \geq h(n)$ for infinitely many n. For all k, set $I_k = \{n \in \mathbb{N} : h(n) \in [2^k, 2^{k+1})\}$ (clearly the I_k form a partition of $\mathbb{N}$) and set $D_k = \{w : |w| \in I_k \wedge d(w) \geq h(|w|)\}$. Since for $w \in I_k$, $h(|w|) \geq 2^k$, one can apply Ville's theorem (Theorem 1.5.1) to get $\lambda([D_k]) \leq 2^{-k}$. The D_k are finite and uniformly computable, and α belongs to infinitely many D_k by definition. The result follows from Lemma 1.5.9. ■

If $d(\alpha\restriction_n) \geq h(n)$ for some computable order h and infinitely many n, then in particular d succeeds against α. Hence:

Corollary 1.5.11. *If $\alpha \in 2^\omega$ is computably random, it is Schnorr random.*

It seems from Theorem 1.5.10 that Schnorr randomness should be weaker than computable randomness. This fact however is not immediate, and was actually an open question for some time (see Lutz [39]). It was finally proved by Wang [62]; in Chapter 2, we will use Kolmogorov complexity to separate these concepts.

Weak randomness via martingales

Weak randomness admits a characterization in terms of martingales that is the dual of the one for Schnorr randomness:

Theorem 1.5.12 (Wang [61])**.** *The following are equivalent for every $\alpha \in 2^\omega$:*
(a) α is not weakly random.
(b) There exists a computable martingale d and a computable order h such that $d(\alpha\restriction_n) \geq h(n)$ for all n.

In order to prove this theorem, let us prove the analogue of Lemma 1.5.9:

Lemma 1.5.13. *The following are equivalent for every $\alpha \in 2^\omega$:*
(a) α is not weakly random.
(b) There exists a uniformly computable sequence $(D_n)_{n\in\mathbb{N}}$ of finite sets of strings such that $\lambda([D_n]) \leq 2^{-n}$ for all n, and such that α belongs to all D_n.

Proof. $(a) \Rightarrow (b)$. Suppose α is not weakly random. There exists a c.e. open set $\mathcal{U}$ of measure 1 such that $\alpha \notin \mathcal{U}$. Let A be a c.e. prefix-free set such that $\mathcal{U} = [A]$. For all i, let t_i be the first stage in the enumeration of A such that $\lambda\Big(\big[A[t_i]\big]\Big) \geq 1 - 2^{-i}$. The t_i can be found effectively. For all i, the set $2^\omega \setminus \big[A[t_i]\big]$ is clopen hence can be written as a finite union of cylinders $[D_i]$ with D_i a finite subset of $2^{<\omega}$. D_i can be computed from $A[t_i]$ hence can be computed effectively. Moreover, for all i, $\lambda\Big(\big[D_i\big]\Big) \leq 2^{-i}$ and $\alpha \in [D_i]$.

$(b) \Rightarrow (a)$. The $[D_n]$ form a uniform sequence of clopen sets. Their intersection is thus an effectively closed set of measure 0. Since α belongs to this intersection, by definition of weak randomness, α is not weakly random. □

The rest of the proof is identical to the proof of Theorem 1.5.10 using Lemma 1.5.13 instead of Lemma 1.5.9 and replacing all the "for infinitely many n" by "for all n".

Corollary 1.5.14. *If $\alpha \in 2^\omega$ is Schnorr random, it is weakly random.*

Proof. This immediately follows from Theorem 1.5.10 and Theorem 1.5.12. ■

It seems rather clear from the respective characterizations of Schnorr randomness and weak randomness by martingales that the second one should be stricly weaker than the first. We will show this in Section 1.8.

Effective Hausdorff dimension via martingales

It turns out that Hausdorff dimension – both effective and non-effective – on the Cantor space also admits a very natural characterization in terms of martingales. Let us first state it for the non-effective version:

Theorem 1.5.15 (Lutz [40]). *For every $\mathcal{X} \subseteq 2^\omega$:*

$$\dim(\mathcal{X}) = \inf\{s : \exists d \text{ martingale which } s\text{-succeeds on all } \alpha \in \mathcal{X}\}$$

Proof. Let d be a martingale – which we can assume to be normed – that s-succeeds on all $\alpha \in \mathcal{X}$, and let us show that $\dim(\mathcal{X}) \leq s$. Let $s' > s$. For all n, set

$$A_n = \big\{w \in 2^{<\omega} \text{ minimal s.t. } d(w) \geq 2^{(1-s')|w|+n}\big\}$$

By definition of s-success $\mathcal{X} \subseteq [A_n]$ for all n. Moreover, since A_n is by definition prefix-free, we can apply Lemma 1.5.2 and we get

$$\sum_{w \in A_n} 2^{-|w|} d(w) \leq d(\epsilon) = 1$$

Hence, by definition of A_n:

$$\sum_{w \in A_n} 2^{-|w|} 2^{(1-s')|w|+n} \leq 1 \quad \text{i.e.} \quad \sum_{w \in A_n} 2^{-s'|w|} \leq 2^{-n}$$

which by Proposition 1.3.8 implies that $H^{s'}(\mathcal{X}) = 0$. We have proven this for all $s' > s$, hence $\dim(\mathcal{X}) \leq s$.

Conversely, suppose that $\dim(\mathcal{X}) \leq s$. We need to show that for all $s' > s$, some martingale d s'-succeeds on all elements of $\mathcal{X}$. Let $s' > s$. We have $H^{s'}(\mathcal{X}) = 0$, hence for all n, by Proposition 1.3.8, there exists a prefix-free set of strings A_n such that $\mathcal{X} \subseteq [A_n]$ and $\sum_{w \in A_n} 2^{-s'|w|} \leq 2^{-n}$. Set

$$d = \sum_{n \in \mathbb{N}} \sum_{w \in A_n} 2^{(1-s')|w|} d_w$$

d is a weighted sum of martingales and

$$d(\epsilon) = \sum_{n \in \mathbb{N}} \sum_{w \in A_n} 2^{(1-s')|w|} 2^{-|w|} \leq \sum_{n \in \mathbb{N}} \sum_{w \in A_n} 2^{-s'|w|} \leq \sum_{n \in \mathbb{N}} 2^{-n} \leq 2$$

Thus d is a martingale, and for all $\alpha \in \mathcal{X}$, for all n, α has a prefix v in A_n. For that prefix, $d(v) \geq 2^{(1-s')|v|+n} d_v(v) = 2^{(1-s')|v|+n}$. Hence, for all n, for infinitely many k, $d(\alpha\restriction_k) \geq 2^{(1-s')k+n}$. This means that d s'-succeeds against α. ■

This theorem can be effectivized both for constructive and computable dimension.

Theorem 1.5.16 (Lutz [40]). *For every $\mathcal{X} \subseteq 2^\omega$:*

$$\mathrm{cdim}(\mathcal{X}) = \inf \{s : \exists d \text{ left-c.e. martingale which } s\text{-succeeds on all } \alpha \in \mathcal{X}\}$$

$$\dim_{\mathrm{comp}}(\mathcal{X}) = \inf \{s : \exists d \text{ comp. martingale which } s\text{-succeeds on all } \alpha \in \mathcal{X}\}$$

(the first part of this theorem is due to Lutz [41], where the author actually uses the characterization by left-c.e. martingales as a definition).

Proof. The proof is almost the same as for Theorem 1.5.15. Just notice that, for s' rational, if the A_n are uniformly c.e then d is left-c.e. The case of computable dimension requires a little more work. In the above proof, if s' is rational, and if the familiy A_n is uniformly computable, take any $s'' > s'$. Let then d' be the martingale defined by

$$d' = \sum_{n \in \mathbb{N}} \sum_{w \in A_n} 2^{(1-s'')|w|} d_w$$

d' is a martingale that s''-succeeds on all $\alpha \in \mathcal{X}$ for the same reason as above. It remains to show that d' is computable. Let u be a string and $k > 0$. We approximate $d'(u)$ by

$$\sum_{n \leq k} \sum_{\substack{w \in A_n \\ |w| \leq k}} 2^{(1-s'')|w|} d_w(u)$$

(which is computable uniformly in (u, k) as a finite sum of uniformly computable terms). It remains to estimate the difference between $d'(u)$ and this approximation.

First, truncating the fisrt sum results in an error of at most $2^{-k+|u|}$ since for all n, $\sum_{w\in A_n} 2^{(1-s'')|w|}d_w(u) \leq 2^{-n+|u|}$. For the second sum, note that

$$\begin{aligned}\sum_{\substack{w\in A_n\\ |w|>k}} 2^{(1-s'')|w|}d_w(u) &\leq \sum_{\substack{w\in A_n\\ |w|>k}} 2^{(-s'')|w|+|u|}\\ &\leq 2^{-(s''-s')k+|u|}\sum_{w\in A_n} 2^{-s'|w|}\\ &\leq 2^{-(s''-s')k+|u|-n}\end{aligned}$$

Hence the total error we are making is of order $O(2^{-(s''-s')k})$ which computably tends to 0. And since we can construct such a martingale d' for s', s'' arbitrarily close to s, we get the desired result.

■

Corollary 1.5.17. *For every Schnorr random sequence α, $\dim_{\text{comp}}(\alpha) = 1$.*

Proof. Suppose $\dim_{\text{comp}}(\alpha) < 1$. By Theorem 1.5.16, there exists a rational number $s < 1$ and a computable martingale d such that $d(\alpha\restriction_n) \geq 2^{(1-s)n}$ for infinitely many n. And since $n \mapsto 2^{(1-s)n}$ is clearly a computable order, by Theorem 1.5.10, α is not Schnorr random. ■

Recalling our previous discussion on how the Law of Large Numbers, Church stochasticity and computable martingales relate, we get from the above theorem:

Proposition 1.5.18. *(i) Let $\alpha \in 2^\omega$ such that $\text{Bias}(\alpha) \geq \delta > 0$. We have $\dim_{\text{comp}}(\alpha) \leq \mathcal{H}\left(\frac{1}{2}+\delta\right) < 1$.*
(ii) Let $\alpha \in 2^\omega$. α is Church stochastic if and only if for every infinite subsequence β extracted from α by a computable selection rule, $\dim_{\text{comp}}(\beta) = 1$ (in particular, if α is Church stochastic, then $\dim_{\text{comp}}(\alpha) = 1$).

Proof. (i) Suppose $\text{Bias}(\alpha) \geq \delta$. By Corollary 1.4.14, for all $s > \mathcal{H}\left(\frac{1}{2}+\delta\right)$, there exists a computable martingale that s-succeeds against α, hence $\dim_{\text{comp}}(\alpha) \leq s$. Since s can be taken arbitrarily close to $\mathcal{H}\left(\frac{1}{2}+\delta\right)$, we get the result.

(ii) First assume that α is not Church stochastic. By definition, there exists a computable selection rule that extracts from α an infinite sequence β with $\delta = \text{Bias}(\beta) > 0$. By the above part (a), this implies $\dim_{\text{comp}}(\beta) \leq \mathcal{H}\left(\frac{1}{2}+\delta\right) < 1$.

Conversely, suppose that some computable selection rule σ extracts from α an infinite sequence $\beta = \sigma[\alpha]$ such that $s = \dim_{\text{comp}}(\beta) < 1$. Take $s < s' < 1$ computable. By Corollary 1.4.18, there exists a computable selection rule σ' such that $\text{Bias}(\sigma'[\beta]) \geq \delta'$ where δ' is such that $\mathcal{H}\left(\frac{1}{2}+\delta'\right) = s'$. We claim that the composition of two computable selection rules is itself a computable selection rule. Indeed, given two computable selection rules σ and σ', define σ'' by

$$\sigma''(u) = \begin{cases} \texttt{select} \text{ if } \sigma(u) = \texttt{select} \text{ and } \sigma'(\sigma[u]) = \texttt{select}\\ \texttt{scan} \text{ otherwise}\end{cases}$$

then σ'' is computable, and $\sigma''[\alpha] = \sigma'\big[\sigma[\alpha]\big]$. Here, if we set $\sigma'' = \sigma' \circ \sigma$, we have $\mathrm{Bias}(\sigma''[\alpha]) = \mathrm{Bias}(\sigma'[\beta]) \geq \delta' > 0$, hence α is not Church random. ■

1.6 Schnorr randomness and normal numbers

In this section, we present an interesting application of the notions and theorems introduced in this chapter. We will see how Schnorr randomness can be applied to prove the following result: there exists a computable absolutely normal number[1]. A number $x \in [0,1]$ is said to be NORMAL IN BASE b if in its decimal representation in base b, every word w over the alphabet $\{0,1,\ldots,b-1\}$ appears with limit frequency $b^{-|w|}$. We say that x is ABSOLUTELY NORMAL if it is normal in base b for all integer $b \geq 2$. By the Law of Large Numbers, we know that the set of absolutely normal numbers has measure 1. It is also easy to construct in a computable way a real x that is normal in a given base. However, proving the existence of a *computable absolutely normal number* is non-trivial. This was first achieved by Becher and Figueira [5], although it seems that Turing [56] had given an almost complete proof of that result (see Becher et al. [6] for an account of the history of this problem). In order to avoid heavy notation and tedious details, we will only give a high-level argument, decomposing it into several steps.

Step 1. Suppose x is not normal in some base b, i.e. some word w (denote its length by k) appears in the expansion x_b of x in base b with a 'limsup frequency" greater than b^{-k}. Grouping the digits of x in base b by blocks of size k, this is equivalent to say that some digit in the expansion of x in base $b' = b^k$ appears with "limsup frequency" greater than $1/b'$.

Step 2. We now have the expansion of $x_{b'}$ of x in base b', which is an infinite sequence of digits in $\Sigma = \{0,\ldots,b'-1\}$, with a digit that infinitely often appears significatively more than it should. We now use Proposition 1.4.13 to design a strategy that makes money on this type of sequence. Let us just extend the notion of martingale to Σ^ω as a function $d : \Sigma^\omega \to \mathbb{R}_+$ satisfying $d(w)(\#\Sigma) = \sum_{\sigma\in\Sigma} d(w\sigma)$ for all strings w over the alphabet Σ. The method used in the proof of Proposition 1.4.13 can be adapted as follows: if a symbol $\tau \in \Sigma$ appears with a "limsup frequency" at least $1/b' + \delta$ for some rational $\delta > 0$, we use the strategy that at each move bets a fraction $\left(\frac{b-1}{b}\right)\delta$ of its capital on the value τ. A simple calculation shows that this strategy succeeds exponentially fast.

Step 3. Recall the characterization of Schnorr randomness in terms of martingales (Theorem 1.5.10): a sequence is not Schnorr random if some computable martingale succeeds on it with some fixed computable speed (represented by a computable order). The exponential functions are in particular orders, hence the martingale we constructed in Step 2 asserts that the expansion of x in base b' is not

[1] at the time of writing this thesis, the proof of this theorem which I present here could not be found in the literature, but some people I discussed with were aware of it. I therefore make no paternity claim.

a Schnorr random element of Σ^ω. Hence, one can construct, like in Theorem 1.5.10, a Schnorr test in Σ^ω that covers $x_{b'}$.

Step 4. Since Σ^ω and 2^ω are more or less isomorphic as representations of $[0,1]$ (it is at this point of the proof that we skip some tedious details, since technically one needs to take care of the "more or less" part of this statement), in an effective way, i.e. one can go from one to the other via a computable isomorphism. Hence, the Schnorr test constructed in Step 3 can effectively be turned into a Schnorr test in 2^ω that covers the expansion x_2 of x is base 2.

Step 5. We now use again Theorem 1.5.10, but in the opposite direction, i.e. from the Schnorr test of 2^ω we got in Step 4, we can construct a computable martingale – which can be taken normed – that succeeds against the representation x_2 of x in 2^ω.

Step 6. The construction we made during the last 5 steps was made for a particular base b', a particular digit τ of $\{0,\dots,b'-1\}$ and a particular rational δ. However, this construction is uniform i.e. at each step the constructions are effective given these three parameters. Hence, from such a triple (b',τ,δ) one can construct a normed martingale $d_{(b',\tau,\delta)}$ that succeeds against all $x \in [0,1]$ whose representation $x_{b'}$ in base b' has a limsup frequency of the digit τ that is greater than $1/b'+\delta$. The effective nature of the construction allows us to compute a mixture of all such strategies: let (b'_n,τ_n,δ_n) be a computable enumeration of all triples (b',τ,δ) with $b' \geq 2$, $\tau \in \{0,\dots,b'-1\}$ and $\delta \in \mathbb{Q}$; consider:

$$d^* = \sum_{n\in\mathbb{N}} 2^{-n}\, d_{(b'_n,\tau_n,\delta_n)}$$

Then d^* is a martingale (as a weighted sum of martingales with initial capital no greater than 2) that is computable (by the above argument) and succeeds against all representations in base 2 of reals in $[0,1]$ that are not absolutely normal.

Step 7. To conclude the argument, we apply Proposition 1.4.10: there exists a computable sequence $\alpha \in 2^\omega$ such that $\alpha \notin \mathrm{Succ}(d^*)$. Hence, by construction of d^*, α is the representation in base 2 of a number $y \in [0,1]$ that is absolutely normal. Moreover, y is computable since α is.

1.7 Non-monotonicity for selection rules and martingales

A stronger model of selection rules and martingales was introduced by Kolmogorov [29] and Loveland [38]: in that model, the player is allowed to read (resp. select, bet on) the bits in any order, with however the restriction that he should not select (resp. bet on) a bit that he has already seen. This insight leads to the notions of non-monotonic rules and non-monotonic martingales which, as we will see later on, can have for some purposes much more power than their monotonic counterparts.

Definition 1.7.1. *A* NON-MONOTONIC SELECTION RULE *is a function* $\sigma : 2^{<\omega} \to \mathbb{N} \times \{\texttt{select},\texttt{scan}\}$.

$\sigma(w) = (k, \texttt{select})$ (resp. $\sigma(w) = (k, \texttt{scan})$) means that having sequentially read the bits $w_{(0)}, \dots, w_{(|w|-1)}$, the selection rule decides to select (resp. scan) the n-th bit of the sequence. To describe how a select rule runs on a sequence, we adapt the formalism of monotonic select rules by adding an history of the previously read bits, together with their positions in the sequence (which are forbidden for selection for the rest of the game).

Let σ be a non-monotonic selection rule. We run σ on a sequence α as follows. Set $\beta^{(0)} = \epsilon$, $w^{(0)} = \epsilon$, $h^{(0)} = \emptyset$. The $\beta^{(n)}$, $w^{(n)}$ and $h^{(n)}$ will represent respectively the bits selected, the bits seen, and the positions visited before stage n. By induction, for all $n \geq 0$:

- if $\sigma(w^{(n)}) = (k, \texttt{scan})$, set $\beta^{(n+1)} = \beta^{(n)}$, $w^{(n+1)} = w^{(n)}\alpha_k$ and $h^{n+1} = h^{(n)} \cup \{k\}$
- if $\sigma(w^{(n)}) = (k, \texttt{select})$, and $k \in h^{(n)}$ (forbidden selection), set $\beta^{(n+1)} = \beta^{(n)}$, $w^{(n+1)} = w^{(n)}\alpha_k$ and $h^{n+1} = h^{(n)} \cup \{k\}$
- if $\sigma(w^{(n)}) = (k, \texttt{select})$, and $k \notin h^{(n)}$, set $\beta^{(n+1)} = \beta^{(n)}\alpha_n$, $w^{(n+1)} = w^{(n)}\alpha_k$ and $h^{n+1} = h^{(n)} \cup \{k\}$

Like in the case of monotonic selection rules, the sequence of strings $(\beta^{(n)})_{n\in\mathbb{N}}$ is non-decreasing for the prefix order $\sqsubseteq$. Hence, either it is stationnary, in which case we set β to be the limit of the sequence. Or the sequence is not stationnary, in which case the $\beta^{(n)}$ are all prefixes of an infinite binary sequence, which we call β. In both cases, β is called the SUBSEQUENCE OF α SELECTED BY σ, and we denote it by $\sigma[\alpha]$.

Definition 1.7.2. *A sequence α is* KOLMOGOROV-LOVELAND STOCHASTIC *(KL-stochastic for short) if for every total computable non-monotonic selection rule, either $\beta = \sigma[\alpha]$ is finite, or it satisfies*

$$\lim_{n\to+\infty} \frac{\#0(\beta\restriction_n)}{n} = \frac{1}{2}$$

We denote by **KLStoch** *the set of Kolmogorov-Loveland stochastic sequences.*

Similarly to non-monotonic selection rules, we could define non-monotonic martingales for non-monotonic betting games. However, martingales are classically defined as being real-valued functions. Since at each move of the game we also need to specify the position of the bit we are going to bet on, we prefer the term *strategy*. And since we have in mind a characterization of KL-stochasticity via strategies, we define the notion of strategy as follows:

Definition 1.7.3. *A* STRATEGY *is a function*

$$S : 2^{<\omega} \to \mathbb{N} \times \Big(\{\text{scans}\} \cup [-1, 1]\Big)$$

$S(w) = (n, \rho)$, with $\rho \in [-1, 1]$ means that, having sequentially read the bits $w_{(0)}, \ldots, w_{(|w|-1)}$, the strategy decides to bet a fraction ρ of its current capital on the value of the n-th bit of the sequence to be 0 (with the same convention as before that a negative value of ρ means a bet on the value 1 of a fraction $(-\rho)$ of the current capital). $S(w) = (n, \texttt{scan})$ means that S simply decides to read the n-th bit without betting anything. It first seems that this is equivalent to $S(w) = (n, 0)$, but in the sequel, we will need to make the distinction as we will be interested in strategies with a capital that is exponential in the number of bets, but not necessarily in the number of moves. In order to run a strategy on a sequence α, we will need the following objects. Before the n-th move is made, $w^{(n)}$ will denote the string made of bits previously seen (with $w^{(0)} = \epsilon$), $h^{(n)}$ the set of positions that were previously visited (with $h^{(0)} = \emptyset$), $W^{(n)}$ the current capital (with $W^{(0)} = 1$) and $N^{(n)}$ the number of bets previously made (i.e. the number of moves among the first n when the strategy did *not* choose to scan a bit but rather to bet money on it), with $N^{(0)} = -1$ by convention. We proceed by induction:

- if $S(w^{(n)}) = (k, \texttt{scan})$, set $w^{(n+1)} = w^{(n)}\alpha_k$, $h^{n+1} = h^{(n)} \cup \{k\}$, $W^{(n+1)} = W^{(n)}$ and $N^{(n+1)} = N^{(n)}$.
- if $S(w^{(n)}) = (k, \rho)$, and $k \in h^{(n)}$ (forbidden position) set $w^{(n+1)} = w^{(n)}\alpha_k$, $h^{n+1} = h^{(n)} \cup \{k\}$, $W^{(n+1)} = W^{(n)}$ and $N^{(n+1)} = N^{(n)}$.
- if $S(w^{(n)}) = (k, \rho)$, and $k \notin h^{(n)}$, set $w^{(n+1)} = w^{(n)}\alpha_k$, $h^{n+1} = h^{(n)} \cup \{k\}$, $N^{(n+1)} = N^{(n)} + 1$ and $W^{(n+1)} = (1 + \rho)W^{(n)}$ if $\alpha_k = 0$, and $W^{(n+1)} = (1 - \rho)W^{(n)}$ if $\alpha_k = 1$.

We will write $W_n(\alpha, S)$ for the CAPITAL of S after the $(n-1)$-th move when playing against α. We will write $V_n(\alpha, S)$ for the capital of S after the $(n-1)$-th bet during the game against α i.e. formally $V_n(\alpha, S) = W_i(\alpha, S)$ where i is the smallest integer such that $N^{(i)} = n - 1$.

A strategy S SUCCEEDS on a sequence α if $\limsup_n W_n(\alpha, S) = +\infty$, or equivalently if $\limsup_n V_n(\alpha, S) = +\infty$. This allows us to define Kolmogorov-Loveland randomness:

Definition 1.7.4. *A sequence $\alpha \in 2^\omega$ is* KOLMOGOROV-LOVELAND RANDOM *(KL-random for short) if no total computable strategy succeeds on α. We denote by* **KLR** *the set of KL-random sequences.*

Remark 1.7.5. *The analogue of Remark 1.4.8 holds for Kolmogorov-Loveland randomness as well: replacing* lim sup *by* lim *in the definition of success for a strategy does not affect the notion of KL-randomness.*

Since non-monotonic selection rules and strategies generalize respectively monotonic selection rules and martingales, it is clear that KL-stochasticity implies Church stochasticity and that KL-randomness implies computable randomness.

Proposition 1.7.6. *If $\alpha \in 2^\omega$ is Kolmogorov-Loveland stochastic, it is Church stochastic. If $\alpha \in 2^\omega$ is Kolmogorov-Loveland random, it is computably random.*

Remark 1.7.7 (Merkle)**.** *Another important fact about KL-randomness and stochasticity is that if one allows partial computable strategies (resp. partial computable non-monotonic selection rules), the definition remains unchanged. To see this, consider a partial computable strategy S that succeeds on a sequences α, i.e. S makes an infinite amount of money by betting against α. Thus, either S makes an infinite amount of money by betting on bits in even positions or it makes an infinite amount of money by betting on the bits in odd positions (or both). Suppose, without loss of generality the first holds. Then, turn S into a strategy S' which simulates S and does the following:*

- *when S bets on an even bit, S' bets the same thing (same fraction of its capital on the same bit)*
- *when S bets on an odd bit, S' scans this bit*
- *between two actions of the above two types, when waiting for the computation of S to terminate, S' scans odd bits of the sequence, "for free"*

Then the strategy S' is total (if S does not terminate, it keep scanning odd bits forever, hence always does something), and makes an infinite amount of money on even bits as S does. A similar argument holds for KL-stochasticity, see Merkle [45] for a detailed discussion.

Also, the fact that stochasticity can be expressed via an exponential winning condition is still true in the non-monotonic setting. Let us extend the concept of s-success to strategies:

Definition 1.7.8. *A strategy S s-*SUCCEEDS *on a sequence $\alpha \in 2^\omega$ if*

$$\limsup_{n\to+\infty} \frac{W_n(\alpha, S)}{2^{(1-s)n}} = +\infty$$

S WEAKLY *s-*SUCCEEDS *on α if*

$$\limsup_{n\to+\infty} \frac{V_n(\alpha, S)}{2^{(1-s)n}} = +\infty$$

The notion of s-success expresses a gain of money that is exponential in the number of moves, while for weak s-success it is only exponential in the number of bets, all the moves of type `scan` being made for free.

The following theorem expresses how the notion of weak s-success relates to selection rules:

Theorem 1.7.9. *(i) Let $\alpha \in 2^\omega$. If there exists a computable non-monotonic selection rule σ that selects an infinite $\beta = \sigma[\alpha]$ with $\mathrm{Bias}(\beta) \geq \delta > 0$ then there exists a strategy S, computable with oracle δ, such that for all $s > \mathcal{H}\left(\frac{1}{2} + \delta\right)$, S weakly s-succeeds against α.*

(ii) If there exists a computable strategy S that weakly s-succeeds against a sequence $\alpha \in 2^\omega$ then there exists a non-monotonic selection rule σ, computable with oracle s, such that $\mathrm{Bias}(\sigma[\alpha]) \geq \delta$ where δ is such that $\mathcal{H}\left(\frac{1}{2} + \delta\right) = s$.

Proof. (i) Let σ be a computable non-monotonic selection rule selecting a subsequence β such that $\mathrm{Bias}(\beta) \geq \delta > 0$. Without loss of generality, suppose that

$$\limsup_{n \to +\infty} \frac{\#0(\beta\restriction_n)}{n} \geq \frac{1}{2} + \delta$$

Let then S be the strategy that follows σ and:

- when σ scans a bit, S scans it too
- when σ selects a bit, S bets a fraction $\rho = 2\delta$ on the value of this bit to be 0

S is computable with oracle δ and

$$V_n(\alpha, S) = (1 + 2\delta)^{\#0(\beta\restriction_n)}(1 - 2\delta)^{\#1(\beta\restriction_n)}$$

and this, by the same computation as in the proof of Proposition 1.4.13, implies that S weakly s-succeeds on α for all $s > \mathcal{H}\left(\frac{1}{2} + \delta\right)$.

(ii) This is just an easy adaptation of Theorem 1.4.16 to the non-monotonic setting. ■

The second part of this theorem shows, similarly to the monotonic case, that KL-randomness implies KL-stochasticity.

Corollary 1.7.10. *If $\alpha \in 2^\omega$ is Kolmogorov-Loveland random, it is Kolmogorov-Loveland stochastic.*

We will see in Chapter 3 that this inclusion cannot be reversed. Another interesting question is how these new notions compare to Martin-Löf randomness. Here is a partial answer to this question:

Proposition 1.7.11. *If $\alpha \in 2^\omega$ is Martin-Löf random, it is Kolmogorov-Loveland random.*

Proof. This is because Theorem 1.5.1 also holds true in the non-monotonic setting, i.e. for a fixed strategy S and constant $c > 0$, the measure of $\mathcal{U}_n = \{\alpha \in 2^\omega : \exists n\, W_n(\alpha, S) \geq 2^n\}$ is bounded by 2^{-n} (since by convention a strategy starts with capital 1). From a measure-theoretic point of view, this is rather obvious: if the bits of α are chosen randomly and independently, by symmetry, the order of

the bets should not influence the expectancy of the game (for the reader who does not find this convincing, one can also notice that given α and S, the sequence of $W_n(\alpha, S)$ is a nonnegative martingale in the classical sense and then it suffices to apply Doob's inequality – see for example Jacod and Protter [25] – to get the result). And as for the monotonic case, if S is computable $\mathcal{U}_n$ is effectively open. The $\mathcal{U}_n$ are thus a Martin-Löf test covering all the sequences on which S succeeds. ■

The question whether the converse of this proposition holds true is one of the most fundamental open questions in the field of algorithmic randomness.

1.8 Randomness and Baire category

Both (Lebesgue) measure and Hausdorff dimension can be seen as evaluations of how big a set is. Another well-known approach is Baire category. The central theorem of this theory is the so-called Baire category theorem, which asserts that given a compact (or complete metric) space $\mathcal{X}$, any countable intersection of dense open subsets of $\mathcal{X}$ is dense in $\mathcal{X}$. Equivalently, if we call *nowhere dense* a subset of $\mathcal{X}$ whose closure has empty interior, the Baire category theorem tells us that a countable union of nowhere dense subsets of $\mathcal{X}$ has empty interior. Any countable union of nowhere dense subsets of $\mathcal{X}$ is then said to be *meager* in $\mathcal{X}$. In a sense, a meager set is the topological analogue of a nullset. In particular, a subset of a meager set is meager, and a countable union of meager sets is meager. The complement of a meager set is said to be *comeager*.

In the same way we defined random sequences as sequences satisfying all "effective" properties of measure 1 (getting different classes for different interpretations of the word "effective"), we can define various classes of sequences which satisfy all "effective" co-meager properties. The following notion will be sufficient for our purposes (for a more complete discussion on genericity, see Jockusch [26]):

Definition 1.8.1 (Kurtz [32]). *A sequence α is* WEAKLY GENERIC *if it belongs to every dense c.e. open set. We denote by* **WG** *the set of weakly generic sequences.*

As there are only countably many effectively open sets, the intersection of all dense effectively open sets is non-empty. More precisely, with the above terminology, the set of weakly generic sequences is comeager, i.e. intuitively "topologically big". How random are (or can be) the elements of this set? The answer is: they are not random at all! To see this, define for example the set

$$\mathcal{U}_k = \{\alpha \in 2^\omega : (\exists n \geq k)\ \#0(\alpha\restriction_n) > 2\,\#1(\alpha\restriction_n)\}$$

It is clear that for all k, $\mathcal{U}_k$ is a c.e. open subset of 2^ω. Moreover, it is dense since for all cylinder $[w]$, $w0^\omega \in [w] \cap \mathcal{U}_k$. Hence, any weakly generic sequence α belongs to all $\mathcal{U}_k$, which means that for infinitely many n, $\alpha\restriction_n$ contains twice more zeros than ones. This is enough to prove that no weakly generic sequence α can be Church stochastic. Refining this argument, we can prove:

Proposition 1.8.2. $\dim_{\text{comp}}(\mathbf{WG}) = 0$

Proof. To prove this, we use the characterization of computable dimension presented in Theorem 1.5.16. We need to prove that for all $s > 0$, there exists a computable martingale d which s-succeeds on every weakly generic sequence. Let s be a positive rational. Let $\rho < 1$ be a rational such that $(1+\rho) > 2^{(1-s)}$. Let d be the computable martingale which at each moves bet the fraction ρ of its capital on the value 0, i.e. formally:

$$d(w) = (1+\rho)^{\#0(w)}(1-\rho)^{\#1(w)}$$

for all $w \in 2^{<\omega}$. For all k, set

$$\mathcal{U}_k = \{\alpha \in 2^\omega : (\exists n \geq k)\, d(\alpha\restriction_n) \geq 2^{(1-s)n+k}\}$$

It is clear that $\mathcal{U}_k$ is a c.e. open set. We claim that is it dense. For each cylinder $[w]$, the sequence $w0^\omega$ is in $\mathcal{U}_k \cap [w]$ because:

$$d(w0^n) \geq (1-\rho)^{|w|}(1+\rho)^n$$

and since $2^{(1-s)n} = o((1+\rho)^n)$, for n large enough the right-hand side is greater than $2^{(1-s)(|w|+n)+k}$. Hence, $\mathcal{U}_k$ intersects every open cylinder hence is dense. Now, if α is any weakly generic sequence, it must belong to all $\mathcal{U}_k$, which by definition means that d s-succeeds on α. Since this is true for all $\alpha \in \mathbf{WG}$, this proves $\dim_{\text{comp}}(\mathbf{WG}) \leq s$. And since this is true for all s, $\dim_{\text{comp}}(\mathbf{WG}) = 0$. ■

Since any Church stochastic (a fortiori: KL-stochastic, Schnorr random, computably random, Martin-Löf random) sequence has computable dimension 1, this means that the classes **MLR**, **CR** and **SR**, **ChStoch**, **KLStoch** are all disjoint from **WG**. And since **WG** is comeager, this means that they are all meager. You may have noticed that the class **WR** is not in this list. Here is why:

Proposition 1.8.3. *Every weakly generic sequence is weakly random.*

Given how non-random weak generic sequences are, this definitely rules out weak randomness as a suitable notion of randomness.

Proof. If α is weakly generic, it belongs to all dense c.e. open sets. Since all c.e. open sets of measure 1 are dense, α belongs to all of them, i.e. α is weakly random by definition. ■

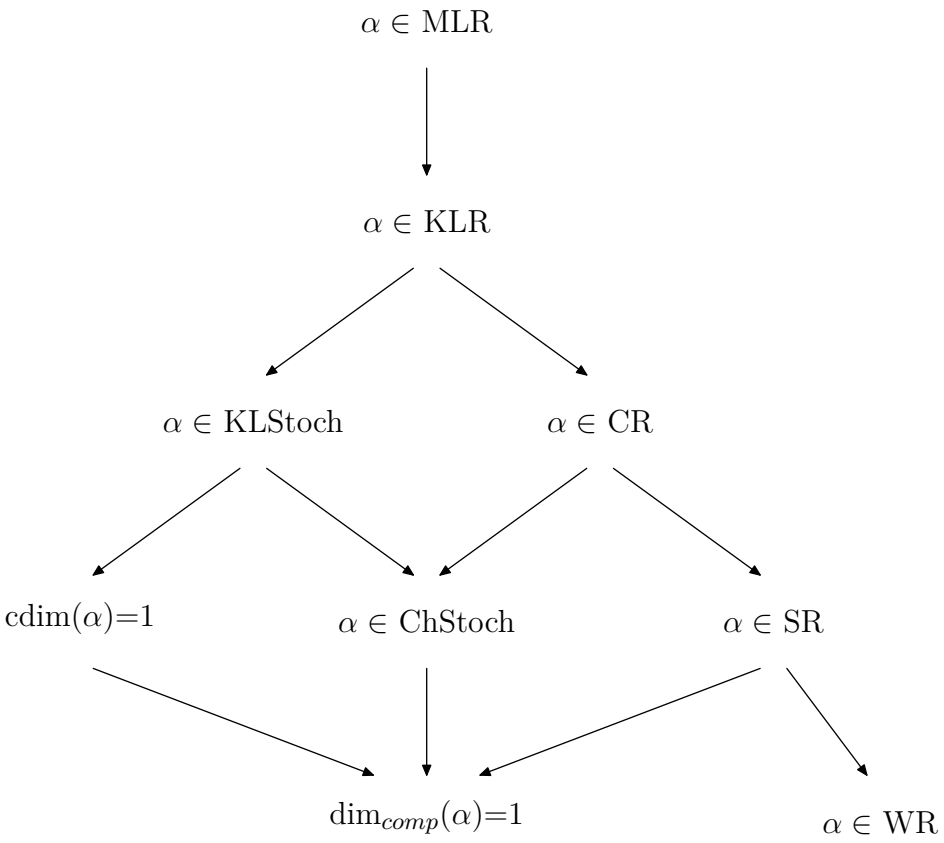
Corollary 1.8.4. *The class* **WR** *is comeager, and some of its elements have computable dimension* 0.

Proof. Since **WG** is comeager and $\mathbf{WG} \subseteq \mathbf{WR}$, **WR** is comeager. And all elements of **WG** have computable dimension 0 by Proposition 1.8.2. ■

1.9 Relations between randomness notions

Before moving on to the next chapter, where we are going to introduce the fundamental notion of *Kolmogorov complexity*, we sum up all the relations between the randomness notions we have defined in this chapter.

For a sequence $\alpha \in 2^\omega$, the following implications hold. Moreover, no other implication is true in general, except a possible equivalence between ML-randomness and KL-randomness. As we have not proven all these implications (and non-implications) yet, we also provide a table indicating where to find the proofs in this thesis, together with the reference where the result was originally proven.



MLR	$\Longrightarrow$	**KLR**	Proposition 1.7.11	Muchnik et al. [50]
KLR	$\Longrightarrow$	**KLStoch**	Corollary 1.7.10	folklore
KLStoch	$\Longrightarrow$	cdim $= 1$	Theorem 2.2.31	Merkle et al. [47]
KLStoch	$\Longrightarrow$	**ChStoch**	Proposition 1.7.6	trivial
CR	$\Longrightarrow$	**ChStoch**	Corollary 1.4.15	folklore
CR	$\Longrightarrow$	**SR**	Corollary 1.5.11	Schnorr [52]
cdim $= 1$	$\Longrightarrow$	$\dim_{\text{comp}} = 1$	page 11	trivial
ChStoch	$\Longrightarrow$	$\dim_{\text{comp}} = 1$	Proposition 1.5.18	folklore
SR	$\Longrightarrow$	$\dim_{\text{comp}} = 1$	Proposition 1.5.17	folklore
SR	$\Longrightarrow$	**WR**	Corollary 1.5.14	folklore
KLStoch	$\not\Longrightarrow$	**WR**	Corollary 3.2.12	Merkle et al. [47]
CR	$\not\Longrightarrow$	cdim $= 1$	Proposition 2.2.29	Muchnik et al. [50]
cdim $= 1$	$\not\Longrightarrow$	**ChStoch**	Proposition 2.2.29	folklore
SR	$\not\Longrightarrow$	**ChStoch**	Corollary 2.2.23	Wang [61]
WR	$\not\Longrightarrow$	$\dim_{\text{comp}} = 1$	Corollary 1.8.4	folklore

Chapter 2

Randomness and Kolmogorov complexity

In this chapter, we introduce the fundamental notion of Kolmogorov complexity and discuss some of its important properties (for a very complete survey, see Li and Vitanyi [37]). Intuitively, Kolmogorov complexity measures the "random content" of finite discrete objects, hence allowing us to define which ones are random and which ones are not. We will discuss two versions of Kolmogorov complexity: the plain version and the prefix version. We then discuss how the randomness notions we have studied in the previous chapter are related to Kolmogorov complexity. As we will see some of these notions (like Martin-Löf randomness) admit elegant characterizations in terms of Kolmogorov complexity, while others (like computable randomness, Schnorr randomness) do not. In the last part of the chapter, we will consider computable upper bounds of Kolmogorov complexity. As we will see, this turns out to be a rather unifying point of view, as one can express many more notions of randomness with this setting than with Kolmogorov complexity alone.

2.1 Kolmogorov complexity

2.1.1 Plain Kolmogorov complexity

Definition of plain Kolmogorov complexity

In parallel to the attempts made to define randomness for infinite binary strings, a very elegant and powerful theory of complexity and randomness for finite objects emerged in the late 1960's from the work of Solomonoff, Kolmogorov and Chaitin. The central notion that came out of their work is now known as *Kolmogorov complexity* (or *algorithmic complexity*). The intuition behind this notion is the following. A finite object is not random if it has some kind of pattern, i.e. part of the information it contains is redundant. Such an object should then admit a description shorter than itself (the description taking advantage of the redundancy). And conversely, if an object admits a description shorter than itself, this means that it contains some redundant information, hence is not random. This is

what we call the *uncompressibility paradigm*:

Uncompressibility paradigm.
A finite object is random if there exists no shorter description of it than itself.

More generally, to measure how random (or non-random) an object is, it is natural to define the *complexity* of an object by the size of its shortest description.

All this is highly informal though, and we need to give a precise definition of what we mean by "description", as an informal one leads to the famous *Berry's paradox*. To understand this paradox, consider the following:

"the smallest positive integer that cannot be described by less than three-hundred ASCII characters"

Since there are only finitely many descriptions of less than 300 characters (at most 256^{300}), there must be some positive integers that cannot be described with less than that. Hence, we can consider the smallest of them. But then, the definition we wrote above describes it completely and yet has less than 300 characters, a contradiction.

Once again, computability theory allows us to state a rigorous definition.

Definition 2.1.1. *A* MACHINE *is a partial computable function* $M : 2^{<\omega} \to 2^{<\omega}$. *We usually call "programs for* M*" the elements of* $dom(M)$. *Let* M *be a machine, and* $w \in 2^{<\omega}$. *We set:*

$$C_M(w) = \min\{|p| : p \in 2^{<\omega} \ \wedge \ M(p) = w\}$$

with the convention that $C_M(w) = +\infty$ *if there is no* p *such that* $M(p) = w$. *We call* $C_M(w)$ *the* KOLMOGOROV COMPLEXITY OF w WITH RESPECT TO THE MACHINE M.

In this definition, M can be seen as a decompression algorithm, and a p such that $M(p) = w$ as a compressed version of w for this decompression algorithm. The Kolmogorov complexity of w with respect to M measures how much remains of w after compression. Notice that for the moment, we do not worry about the feasibility of the compression (we will come back to this in Section 2.3)

Of course, the above definition depends on the chosen machine M whereas we would like to have a universal definition of complexity. The following easy but fundamental theorem will allow us to give one.

Theorem 2.1.2 (Additive Optimality for C)**.** *There exists an* ADDITIVELY OPTIMAL MACHINE, *i.e., a machine* $\mathbb{U}$ *such that for every machine* M*:*

$$C_{\mathbb{U}} \leq C_M + O(1)$$

In other words, the machine $\mathbb{U}$ is better than any other at describing any string, up to an additive constant. Let us prove the above theorem briefly:

Proof. Let $\{M_e\}_{e\in\mathbb{N}}$ be a computable enumeration of machines. Let $\mathbb{U}$ be the machine defined on strings of type 0^e1p (with $p \in 2^{<\omega}$) by $\mathbb{U}(0^e1p) = M_e(p)$. It is clear that $\mathbb{U}$ is partial computable, and for all strings $w \in 2^{<\omega}$: if for some $e \in \mathbb{N}$ and $p \in 2^{<\omega}$ one has $M_e(p) = w$, then $\mathbb{U}(0^e1p) = w$. This means that $C_{\mathbb{U}} \leq C_{M_e} + (e+1)$ for all $e \in \mathbb{N}$.

■

For the rest of this thesis, we fix an additively optimal machine $\mathbb{U}$ and we define the Kolmogorov complexity as follows:

Definition 2.1.3. *For all strings $w \in 2^{<\omega}$ we set $C(w) = C_{\mathbb{U}}(w)$, and we call this quantity the* Kolmogorov complexity of w.

We sometimes use the term plain Kolmogorov complexity to distinguish it from prefix Kolmogorov complexity which we will define later. Also, we will sometimes talk about the complexity $C(n)$ of an integer n: this is just an abbreviation of $C(\mathrm{Bin}(n))$.

The Kolmogorov complexity function depends on the particular machine $\mathbb{U}$ we fixed, but by Theorem 2.1.2, if $\mathbb{U}'$ is another optimal machine, the difference $|C_{\mathbb{U}} - C_{\mathbb{U}'}|$ is bounded by a constant. It can seem strange at first to define an integer-valued function $w \mapsto C(w)$ up to a bounded additive term, but we have to keep in mind that Kolmogorov complexity is an *asymptotic* theory. For example, it does not really make sense to say that $C(10101101001001) = 7$ because once again the value $C(10101101001001)$ depends on $\mathbb{U}$ which we did not define explicitely. Hence, in the sequel, all the statements we will make that involve Kolmogorov complexity will be "up to an additive constant".

Some important properties of plain Kolmogorov complexity

One first important property that should be noticed from the above optimality theorem is that the plain Kolmogorov complexity of a string w never exceeds its length (up to an additive constant):

Proposition 2.1.4. *There exists a constant $c > 0$ such that $C(w) \leq |w| + c$ for all strings w.*

Proof. This is just an application of Theorem 2.1.2, with M being the identity machine (for which $C_M(w) = |w|$ for all w). ■

This is of course rather intuitive since one can always "describe" a string by giving it explicitly.

Here is another application of the optimality theorem. It states that the image $f(w)$ of w via a computable function f has complexity no bigger that w, up to an additive constant. This is natural, as to describe w, one only needs to give f and w (and the description of f is independent of w).

Proposition 2.1.5. *Let $f : 2^{<\omega} \to 2^{<\omega}$ be a computable function. For all strings w, $C(f(w)) \leq C(w) + O(1)$.*

Proof. Let $\mathbb{U}$ be the above optimal machine and let M be the machine defined by $M(p) = f(\mathbb{U}(p))$. Then, if $\mathbb{U}(p) = w$, then $M(p) = f(w)$. This proves that $C_M(f(w)) \leq C(w)$ for all w. Applying Theorem 2.1.2, we get $C(f(w)) \leq C_M(f(w)) + O(1) \leq C(w) + O(1)$. ∎

Although simple, this proposition is very useful. We will frequently use it in the sequel, most of the time implicitly: whenever we show that a string w' can be obtained by a simple transformation of w, where the function performing the transformation does not depend on w, w', we will conclude that $C(w') \leq C(w) + O(1)$. Here is an example of the usefulness of this argument. We would like to define the Kolmogorov complexity of a pair of strings u, v. To do so, we define it as being the quantity $C(\langle u, v\rangle)$ where $\langle ., .\rangle$ is a computable encoding (i.e. injection) of $2^{<\omega} \times 2^{<\omega}$ into $2^{<\omega}$ (the Cantor bijection is such an encoding). One can be concerned about the fact that it then depends of the particular encoding we chose. Well, it does, but only up to an additive constant. Indeed, if $\langle ., .\rangle_1$ and $\langle ., .\rangle_2$ are two different encodings, given two strings u and v, one can get $\langle u, v\rangle_2$ from $\langle u, v\rangle_1$, simply by a decoding $\langle u, v\rangle_1$ to get u and v and then re-encoding them as $\langle u, v\rangle_2$. Hence, $C(\langle u, v\rangle_2) \leq C(\langle u, v\rangle_1) + O(1)$. By symmetry, $C(\langle u, v\rangle_2) = C(\langle u, v\rangle_1) + O(1)$. Hence, we fix a particular encoding $\langle ., .\rangle$ once and for all and we define the complexity of a pair as the complexity of its encoding.

By the way, how large is the complexity of a pair of strings? As one can expect, not much bigger than the sum of their complexities. However, the precise statement of this fact involves a additional logarithmic factor.

Proposition 2.1.6. *Let $u, v \in 2^{<\omega}$. One has $C(u, v) \leq C(u) + C(v) + O\big(\min(\log C(u), \log C(v))\big)$*

Proof. Where does this logarithmic term come from? It seems that if we have a shortest program q' for u and a shortest program q'' for v, then $q = q'q''$ contains enough information to retrieve u and v, hence $C(u, v) \leq |q'| + |q''| = C(u) + C(v)$. However, this is too naive: if we are given q and we want to compute u and v, we first need to split q into its two parts q' and q''. But, not knowing q' nor q'', we do not know where to split q. There could exist another way to split q into two parts p' and p'' (with p' and p'' in $dom(\mathbb{U})$) leading to a different pair $\mathbb{U}(p'), \mathbb{U}(p'')$. Hence, some additional information is needed to decode $q = q'q''$. Providing $|q'|$ or $|q''|$ is enough, as if we know either of those, we know where to split r. Let M be the machine whose domain is contained in $\{0^k 1 r : k \in \mathbb{N}, r \in 2^{<\omega} \text{ and } |r| \geq k\}$ and which does the following. On input $0^k 1 r$, M splits r into two parts: $r = pq$ with $|p| = k$. Then, it computes $l = \mathrm{Bin}^{-1}(p)$. Then, it splits q into two parts $q = q'q''$ with $|q| = l$. Finally, it outputs $\langle \mathbb{U}(q'), \mathbb{U}(q'')\rangle$. It is easy to see that M is indeed computable. Let $u, v \in 2^{<\omega}$. Let q' be a shortest program for u and q'' a shortest

program for v (i.e. $\mathbb{U}(q') = u$, $\mathbb{U}(q'') = v$, $C(u) = |q'|$, and $C(v) = |q''|$). Then, on input $0^{|\mathrm{Bin}(|q'|)|}1\mathrm{Bin}(|q'|)q'q''$, the machine M exactly outputs $\langle u, v\rangle$. Thus:

$$\begin{aligned} C(u,v) &\leq C_M(u,v) + O(1) \\ &\leq 2\left|\mathrm{Bin}\left(|q'|\right)\right| + |q'| + |q''| + O(1) \\ &\leq C(u) + C(v) + O(\log C(u)) \end{aligned}$$

By symmetry, $C(v,u) \leq C(u) + C(v) + O(\log C(v))$. And since $C(u,v) = C(v,u) + O(1)$, the theorem is proved. ■

Incompressible strings

Now that we have a universal measure of complexity, it is natural to look back at our original goal: to give a good definition of randomness for finite objects. Kolmogorov complexity allows us to do this: an finite string is *random* if its Kolmogorov complexity is maximal. However, since the Kolmogorov complexity of a given string w depends on the particular choice of the machine $\mathbb{U}$, we need to introduce a parameter in the defintion.

Definition 2.1.7. *We say that a string w is k-incompressible$_C$ (where $k \in \mathbb{N}$) if $C(w) \geq |w| - k$.*

This definition is nice, but there is one thing we need to check, namely that most strings are incompressible (it would be quite strange to have a definition of "randomness" which a string chosen at random has a high probability not to have). The next proposition is comforting: this is indeed the case.

Proposition 2.1.8 (C-counting theorem)**.**
(i) There are at most $2^k - 1$ strings w such that $C(w) < k$.
(ii) If the bits of a string w of length n are chosen independently with probabilities $(1/2, 1/2)$, then for all $k \in \mathbb{N}$, the probability that $C(w) \geq |w| - k$ is greater than $1 - 2^{-k}$.

Proof. (i) is a simple counting argument: there are 2^i programs for $\mathbb{U}$ of length i. A fortiori there are at most 2^i programs of length i in the domain of $\mathbb{U}$. Hence, there are at most $2^0 + 2^1 + 2^2 + ... + 2^{k-1} = 2^k - 1$ programs of length smaller than k in the domain of $\mathbb{U}$. Consequently, the image set of these programs by $\mathbb{U}$, which is by definition the set of strings w with complexity lower than k, has cardinality at most $2^k - 1$.
(ii) is an easy consequence of (i). For a string w of size n, the probability that $C(w) < n - k$ is equal to $\frac{1}{2^n}\#\{w : |w| = n \wedge C(w) < n - k\}$ which by (i) is smaller than $\frac{1}{2^n}\left(2^{n-k} - 1\right)$ that is smaller than 2^{-k}. ■

Non-computability of Kolmogorov complexity

Well, all this theory sounds very nice, but before we continue on, the reader should be warned that it has a major drawback.

Proposition 2.1.9. *The Kolmogorov complexity function $w \mapsto C(w)$ is not computable.*

Proof. The proof is inspired by Berry's paradox. Suppose C is computable. This would allow us to find effectively for every k, the least string (in the length-lexicographic order, say) that has Kolmogorov complexity greater than k (in some sense,"the smallest string that cannot be described in less than k bits"). Call this string w_k. Since w_k can be computed using k only, and since k written in binary has length about $\log k$, it follows that $C(w_k) \leq \log k + O(1)$. But for k large enough, this contradicts the definition of w_k. The completely formal argument is as follows: let M be the machine such that $M(\mathrm{Bin}(k)) = w_k$ for all k. We have for all k, $C_M(w_k) \leq |\mathrm{Bin}(k)| \leq \log k + O(1)$. By the optimality theorem (Theorem 2.1.2) we have $C(w_k) \leq C_M(w_k) + O(1) \leq \log k + O(1)$ for all k. But this contradicts $C(w_k) > k$ for k large enough. ■

The same kind of argument allows us to prove the following result that will be useful in the sequel:

Proposition 2.1.10. *The function B defined by $B(k) = \max\{n \in \mathbb{N} : \ C(n) \leq k\}$ dominates any computable function f.*

Proof. Suppose the contrary, that is there exists a computable function f such that $f(k) > B(k)$ for infinitely many k. Then for infinitely many k, $f(k) + 1$ has complexity greater than k. But since f is computable, for all k, $C(f(k) + 1) \leq \log k + O(1)$, a contradiction. ■

Fortunately, the function C can be "computed" in a weaker sense. Namely, it is *enumerable from above*.

Proposition 2.1.11. *The function C is enumerable from above, i.e., there exists a total computable $C(.)[.] : 2^{<\omega} \times \mathbb{N} \to \mathbb{N}$ such that for all $w \in 2^{<\omega}$ the sequence $C(w)[0], C(w)[1], C(w)[2], \ldots$ is non-increasing and converges to $C(w)$.*

Proof. Let c be a constant such that $C(w) \leq |w| + c$ for all c (such a constant exists by Proposition 2.1.4). For all $w \in 2^{<\omega}$, set $C(w)[0] = |w| + c$. Now, given a string w, run all the programs p for $\mathbb{U}$ (the additively optimal machine) in parallel. For all $t > 0$, define $C(w)[t]$ to be the minimal length of all programs which output w in less than t steps of computation (if there is no such program, set $C(w)[t] = |w| + c$). Clearly, $C(w)[t]$ is uniformly computable in k, is nonincreasing, and converges to $C(w)$ as t tends to $+\infty$ as the shortest program will eventually be found. ■

The non-computability of C can be stated in a more precise way: C is exactly as uncomputable as the halting problem.

Theorem 2.1.12. *The Turing degree of the Kolmgorov complexity function C is $\mathbf{0}'$ (i.e. $C \leq_T \mathbf{0}'$ and $\mathbf{0}' \leq_T C$).*

Proof. Let B be the function defined in Proposition 2.1.10. We will prove:

$$\mathbf{0}' \leq_T B \leq_T C \leq_T \mathbf{0}'$$

($C \leq_T \mathbf{0}'$). This part is easy. Let $\mathbb{U}$ be the additively optimal machine for C. We know that there exists a constant c such that $C(w) \leq |w| + c$ for all $w \in 2^{<\omega}$ (Proposition 2.1.4) Given a string w, one can compute with oracle $\mathbf{0}'$ the set of $p \in 2^{<\omega}$ of length at most $|w| + c$ such that $\mathbb{U}(p) \downarrow = w$. This set is non-empty and the length of its shortest element gives us exactly $C(w)$.

($B \leq_T C$). By Proposition 2.1.11, for any $w \in 2^{<\omega}$, $C(w)$ is the limit of a non-decreasing sequence (hence eventually constant) sequence $C(w)[t]$ where $C(.)[.]$ is computable. Thus, for all n, using C as an oracle, one can find a t_n such that $C(w)[t_n] = C(w)$ for all w of length n. Moreover, for all $t' \geq t(n)$, $C(w)[t'] = C(w)$ for all w of length n. Thus, given $t' \geq t_n$, one can compute $C(w)$ for all w of length n and then find one of complexity at least n (there must exist one by Proposition 2.1.8). Thus, for all $t' \geq t_n$, $C(t') > n - c$ for some constant c independent of n. By definition of B, this implies $B(n) \leq t_{n+c}$ for all n. Hence, for all n:

$$B(n) = \max\{k \leq t_{n+c} : \ C(k) \leq n\}$$

since t_n is computable (uniformly in n) from C, this proves that B is computable from C.

($\mathbf{0}' \leq_T B$). Suppose we know B. Given $e \in \mathbb{N}$ and $q \in 2^{<\omega}$, we want to know whether $M_e(q)$ halts (M_e being the e-th machine in some standard enumeration). Suppose it does; let $t(e, q)$ be the computation time of $M_e(q)$. It is clear that $t(e, q)$ can be computed from e and q hence $C(t(e, p)) \leq C(e, p) \leq |e| + |q| + c \log |e|$ for some constant $c > 0$ indepedent of e, p. Hence, $t(e, p) \leq B(|e| + |q| + c \log |e|)$ by definition of B. Hence, for all e, q, $M_e(q)$ halts if and only if it halts in at most $B(|e| + |q| + c \log |e|)$ steps of computation. Hence, with B as an oracle, one can decide the halting problem.

■

Conditional complexity

In Proposition 2.1.5, we have seen that if we can transform a string u into a string v by a simple computable function f, then one roughly gets $C(u) \leq C(v)$. But our intuition tells us that there is a much stronger link between u and v than a simple inequality on their complexities. Indeed, in this situation, u is *simple when v is known*. This leads to the notion of conditional Kolmogorov complexity.

Definition 2.1.13. *Let $M : 2^{<\omega} \times 2^{<\omega} \to 2^{<\omega}$ be a two-variable machine. The* CONDITIONAL KOLMOGOROV COMPLEXITY OF u KNOWING v, WITH RESPECT TO M, *denoted $C_M(u|v)$ is defined as:*

$$C_M(u|v) = \min\{|p| : \ p \in 2^{<\omega} \ \wedge \ M(p, v) = u\}$$

It is easy to prove, similarly to the unconditional case, that there exists a two-variable machine M such that for all M', for all $u, v \in 2^{<\omega}$, $C_M(u|v) \leq C_{M'}(u|v) + O(1)$. We chose such a machine once for all, calling it $\mathbb{U}_2$, and for all $u, v \in 2^{<\omega}$ we define the *Kolmogorov complexity of u knowing v* as the quantity $C_{\mathbb{U}_2}(u|v)$.

This way, if f is a computable function, we have for all $v \in 2^{<\omega}$: $C(f(v)|v) = O(1)$. Indeed, let M be the two-variable function defined by $M(p, v) = f(v)$. We have for all v: $C_M(f(v)|v) = 0$ as $M(\epsilon, v) = f(v)$. Thus, by additive optimality: $C(f(v)|v) = O(1)$.

The most important theorem regarding conditional complexity is the so-called *theorem of symmetry of information*, which refines Proposition 2.1.6

Theorem 2.1.14 (Symmetry of information, Levin and Kolmogorov [63])**.** *For all $u, v \in 2^{<\omega}$:*

$$C(u, v) = C(u) + C(v|u) + O(\log C(u, v))$$

To prove this theorem, we need the following (simple yet important) lemma. It states that given a finite set of strings A, all elements of A have complexity at most $\log(\#A)$ plus the complexity of A seen as a c.e. set.

Lemma 2.1.15. *Let A be a finite subset of $2^{<\omega}$. Let a be an index of A as a c.e. subset of $2^{<\omega}$. For all $w \in A$, we have $C(w) \log(\#A) + O(C(a))$.*

Proof. Let $w \in A$. To describe w, it suffices to give a and the position i of w within the c.e. enumeration of A. Hence, $C(w) \leq C(i, a) \leq \log(i) + O(C(a)) \leq \log(\#A) + O(C(a))$. □

Proof of Theorem 2.1.14. We first prove $C(u, v) \leq C(u) + C(v|u) + O(\log C(u, v))$. The proof is very similar to the proof of Proposition of 2.1.6. Let M be the machine of domain $\{0^k 1 r : k \in \mathbb{N}, r \in 2^{<\omega} \text{ and } |r| \geq k\}$ which does the following. On input $0^k 1 r$, M splits r into two parts: $r = pq$ with $|p| = k$. It computes $l = \text{Bin}^{-1}(p)$, and splits q into two parts $q = q'q''$ with $|q| = l$. Then, it computes $w = \mathbb{U}(q')$, and then $w' = \mathbb{U}_2(q'', w)$. Finally, it outputs $\langle w, w' \rangle$. Let $u, v \in 2^{<\omega}$. Let q' be a shortest program for u and q'' a shortest program for v (i.e. $\mathbb{U}(q') = u$, $\mathbb{U}_2(q'', u) = v$, $C(u) = |q'|$, and $C(v|u) = |q''|$). Then, on input $0^{|\text{Bin}(|q'|)|} 1 \text{Bin}(|q'|) q' q''$, the machine M exactly outputs $\langle u, v \rangle$. Thus:

$$\begin{aligned} C(u, v) \leq C_M(u, v) + O(1) &\leq 2\left|\text{Bin}\left(|q'|\right)\right| + |q'| + |q''| + O(1) \\ &\leq C(u) + C(v|u) + O(\log C(u)) \\ &\leq C(u) + C(v|u) + O(\log C(u, v)) \end{aligned}$$

We now prove the more difficult part: $C(u, v) \geq C(u) + C(v|u) + O(\log C(u, v))$. Let $u, v \in 2^{<\omega}$. Set $k_0 = C(u, v)$ and $k_1 = C(u)$. Let A be the set

$$A = \{(w, w') \in 2^{<\omega} \times 2^{<\omega} : \ C(w, w') \leq k_0\}$$

Notice that $\#A \leq 2^{k_0+1}$, and that an index for A as a c.e. set can be computed from k_0. Notice also that by definition, $(u, v) \in A$. For all $w \in 2^{<\omega}$, set $A_w = \{w' \in 2^{<\omega} : (w, w') \in A\}$, and set $l = \log(\#A_u)$. Then, set $B = \{w \in 2^{<\omega} : \#A_w \geq 2^l\}$. An index for B a c.e. set can be computed from k_0 and l. Also, by definition:

$$2^l \#B \leq \#A \quad \text{hence} \quad \#B \leq 2^{k_0+1-l}$$

Since $u \in B$, we have, by Lemma 2.1.15:

$$C(u) \leq \log(\#B) + O(C(k_0, l)) \leq k_0 + 1 - l + O(\log k_0 + \log l) \leq k_0 - l + O(\log k_0)$$

(the last inequality comes from the fact that $l = \log(\#A_u) \leq \log(\#A) \leq k_0 + 1$). Hence:

$$l \leq k_0 - C(u) + O(\log k_0) = k_0 - k_1 + O(\log k_0) \tag{2.1}$$

Now, suppose we know k_0, u, and $\#A_u$. To find v, it suffices to enumerate A until we find all the w' such that $(u, w') \in A$ i.e. all the w' in A_u (we know when to stop because we know the cardinal of A_u), and then give the position of v among the elements of A_u. If u is known, an index for A_u can be found from an index for A, which itself can be computed from k_0. Hence, applying Lemma 2.1.15 with condition u:

$$C(v|u) \leq \log(\#A_u) + O(\log k_0) \leq l + O(\log k_0) \leq k_0 - k_1 + O(\log k_0)$$

(the second inequality follows from (2.1)). By definition of k_0 and k_1:

$$C(v|u) \leq C(u, v) - C(u) + O(\log C(u, v))$$

■

2.1.2 Prefix-free Kolmogorov complexity

In the proof of Proposition 2.1.6, we explained that given two programs p and q for a machine M which output respectively u and v, the simple concatenation $r = pq$ does not contain enough information to compute the pair (u, v). To do so, some extra information on where to split r is necessary. Suppose now that M has the particular property to have a prefix-free domain: given two programs $p \sqsubset p'$ for M, at most one of them halts. In this case, $r = pq$ can be decoded without extra information: it suffices to consider all the possible splittings $r = p'q'$ of r and run in parallel all the computations of $(M(p'), M(q'))$. At most one of them can halt, since otherwise this would contradict the prefix-freeness of the domain of M. Hence, the programs for M are "auto-delimited": any concatenation of such programs can be decoded without extra information. Restricting our attention to this particular kind of machines, we can define the notion of *prefix-free Kolmogorov complexity*.

Definition 2.1.16. *A machine $M : 2^{<\omega} \to 2^{<\omega}$ is said to be* PREFIX-FREE *if its domain is prefix-free. For a prefix-free machine M, the Kolmogorov complexity with respect to M is denoted by K_M instead of C_M.*

It is easy to see that prefix-free machines can effectively be enumerated. Indeed, given an index for a machine, one can enumerate its graph, stopping the enumeration if at some point adding a new $(p, M(p))$ to the graph contradicts the prefix-freeness property (the enumeration is stopped before this happens). Hence, using the same argument as for plain complexity, we get the existence of an additively optimal machine among those that are prefix-free:

Theorem 2.1.17 (Additive Optimality for K)**.** *There exists a prefix-free machine $\mathbb{V}$ such that for every prefix-free machine M, $K_{\mathbb{V}} \leq K_M + O(1)$*

We fix an additive machine $\mathbb{V}$ once for all, and we set $K = K_{\mathbb{V}}$. For all $w \in 2^{<\omega}$, $K(w)$ is called PREFIX KOLMOGOROV COMPLEXITY of w.

Since a prefix-free machine is a particular kind of machine, we have by the additive optimality theorem for C:

$$C(w) \leq C_{\mathbb{V}}(w) + O(1) = K(w) + O(1)$$

for all $w \in 2^{<\omega}$. Hence the prefix-free complexity is always bigger (up to an additive constant) than the plain one. However, not much bigger:

Lemma 2.1.18. *For all $w \in 2^{<\omega}$, $K(w) \leq C(w) + 2\log(C(w)) + O(1)$*

Proof. Let M be the machine whose domain is contained in $\{0^k 1r : k \in \mathbb{N}, r \in 2^{<\omega} \text{ and } |r| \geq k\}$ and which does the following. On an input $0^k 1r$, it splits r into two parts $r = pq$ with $|p| = k$. Then, it computes $l = \mathrm{Bin}^{-1}(p)$. If $|q| = l$, it outputs $\mathbb{U}(q)$ (if defined). If not, $M(0^k 1r)$ is undefined. The machine M is prefix-free: if it halts on two programs $0^k 1r \sqsubseteq 0^{k'} 1r'$, then one must have $k = k'$, hence $r \sqsubseteq r'$. Moreover, M splits r into $r = pq$ with $|p| = k$. Hence $r' = pq'$ with $q \sqsubseteq q'$. But by definition of M, the fact that M halts on $0^k 1r$ (resp.$0^{k'} 1r'$) means that $|q| = \mathrm{Bin}^{-1}(p)$ (resp. $|q'| = \mathrm{Bin}^{-1}(p)$). Hence q is a prefix of q' and has the same length, hence $q = q'$. We have proven that $0^k 1r = 0^{k'} 1r'$.

Now, let $w \in 2^{<\omega}$ and q a shortest program for $\mathbb{U}$ which outputs w. Let $p = \mathrm{Bin}(|q|)$ (this implies $|p| = \log(|q|) + O(1)$) and $k = |p|$. Then by definition $M(0^k 1pq) = \mathbb{U}(q) = w$. Thus:

$$\begin{aligned} K_M(w) &\leq k + 1 + |p| + |q| \\ &\leq 2|p| + |q| + 1 \\ &\leq |q| + 2\log|q| + O(1) \\ &\leq C(w) + 2\log C(w) + O(1) \end{aligned}$$

■

Prefix Kolmogorov complexity shares many properties with C, including the following:

Theorem 2.1.19. *K is not computable, its Turing degree is $\mathbf{0}'$.*
K is enumerable from above, i.e. there exists a computable function $K(.)[.]$ such that for all $w \in 2^{<\omega}$, $K(w)$ is the limit of the nonincreasing sequence (hence eventually constant) $(K(w)[t])_{t\in\mathbb{N}}$.

Proof. The proof is the same as for C. ■

Proposition 2.1.20. *Let $f : 2^{<\omega} \to 2^{<\omega}$ be a computable function. For all w, $K(f(w)) \leq K(w) + O(1)$.*

Proof. The proof is the same as for C. ■

Conditional prefix complexity

We can also define conditional prefix complexity similarly to the plain complexity case. It suffices to consider the class of two-variable machines M having the property that for all v, the set of $\{p : (p, v) \in dom(M)\}$ is prefix-free. Then taking an additively optimal machine $\mathbb{V}_2$ among those, we can set

$$K(u|v) = \min\{|p| : \mathbb{V}_2(p, v) = u\}$$

As C and K are equal up to a logarithmic factor (Lemma 2.1.18), one can rewrite the symmetry of information theorem (Theorem 2.1.14):

Theorem 2.1.21 (Symmetry of information)**.** *For all $u, v \in 2^{<\omega}$:*

$$K(u, v) = K(u) + K(v|u) + O(\log K(u, v))$$

In fact, a much more precise theorem holds for K: for all $u, v \in 2^{<\omega}$, $K(u, v) = K(u) + K(v|u, K(u)) + O(1)$ (see Gacs [18]). However, Theorem 2.1.21 will be sufficient for our purposes.

Kraft-Chaitin's theorem

Dealing with prefix Kolmogorov complexity can seem a little tedious. For plain complexity, when we want to give an upper bound on $C(w)$, we typically construct a particular machine M and estimate the complexity $C_M(w)$ (see section on plain complexity). For prefix complexity, the same approach requires to build a machine M which is additionally prefix-free, a property that is not so easy to check. There is a way to avoid this problem, using the so-called *Kraft-Chaitin theorem*, which we now present.

In information and coding theory, Kraft's theorem (first proven in [30]) asserts that given a prefix-free set of strings $\{p_i : i \in \mathbb{N}\}$, it holds that $\sum_i 2^{-|p_i|} \leq 1$. This is easy to see as $\sum_i 2^{-|p_i|}$ is the measure of the open set generated by the p_i. The Kraft-Chaitin theorem tells us that conversely, given a subset $\{n_i : i \in \mathbb{N}\}$ of $\mathbb{N}$ such that $\sum_i 2^{-n_i} \leq 1$, there exists a prefix-free set of words $\{p_i : i \in \mathbb{N}\}$ such that $|p_i| = n_i$ for all i and the p_i can be found effectively given the n_i:

Theorem 2.1.22 (Kraft-Chaitin theorem; Levin [34], Chaitin [13]). *Let $(n_i)_{i\in\mathbb{N}}$ be a computable sequence of nonnegative integers, such that $\sum_{i\in\mathbb{N}} 2^{-n_i} \leq 1$. There exists a computable sequence of strings $(p_i)_{i\in\mathbb{N}}$ such that the set of p_i is prefix-free and for all i: $|p_i| = n_i$.*

Proof. This theorem can be seen as a two-player infinite game. Player (I) will play the role of the sequence $(n_i)_{i\in\mathbb{N}}$ and player (II) the role of the sequence $(p_i)_{i\in\mathbb{N}}$. At stage s of the game, player (I) plays an integer n_s, and player (II) answers by a string p_s. There are constraints on this game: for all s, one must have $\sum_{i=0}^{s} 2^{-n_i} \leq 1$ and the set $\{p_i : 0 \leq i \leq s\}$ must be prefix-free. Player (I) wins the game if he plays according these rules and at some stage Player (II) has to break them. Player (II) wins if he can keep the game going forever without breaking the rules. We claim that Player (II) has a computable winning strategy (which proves the theorem). It works as follows. Suppose that at the beginning of stage s, (II) has already computed a finite list of cylinders

$$L_s = [u_0^s], [u_1^s], \ldots, [u_{k_s}^s]$$

satisfying the following properties:

(a) $|u_0^s| < |u_1^s| < \ldots < |u_{k_s}^s|$

(b) the cylinders $[u_0^s], \ldots, [u_{k_s}^s], [p_0], \ldots, [p_{s-1}]$ are pairwise disjoint.

(c) $\sum_{j=0}^{k_s} 2^{-|u_j^s|} = 1 - \sum_{i=0}^{s-1} 2^{-|p_i|} = 1 - \sum_{i=0}^{s-1} 2^{-n_i}$

(initially, $L_0 = \emptyset$). Now, at stage s, (I) plays an integer n_s. We distinguish two cases.

Case 1. There exists a j such that $|u_j^s| = n_s$. In this case, (II) plays $p_s = u_j^s$ and remove the cylinder $[u_j^s]$ from the list L_s to get L_{s+1}. It is clear that the properties (a), (b) and (c) are preserved by this operation.

Case 2. Otherwise, let $j \leq 0$ be the biggest integer such that $|u_l^s| \leq n_s$. Note that there must exist such a l. If not, this would mean that n_s is stricly smaller than all $|u_j^s|$, hence:

$$2^{-n_s} > \sum_{j=0}^{k_s} 2^{-|u_j^s|} = 1 - \sum_{i=0}^{s-1} 2^{-n_i}$$

which implies $\sum_{i=0}^{s} 2^{-n_i} > 1$, i.e. (I) would have broken the rules.

In this case, (II) plays $p_s = u_l^s 0^{n_s - |u_l^s|}$, removes u_l^s from L_s and replaces it by the list of cylinders

$$L' = \{[u_l^s 0^m 1] : 1 \leq m \leq n_s - |u_l^s| - 1\}$$

to get L_{s+1}. Notice that the above list L' of intervals, together with $[p_s]$, form a partition of the cylinder $[u_l^s]$, hence the properties (b) and (c) are preserved. Moreover, (a) is also preserved because the intervals in L' are of increasing length, and

by definition of l, there is no u_j^s whose length is between $|u_l^s|+1$ and n_s.

Since the property (b) is preserved at all stages and implies in particular that $p_0, \dots, p_s$ is a prefix-free list of strings for all s, this is a winning strategy for (II). Morevover, all the steps of the strategy are effective.

■

This theorem is particularly useful, as it allows us to build prefix-free machines implicitly:

Corollary 2.1.23. *Let $\big((w_i, n_i)\big)_{i\in\mathbb{N}}$ be computable sequence of elements of $2^{<\omega}\times\mathbb{N}$ such that $\sum_i 2^{-n_i} \leq 1$. There exists a computable sequence of strings $(p_i)_{i\in\mathbb{N}}$ such that the p_i form a prefix-free set, and a machine M with domain $\{p_i : i \in \mathbb{N}\}$ such that for all i, $|p_i| = n_i$ and $M(p_i) = w_i$.*

Proof. Since the n_i form a computable sequence, by the Kraft-Chaitin theorem, there exists a computable sequence $(p_i)_{i\in\mathbb{N}}$ such that p_i form a prefix-free set and for all i: $|p_i| = n_i$. Let M be the machine which works as follows. On an input $p \in 2^{<\omega}$, M enumerates the set $\{p_j : j \in \mathbb{N}\}$ until it finds an i such that $p_i = p$ and then outputs w_i (if no such i is found, $M(p)$ is undefined). ■

Hence, when given a computable sequence $\{(w_i, n_i) : i \in \mathbb{N}\}$ such that $\sum_i 2^{-n_i} \leq 1$, if one is interested in the existence of a prefix-free machine M such that $K_M(w_i) \leq n_i$ for all i, but does not care about the exact behaviour of M, one can invoke the Kraft-Chaitin theorem to get the machine M, without having to give a precise specification for it. From Corollary 2.1.23, one easily gets:

Corollary 2.1.24. *Let L be a c.e. subset of $2^{<\omega}\times\mathbb{N}$ such that*

$$\sum_{(w,n)\in L} 2^{-n} < +\infty$$

Then for all $(w, n) \in L$: $K(w) \leq n + O(1)$.

Proof. Let $i \mapsto (w_i, n_i)$ be a computable enumeration of L. Let c be an integer such that $\sum_i 2^{-n_i} < 2^c$. This means $\sum_i 2^{-(n_i+c)} \leq 1$. Hence, applying Corollary 2.1.23 to $\{(w_i, n_i + c) : i \in \mathbb{N}\}_{i\in\mathbb{N}}$, there exists a prefix-free machine M such that $K_M(w_i) \leq n_i + c$ for all i. By additive optimality, $K(w_i) \leq n_i + O(1)$ for all i. ■

We call Kraft-Chaitin set any c.e. set L satisfying the conditions of Corollary 2.1.24.

Using Kraft-Chaitin theorem, we can prove the following (somewhat surprising) result: among all the sequences $(a_n)_{n\in\mathbb{N}}$ of real numbers that are uniformly left-c.e. (that is, the function $n \mapsto a_n$ is left-c.e) and such that $\sum_n a_n < +\infty$, there are some that majorize all the others up to a multiplicative constant, and $(2^{-K(n)})_{n\in\mathbb{N}}$ is an example of such a sequence.

Theorem 2.1.25 (Levin).
(i) The sequence $(2^{-K(n)})_{n\in\mathbb{N}}$ *is uniformly left-c.e. and* $\sum_{n\in\mathbb{N}} 2^{-K(n)} \leq 1$.
(ii) For every uniformly left-c.e sequence $(a_n)_{n\in\mathbb{N}}$ *of real numbers such that* $\sum_n a_n < +\infty$, *one has* $a_n = O(2^{-K(n)})$ *(i.e.* $K(n) \leq -\log(a_n) + O(1)$ *for all* n*).*

Proof. (i) One has $\sum_n 2^{-K(n)} = \sum_n 2^{-K(\mathrm{Bin}(n))} \leq \sum_w 2^{-K(w)}$. But for all w, $K(w)$ is the length of some program p of $\mathbb{V}$, the additively optimal machine for prefix complexity (namely, p is the shortest element of $dom(\mathbb{V})$ such that $\mathbb{V}(p) = w$). Hence: $\sum_w 2^{-K(w)} \leq \sum_{p\in dom(\mathbb{V})} 2^{-|p|}$. But $dom(\mathbb{V})$ is a prefix-free set by definition of prefix-free machines. Hence, by Kraft's theorem, $\sum_{p\in dom(\mathbb{V})} 2^{-|p|} \leq 1$.
(ii) Let $a : \mathbb{N}\times\mathbb{N} \to \mathbb{R}$ be a witness that $(a_n)_{n\in\mathbb{N}}$ is uniformly left-c.e. that is, a is computable and for all n, the sequence of reals $(a_n(t))_{t\in\mathbb{N}}$ is non-decreasing and converges to a_n. Set for all n, t, $b_n(t) = \lfloor -\log a_n(t)\rfloor$ and $b_n = \lfloor -\log a_n\rfloor$. It is easy to see that the sequence of reals $(2^{-b_n})_{n\in\mathbb{N}}$ is uniformly left-c.e. (witnessed by $\lim_{t\to+\infty} 2^{-b_n(t)} = 2^{-b_n}$, the function being non-decreasing) and that $\sum 2^{-b_n} < +\infty$. We construct a Kraft-Chaitin set L as follows: first enumerate in L all the couples $(\mathrm{Bin}(n), b_n(0))$. Then, each time ones finds n, t such that $b_n(t) < b_n(t-1)$ (i.e. when the approximation of b_n gets better), one enumerates $(\mathrm{Bin}(n), b_n(t))$ into L. Let us check that L is indeed a Kraft-Chaitin set. Clearly, L is computably enumerable. Moreove, notice that for all n, the elements $(\mathrm{Bin}(n), k)$ that are enumerated into L all satisfy $k \geq b_n$ since they are all of type $k = b_n(t)$ for some t, and $b_n(t) \geq b_n$ for all t. Notice also that for all n, $(\mathrm{Bin}(n), b_n) \in L$, as $b_n(t)$ is a nonincreasing sequence of integers that converges to b_n, hence reaches b_n at some point. Thus:

$$\begin{aligned}\sum_{\substack{n,k\\(\mathrm{Bin}(n),k)\in L}} 2^{-k} &\leq \sum_{n\in\mathbb{N}}\sum_{k\geq b_n} 2^{-k}\\ &\leq \sum_{n\in\mathbb{N}} 2^{-b_n+1}\\ &< +\infty\end{aligned}$$

Hence by Corollary 2.1.24, for all $n \in \mathbb{N}$, since $(\mathrm{Bin}(n), b_n) \in L$, we have $K(n) = K(\mathrm{Bin}(n)) \leq b_n + O(1)$. Hence $a_n = O(2^{-b_n}) = O(2^{-K(n)})$. ∎

Remark 2.1.26. *The same argument shows that if* $f : 2^{<\omega} \to \mathbb{R}$ *is a left-c.e. function such that* $\sum_{w\in 2^{<\omega}} f(w) < +\infty$, *then for all* w*:* $K(w) \leq -\log f(w) + O(1)$ *for all* w *(this is just because we can identify* $2^{<\omega}$ *to* $\mathbb{N}$*).*

The K-counting theorem

We have seen that a string of length n has plain complexity at most $n + O(1)$, and that most strings have a complexity close to this value (see Proposition 2.1.8) i.e. most strings are k-incompressible$_C$ with a small k. How high can be the prefix complexity of a string of length n? We already know that it is at most $n + 2\log n + O(1)$ (see Lemma 2.1.18). The precise bound is: $n + K(n) + O(1)$, and most strings of length n have a prefix complexity that is close to this value.

Proposition 2.1.27 (K-Counting theorem)**.**
(i) For all $w \in 2^{<\omega}$, $K(w) \leq |w| + K(|w|) + O(1)$.
(ii) $\#\{w \in 2^{<\omega} : |w| = n \wedge K(w) \leq n + K(n) - k\} = O(2^{n-k})$

Proof. (i) Let $f : 2^{<\omega} \to \mathbb{R}$ defined by $f(w) = 2^{-|w|-K(|w|)}$. It is left-c.e. since K is enumerable from above. Moreover:

$$\begin{aligned}\sum_{w\in 2^{<\omega}} f(w) &= \sum_{w\in 2^{<\omega}} 2^{-|w|-K(|w|)} \\ &= \sum_{n}\sum_{|w|=n} 2^{-n-K(n)} \\ &= \sum_{n} 2^n\, 2^{-n-K(n)} \\ &\leq 1\end{aligned}$$

(ii) We will prove the equivalent statement:

$$\#\{w \in 2^{<\omega} : |w| = n \wedge 2^{-K(w)} \geq 2^{n-K(n)-k}\} = O(2^{n-k})$$

For all n, set $a_n = \sum_{|w|=n} 2^{-K(w)}$. It is clear that a_n is uniformly left-c.e as K is enumerable from above. By Theorem 2.1.25(i), it holds that $\sum_n a_n \leq 1$. Hence, by Theorem 2.1.25(ii), there exists a constant c such that $a_n \leq 2^{-K(n)+c}$ for all n. Thus, for all n

$$\#\{w \in 2^{<\omega} : |w| = n \wedge 2^{-K(w)} \geq 2^{n-K(n)-k}\} \leq 2^{n-k+c}$$

since, if it is not the case, then

$$\begin{aligned}a_n &= \sum_{|w|=n} 2^{-K(w)} \\ &> 2^{n-k+c}\, 2^{n-K(n)-k} \\ &> 2^{-K(n)+c}\end{aligned}$$

which contradicts $a_n \leq 2^{-K(n)+c}$.

■

2.2 Characterizing (or not) infinite random sequences via Kolmogorov complexity

At this stage, we have introduced various notions of randomness for infinite sequences, and a measure of complexity/randomness for finite ones (Kolmogorov complexity). The immediate question that comes to mind is how these notions are related, i.e. if the defintions of randomness for infinite sequences can be expressed in terms of Kolmogorov complexity. This is what we shall discuss in this section. As we will see, Martin-Löf randomness has very elegant characterizations in terms of Kolmogorov complexity (both plain and prefix-free) while the situation is more complicated for other notions, like computable or Schnorr randomness. This is yet another reason for the popularity of Martin-Löf randomness.

2.2.1 Martin-Löf randomness vs Kolmogorov complexity

Martin-Löf's oscillation theorems

Knowing the concept of Kolmogorov complexity, a first attempt to define randomness for infinite sequences could be the following: "an infinite sequence α is random if all the strings $\alpha_{(i)}, \alpha_{(i+1)}, ..., \alpha_{(j)}$ it contains are c-incompressible for some c (which does not depend on i, j)". But this is obviously too naive as a truly random sequence should contain arbitrarily long substrings of zeroes (in particular, a Martin-Löf random sequence does). One could refine this idea by requiring the sequence α to have maximal plain Kolmogorov complexity on all its initial segment, i.e., formally: there exists some $c > 0$ such that $C(\alpha\restriction_n) \geq n - c$ for all n. However, this is still too naive, as there exists no such sequence!

Proposition 2.2.1 (Martin-Löf [42]). *Let $\alpha \in 2^\omega$. There exist infinitely many n such that $C(\alpha\restriction_n) \leq n - \log n$.*

Proof. Let $\alpha \in 2^\omega$ be a fixed sequence. Let n be fixed, and let $w = \alpha\restriction_n$ be the prefix of α of size n. We can see w as the binary writing of some natural number, hence let k be such that $w = \text{Bin}(k)$. Remember that $|w| = \log(k+1)$. Now, consider the prefix $\alpha\restriction_k$ of α. We claim that $\alpha_{(n+1)}...\alpha_{(k)}$ is enough to reconstruct the whole $\alpha\restriction_k$ (up to a finite amount of additional information). Indeed, from $\alpha_{(n)}...\alpha_{(k)}$, one can compute its length, which is equal to $k - n = k - \log(k+1)$. But the function $k \mapsto k - \log(k+1)$ is almost one-to-one, in the sense that every element of $\mathbb{N}$ has at most 2 pre-images. Hence, k can be retrived from $k - \log(k+1)$ together with another bit of information (in the case there are two possible pre-images). Now, from k, one can compute w since Bin is a computable bijection. It only remains to concatenate w and $\alpha_{(n)}...\alpha_{(k)}$ to get $\alpha\restriction_k$. All this means that

$$C(\alpha\restriction_k \, |\alpha\restriction_{[n,k-1]}) \leq O(1)$$

and hence

$$C(\alpha\restriction_k) \leq C(\alpha\restriction_{[n,k-1]}) + O(1) \leq k - \log(k+1) + O(1)$$

■

In fact, Martin-Löf proved a much more general result:

Theorem 2.2.2 (Martin-Löf [42]). *Let $f : \mathbb{N} \to \mathbb{N}$ be a computable function such that $\sum_{n\in\mathbb{N}} 2^{-f(n)} = +\infty$. Then there are infinitely many n for which $C(\alpha\restriction_n |n) \leq n - f(n)$. Moreover, if the function f is such that $C(n|n - f(n)) \leq O(1)$ then for infinitely many n's, $C(\alpha\restriction_n) \leq n - f(n)$.*

We omit the proof of this theorem.

Since no sequence can have maximal plain complexity on all its initial segments, we can try to define a notion of randomness by weakening this naive requirement. We can do this in different ways:

- We could use K in place of C, requiring $K(\alpha\restriction_n) \geq n - O(1)$ for all n. This indeed weakens the condition as K is greater than C.
- We could weaken the maximality condition by requiring $C(\alpha\restriction_n)$ to be big enough for all n, say greater than $n - f(n)$ for some well-chosen function f, computable or not.
- We could require $C(\alpha\restriction_n)$ to be maximal not all the time but only infinitely often.

We will discuss these three approaches.

The Levin-Chaitin theorem

The first of the three ideas we just mentioned does work.

Proposition 2.2.3. *The class* $\mathcal{R} = \{\alpha \in 2^\omega : \ K(\alpha\restriction_n) \geq n - O(1)\}$ *has Lebesgue measure* 1.

Proof. For all $k \in \mathbb{N}$, let $\mathcal{A}_k = \{\alpha \in 2^\omega : \ (\exists n)\ K(\alpha\restriction_n) \leq n - k\}$. A sequence is *not* in $\mathcal{R}$ if and only if it belongs to all $\mathcal{A}_k$. We need to show that the intersection of all A_k has measure 0. In fact:

Lemma 2.2.4. *For all* k*, the measure of the set*

$$\mathcal{A}_k = \{\alpha \in 2^\omega : \ (\exists n)\ K(\alpha\restriction_n) \leq n - k\}$$

is at most 2^{-k}*.*

Subproof. Fix $k \in \mathbb{N}$. Let $\mathcal{A}_k$ is an open set, generated by the set of strings

$$A_k = \{w \in 2^{<\omega} : \ K(w) \leq |w| - k\}$$

Let A'_k be the set of minimal elements of A_k (which as usual ensures that A'_k is prefix-free and $[A'_k] = [A_k]$). For all $w \in A'_k$, let p_w be the shortest program p such that $\mathbb{V}(p) = w$ (by definition of K, this implies $K(w) = |p_w|$). We then have

$$\begin{aligned} \lambda(\mathcal{A}_k) &= \sum_{w \in A'_k} 2^{-|w|} \\ &\leq \sum_{w \in A'_k} 2^{-K(w)-k} \qquad \text{(by definition of } A_k\text{)} \\ &\leq 2^{-k} \sum_{w \in A'_k} 2^{-|p_w|} \end{aligned}$$

Moreover, since the p_w are in the domain of $\mathbb{V}$ they form a prefix-free set which implies $\sum_{w \in A'_k} 2^{-|p_w|} \leq 1$. □

The theorem follows immediately from this lemma.

■

In the above proof, it is easy to see that the sequence $\{A_k : k \in \mathbb{N}\}$ is a computable sequence of c.e. open sets. And since there is a constant c such that $\lambda(A_k) \leq c\,2^{-k}$ for all k, they in fact form a Martin-Löf test. This means that any Martin-Löf random sequence α must satisfy $K(\alpha\restriction_n) \geq n - O(1)$ for all n. In fact, this characterizes Martin-Löf random sequences!

Theorem 2.2.5 (Schnorr [53], Levin [35]). *A sequence $\alpha \in 2^\omega$ is Martin-Löf random if and only if $K(\alpha\restriction_n) \geq n - O(1)$.*

Proof. By the above discussion, it only remains to prove that if α is not Martin-Löf random, then it is not true that $K(\alpha\restriction_n) \geq n - O(1)$ for all n. Let $\{\mathcal{U}_n : n \in \mathbb{N}\}$ be a Martin-Löf test that covers α. We can assume by Lemma 1.3.2 that $\lambda(\mathcal{U}_n) \leq 2^{-2n-1}$ for all n. We can also assume, by Remark 1.2.6, that for all n: $\mathcal{U}_n = \bigcup_i [u_i^{(n)}]$ where the set $\{u_i^{(n)} : i \in \mathbb{N}\}$ is c.e. uniformly in n. Hence, $\sum_i 2^{-|u_i^{(n)}|} = \lambda(\mathcal{U}_n) \leq 2^{-2n-1}$. We claim that the set $\left(u_i^{(n)}, |u_i^{(n)}| - n\right)_{i,n\in\mathbb{N}}$ is a Kraft-Chaitin set. Indeed, it is computably enumerable and:

$$\begin{aligned}\sum_{i,n} 2^{-|u_i^{(n)}|+n} &= \sum_n 2^n \sum_i 2^{-|u_i^{(n)}|} \\ &\leq \sum_n 2^n\, 2^{-2n-1} \\ &\leq 1\end{aligned}$$

Hence, by Corollary 2.1.24, $K(u_i^{(n)}) \leq |u_i^{(n)}| - n + O(1)$ for all n, i. But α is covered by $(\mathcal{U}_n)_{n\in\mathbb{N}}$, hence for all n, there exists i such that $u_i^{(n)} \sqsubseteq \alpha$ i.e. $u_i^{(n)} = \alpha\restriction_k$ for some k. For such an k, one has $K(\alpha\restriction_k) \leq k - n - O(1)$. Since we can do this for arbitrarily large n, this proves that $K(\alpha\restriction_k) \geq k - O(1)$ is not true. ∎

The Miller-Yu theorem

Despite Martin-Löf's oscillation theorem (Theorem 2.2.2), is there still hope that Martin-Löf randomness could be characterized by plain Kolmogorov complexity? This was a longstanding open question, recently answered affirmatively by Miller and Yu.

Theorem 2.2.6 (Miller and Yu [49]). *The following are equivalent for every sequence $\alpha \in 2^\omega$:*
(a) α is Martin-Löf random
(b) $C(\alpha\restriction_n) \geq n - K(n) - O(1)$
(c) For every computable function f such that $\sum_n 2^{-f(n)} < +\infty$, $C(\alpha\restriction_n) \geq n - f(n) - O(1)$.

Remark 2.2.7. *The equivalence of (a) and (b) was proven earlier in [19].*

Proof. $(a) \Rightarrow (b)$: if α is Martin-Löf random random, then by Theorem 2.2.5, $K(\alpha\restriction_n) \geq n - O(1)$. Moreover, by Proposition 2.1.27, $K(\alpha\restriction_n) \leq C(\alpha\restriction_n) + K(n) + O(1)$. It follows that $C(\alpha\restriction_n) + K(n) + O(1) \geq n - O(1)$, hence the result.

$(b) \Rightarrow (c)$: Suppose (b) does not hold. Then for infinitely many n, $C(\alpha\restriction_n) \leq n - K(n)$. For an given n, the set

$$\mathcal{U}_n = [\{w : |w| = n \wedge C(w) \leq n - K(n)\}]$$

is a c.e. open set that contains α and its measure is $O(2^{-n})$ by Proposition 2.1.8. Since $\sum_n 2^{-K(n)} < +\infty$, we can apply Theorem 1.3.4 and we get that α is not Martin-Löf random.

$(c) \Rightarrow (a)$: this is the hard part of the theorem. We will provide a proof in Section 2.3, using computable upper bounds of Kolmogorov complexity. For an elementary proof that requires only very basic knowledge of Kolmogorov complexity, see [11].

■

Kolmogorov randomness and strong Chaitin randomness

Another way to sidestep Martin-Löf's oscillation theorem is to require that a random sequence has maximal plain complexity not for all n but only for infinitely many. This leads to the notion of Kolmogorov-randomness.

Definition 2.2.8. *A sequence $\alpha \in 2^\omega$ is* Kolmogorov random *if $C(\alpha\restriction_n) \geq n - O(1)$ for infinitely many n.*

It is not immediately clear why this is a notion of randomness, i.e. why the class of Kolmogorov random sequences actually has measure 1. This comes from the following classical lemma.

Lemma 2.2.9 (Fatou's lemma)**.** *Let $\varepsilon \in [0,1]$. Let $(\mathcal{A}_n)_{n\in\mathbb{N}}$ be a family of subsets of 2^ω infinitely many of which have measure at most ε. Then the measure of $\liminf_n(\mathcal{A}_n)$ is at most ε, where $\liminf_n(\mathcal{A}_n)$ denotes the class of those α that belong to all but finitely many $\mathcal{A}_n$.*

Subproof. This is because

$$\liminf_n(\mathcal{A}_n) = \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} \mathcal{A}_k$$

Set for all n $\mathcal{B}_n = \bigcap_{k\geq n} \mathcal{A}_k$. The $\mathcal{B}_n$ form a non-decreasing sequence of sets each of which has measure at most ε (this because, by the hypothesis, for all k there exists $k' \geq k$ such that the measure of $\mathcal{A}_{k'}$ is at most ε and $\mathcal{B}_k \subseteq \mathcal{A}_{k'}$ by definition) and whose union is $\liminf_n(\mathcal{A}_n)$. Hence, $\lambda\left(\liminf_n(\mathcal{A}_n)\right) = \lim_n \lambda(\mathcal{B}_n) \leq \varepsilon$ □

Proposition 2.2.10. *The class of Kolmogorov random sequences has measure 1.*

Proof. If α is not Kolmogorov random, then for all c, for almost all n, $C(\alpha\restriction_n) < n - c$. This is equivalent to say that the complement of the class of Kolmogorov random sequences is exactly

$$\bigcap_{c\in\mathbb{N}} \liminf_n (\mathcal{A}_n^c)$$

where $\mathcal{A}_n^c$ is the set of sequences α such that $C(\alpha\restriction_n) < n - c$. For every c, n, the set $\mathcal{A}_n^c$ has measure at most 2^{-c}; this is because $\mathcal{A}_n^c$ is an open set generated by the $w \in 2^{<\omega}$ of length n such that $C(w) \leq n - c$ and by the C-counting theorem (Proposition 2.1.8), there are at most 2^{n-c} such strings. Hence, by Lemma 2.2.9, the measure of $\liminf_n(\mathcal{A}_n^c)$ is at most 2^{-c} for all c. This, by the above formula, proves that the class of non-Kolmogorov random sequences has measure 0. ■

It turns out that Kolmogorov randomness is stronger than Martin-Löf randomness. It is in fact equal to $\mathbf{0}'$-Martin-Löf randomness, which is the class of sequences we obtain if we relativize the definition of Martin-Löf to $\mathbf{0}'$. Adding computational power allows more Martin-Löf tests, hence the notion of $\mathbf{0}'$-Martin-Löf randomness is stronger than Martin-Löf randomness. This was proven independently by Miller [48] and Nies, Stephan and Terwijn [51].

Theorem 2.2.11 (Miller [48], Nies et al. [51])**.** *A sequence $\alpha \in 2^\omega$ is Kolmogorov random if and only if it is $\mathbf{0}'$-Martin-Löf random.*

Proof of $\Rightarrow$. Assume that a sequence $\alpha \in 2^\omega$ is not $\mathbf{0}'$-Martin-Löf random. This means that for all k, α is covered by a $\mathbf{0}'$-c.e. open set $\mathcal{U}_k$ of measure at most 2^{-k}. We fix a k for now and we carry out the proof uniformly in k. Let A_k be a $\mathbf{0}'$-c.e. prefix-free set of strings such that $\mathcal{U}_k = [A_k]$. By a classical characterization of $\mathbf{0}'$-computability, A_k is limit computable i.e. it can be written as the pointwise limit over t of finite prefix-free subsets $A_k(t)$ uniformly computable in k, t (we can moreover assume that the measure of $[A_k(t)]$ is less than 2^{-k} for all t). Since $\alpha \in [A_k]$, $\alpha \in [A_k(t)]$ for all t large enough, say greater than some n. Since $\alpha \in A_k(n)$, we have $[\alpha\restriction_t] \subseteq [A_k(t)]$ for all $t \geq n$. Since $[A_k(t)]$ has measure at most 2^{-k}, there are at most 2^{t-k} strings w of length t such that $[w] \subseteq [A_k(n)]$. Hence, for all $t \geq n$, $\alpha\restriction_t$ can be described by its index N in this set of strings (which written in binary has length $t - k$) and k. Note that this is possible only because $A_k(t)$ is computable in k, t, and that there is no need to give t in the description as it can be retrieved from N and k. Thus, for all $t \geq N$, $C(\alpha\restriction_t) \leq t - k + 2\log k$. This being true uniformly in k, we have proven that $(\exists^\infty n \;\; C(\alpha\restriction_n) \geq n - O(1))$ fails to hold.

■

In order to prove the converse, we first prove an effective version of Lemma 2.2.9:

Proposition 2.2.12. *Let $(A^c_n)_{n,c\in\mathbb{N}}$ be a computable family of finite subsets of $2^{<\omega}$ having the property that for all c, there are infinitely many n such that $\lambda([A^c_n]) \leq 2^{-c}$. Then, the set*

$$\bigcap_{c\in\mathbb{N}} \liminf_n([A^c_n])$$

is a $\mathbf{0}'$-Martin-Löf nullset.

Proof. We already know from the proof of Proposition 2.2.9 that for all c, $\liminf_n(\mathcal{A}^c_n)$ has measure at most 2^{-c}. To prove the result, we need to cover, uniformly in c, the set $\liminf_n(\mathcal{A}^c_n)$ by a $\mathbf{0}'$-c.e. open set of measure, say, 2^{-c+1}.

For all n, c, set

$$\mathcal{B}^c_n = \bigcap_{k\geq n} [A^c_k]$$

For every c, the $\mathcal{B}^c_n$ form a non-decreasing sequence of subsets of 2^ω whose union is $\liminf_n(\mathcal{A}^c_n)$. Hence, $\liminf_n([A^c_n])$ is the disjoint union of the sets $\mathcal{B}^c_{n+1} \setminus \mathcal{B}^c_n$.

Notice that

$$\mathcal{B}^c_{n+1} \setminus \mathcal{B}^c_n = \left(\bigcap_{k\geq n+1} [A^c_k]\right) \setminus [A^c_n] = \bigcap_{k\geq n+1} ([A^c_k] \setminus [A^c_n])$$

All the $[A^c_k] \setminus [A^c_n]$ are clopen sets uniformly computable in c, n, k. Hence, the measure of $\bigcap_{k\geq n+1}([A^c_k]\setminus[A^c_n])$, which is the limit over N the nonincreasing sequence $\bigcap_{k=n+1}^{N}([A^c_k] \setminus [A^c_n])$ (all the term are computable uniformly in c, k, n, N), can be computed with oracle $\mathbf{0}'$. And thus, one can find using oracle $\mathbf{0}'$ (and uniformly in c) an N_c such that

$$\lambda\left(\bigcap_{k=n+1}^{N_c} [A^c_k] \setminus [A^c_n]\right) \leq 2\,\lambda\left(\bigcap_{k=n+1}^{+\infty} [A^c_k] \setminus [A^c_n]\right)$$

Hence, for all n, c, the set $\mathcal{V}^c_n = \bigcap_{k=n+1}^{N_c}([A^c_k] \setminus [A^c_n])$ is a $\mathbf{0}'$-computable clopen set which covers $\mathcal{B}^c_{n+1} \setminus \mathcal{B}^c_n$ and has at at most twice its measure. Hence, setting

$$\mathcal{U}_c = \bigcup_{n\in\mathbb{N}} \mathcal{V}^c_n$$

we obtain a $\mathbf{0}'$-c.e. open set which covers $\liminf_n([A^c_n])$ which has at most twice its measure, that is at most 2^{-c+1}. As we have performed this construction uniformly in c, the sequence of sets $(\mathcal{U}_c)_{c\in\mathbb{N}}$ forms a $\mathbf{0}'$-Martin-Löf test that covers $\bigcap_{c\in\mathbb{N}} \liminf_n([A^c_n])$. ∎

How does this help us proving a $\mathbf{0}'$-Martin-Löf random sequence is Kolmogorov random? Suppose that $\alpha \in 2^\omega$ is not Kolmogorov random. By definition this means that the quantity $C(\alpha\restriction_n) - n$ tends to $-\infty$. In other words, for any given c, α belongs to almost all A^c_n – i.e. belongs to $\liminf_n([A^c_n])$ – where

$$A^c_n = \{w \in 2^{<\omega} : \ |w| = n \wedge C(w) < n - c\}$$

By the C-counting theorem (Proposition 2.1.8), we know that there are at most 2^{n-c} elements in A_n^c, and consequently that the measure of $[A_n^c]$ is at most 2^{-c}. We would like to apply Proposition 2.2.12 to the A_n^c to conclude, but here the A_n^c are *not* computable in c, n (they are only computably enumerable) and the computability of the A_n^c is essential in the proof of Proposition 2.2.12. To overcome this difficulty, following Nies et al. [51], we are going to use a trick to make the A_k^c almost computable. This "trick" involves a classical theorem of computability theory, the Low Basis Theorem, proven by Jockusch and Soare in [27]. Recall that a sequence $\beta \in 2^\omega$ is *low* if $\beta' \leq_T \mathbf{0}'$.

Theorem 2.2.13. *Every non-empty effectively closed subset of 2^ω contains a low element.*

Proof. Let $\mathcal{C}$ be an effectively closed subset of 2^ω, and let $\mathcal{U}$ be its complement. Assume that an oracle machine M and an input x are fixed. The computation of M with oracle β on x may terminate or not depending on oracle β. Let us consider the set $T(M, x)$ of all $\beta \in 2^\omega$ such that $M^\beta(x)$ terminates (for fixed machine M and input x). This set is an effectively open set (if the termination happens, it happens due to finitely many oracle values). This set together with $\mathcal{U}$ may cover the entire 2^ω; this means that $M^\beta(x)$ *terminates for all* $\beta \in \mathcal{C}$. If it is not the case, we can add $T(M, x)$ to $\mathcal{U}$ and get a bigger effectively open set $\mathcal{U}'$ that still has non-empty complement and such that $M^\beta(x)$ *does not terminate for all* $\beta \notin \mathcal{U}'$. This operation guarantees (in one of two ways) that termination of the computation $M^\beta(x)$ does not depend on the choice of β (in the remaining non-empty effectively closed set).

This operation can be performed for all pairs (M, x) sequentially. Note that if $\mathcal{U} \cup T(M, x)$ covers the entire space 2^ω, this happens at some finite stage (compactness), so $\mathbf{0}'$ is enough to find out whether it happens or not, and on the next step we have again some effectively open (without any oracle) set. So $\mathbf{0}'$-oracle is enough to say which of the computations $M^\beta(x)$ terminate (as we have said, this does not depend of the choice of β). Therefore any such β is low (the universal β-enumerable set is $\mathbf{0}'$-decidable). And such an β exists since the intersection of the decreasing sequence of non-empty closed sets is non-empty (by compactness). ■

We can now finish the proof of Theorem 2.2.11.

Proof of part ($\Leftarrow$) of Theorem 2.2.11. Let $\mathcal{F}$ be the class of functions $f : 2^{<\omega} \to 2^{<\omega}$ such that:

- $\forall w \in 2^{<\omega}\ f(w) \leq C(w)$
- $\forall k \in \mathbb{N}\ \ \#\{w \in 2^{<\omega} :\ f(w) < k\} \leq 2^k$

Because the first condition imposes some computable bound on the size of $F(w)$ (one must have $f(w) \leq |w| + O(1)$), the class $\mathcal{F}$ can easily be encoded in 2^ω: for example f can be encoded by concatenating all the strings $0^{f(w)}1^{\mathrm{Bin}^{-1}(w)}$ for all $w \in 2^{<\omega}$. Moreover, this encoding $\mathcal{F}'$ of $\mathcal{F}$ is a effectively closed subset of 2^ω as the two above conditions are both of type $\forall x P(x)$ with P a computable predicate: for

the second one this is obvious; and the first one is equivalent to $\forall t\ f(w) \leq C(w)[t]$. Finally, $\mathcal{F}'$ is not empty as it contains C itself. Hence, the Low Basis Theorem applies, and there exists a function $F \in \mathcal{F}$ such that $F' \leq_T \mathbf{0}'$ (here F' means the halting problem relativized to F). We use this particular F to prove the result. Suppose α is not Kolmogorov random. This means that the quantity $C(\alpha\restriction_n) - n$ tends to $-\infty$. A fortiori, $F(\alpha\restriction_n) - n$ tends to $-\infty$ as by definition $F \leq C$. We set for all integers c, n:

$$A_n^c = \{w \in 2^{<\omega} :\ |w| = n \wedge F(w) < |w| - c\}$$

with this notation, we have, like in the proof of Propostion 2.2.10:

$$\alpha \in \bigcap_{c\in\mathbb{N}} \liminf_{n\to+\infty}([A_n^c])$$

and we know by construction of F that A_n^c has at most 2^{n-c} elements, hence $[A_n^c]$ has measure at most 2^{-c} for all n. The A_n^c are uniformly computable with oracle F. Thus, we can apply Proposition 2.2.12 relativized to F, and we get that

$$\bigcap_{c\in\mathbb{N}} \liminf_{n\to+\infty}([A_n^c])$$

is a F'-Martin-Löf nullset. But since $F' \leq_T \mathbf{0}'$, we are done.

■

Another way to define randomness by having infinitely many prefixes of maximal Kolmogorov complexity is to use prefix complexity. As we saw earlier, if $|w| = n$, the maximum value of $K(w)$ is $n + K(n) + O(1)$. This yields the definition:

Definition 2.2.14. *A sequence $\alpha \in 2^\omega$ is* STRONGLY CHAITIN RANDOM *if $K(\alpha\restriction_n) \geq n + K(n) - O(1)$ for infinitely many n.*

By the K-counting theorem (Proposition 2.1.27), for any given n there are $O(2^{n-c})$ strings w of length n such that $K(w) \leq n + K(n) - c$. Hence, with the same argument as for Proposition 2.2.10, we see that the class of strongly Chaitin random sequences has measure 1. In fact, it is rather easy to see that for all c, there exists a c' such that $K(w) \geq |w| + K(|w|) - c \Rightarrow C(w) \geq |w| - c'$, hence strong Chaitin randomness implies Kolmogorov randomness. The question whether these two classes coincide was open for quite some time. At the time of writing this thesis, a positive answer to this question has been announced by J. Miller. His proof involves some delicate concepts (such as lowness for Ω) which go beyond the scope of this thesis.

A particular ML-random sequence: Chaitin's Ω

The Levin-Schnorr characterization of Martin-Löf random sequences as sequences all of whose prefixes have (almost) maximal prefix Kolmogorov complexity tells us that they do not have any pattern of any kind. Therefore, although these sequences form a class of measure 1, it seems difficult to provide an explicit example. Remarkably, Chaitin was able to give a natural and beautiful such example, whose properties will turn out useful in the sequel.

Definition 2.2.15 (Chaitin [12]). *Let $\mathbb{V}$ be the additively optimal machine for K. The sequence $\Omega \in 2^\omega$ is the (infinite) binary expansion of the real number:*

$$\sum_{p \in dom(\mathbb{V})} 2^{-|p|}$$

Notice that this sum is indeed convergent as $dom(\mathbb{V})$ is prefix-free (by Kraft's theorem). One can also see the sum as the Lebesgue measure of the open set generated by $dom(\mathbb{V})$ (for this reason, Ω is sometimes referred to as a *halting probability*, and we will often identify Ω to the real number it represents). The sequence Ω has the following remarkable properties:

Theorem 2.2.16.
(i) Ω is left-c.e.
(ii) Ω is Martin-Löf random
(iii) The Turing degree of Ω is $\mathbf{0}'$

Proof. (i) This is because $dom(\mathbb{V})$ is a c.e. subset of $2^{<\omega}$, hence:

$$\Omega = \lim_{t \to +\infty} \sum_{p \in dom(\mathbb{V})[t]} 2^{-|p|}$$

meaning that Ω is the limit of a computable non-decreasing sequence of reals (i.e. Ω is left-c.e.)

(ii) To prove that Ω is Martin-Löf random, we are going to prove that for all n, $K(\Omega\restriction_n) \geq n - O(1)$, concluding by the Levin-Schnorr theorem (Theorem 2.2.5). Suppose we know $\Omega\restriction_n$, which seen as a real number is a dyadic r such that $\Omega \in [r, r + 2^{-n}]$. Knowing this r, we enumerate $dom(\mathbb{V})$ until we find a stage t_n such that

$$\sum_{p \in dom(\mathbb{V})[t_n]} 2^{-|p|} \in [r, r + 2^{-n}]$$

Since Ω itself is in $[r, r + 2^{-n}]$, this means:

$$\sum_{p \in dom(\mathbb{V}) \setminus dom(\mathbb{V})[t_n]} 2^{-|p|} \in [0, 2^{-n}]$$

Thus, for all $p \in dom(\mathbb{V}) \setminus dom(\mathbb{V})[t_n]$, we have $|p| \geq n$. In other words, $dom(\mathbb{V})[t_n]$ contains all the elements $p \in dom(\mathbb{V})$ that have length smaller than n. Hence, if we compute $\{\mathbb{V}(p) : p \in dom(\mathbb{V})[t_n]\}$ and we take the first $w_n \in 2^{<\omega}$ which is *not* in this set, we have $K(w_n) \geq n$. And since knowing $\Omega\restriction_n$ is enough to find w, we have

$$K(\Omega\restriction_n) \geq K(w_n) - O(1) \geq n - O(1)$$

This proves that Ω is Martin-Löf random.

(*iii*) Ω being left-c.e., it is in particular $\mathbf{0}'$-computable. Conversely, the argument we used to prove (*ii*) shows that knowing the first n bits of Ω is enough to find all the strings w such that $K(w) < n$. Hence, with oracle Ω, the predicate $K(w) < n$ is computable uniformly in n, w. Hence, $K \leq_T \Omega$. Combining this with Theorem 2.1.19, we get $\mathbf{0}' \leq_T \Omega$. ■

By definition, Ω being left-c.e. means that there exists a computable sequence of strings $(w_s)_{s\in\mathbb{N}}$ that is non-decreasing for $\leq_{lex}$ and that pointwise converges to Ω. Hence, it seems that if we ran a program computing this sequence of strings, we would get after some time a string that coincides with Ω on a long prefix, hence we would have generated true randomness with a computer. The following proposition tells us that this might not be such a good idea: the time needed to get n bits of Ω right is a function of n that majorizes every computable function.

Proposition 2.2.17. *Let $(w_s)_{s\in\mathbb{N}}$ be a computable sequence of strings that is non-decreasing for $\leq_{lex}$ and pointwise converges to Ω. For all n, set*

$$T(n) = \min\{s : \ w_s >_{lex} \Omega\restriction_n\}$$

The function T is non-decreasing and dominates every computable function.

Proof. Let f be a computable function. Let n such that $f(n) \geq T(n)$. Given such an n, one can compute $w_{f(n)}$. Since we have

$$\Omega\restriction_n \leq_{lex} w_{T(n)} \leq_{lex} w_{f(n)} \leq_{lex} \Omega$$

(the first inequality comes from the definition of T, the second one from the fact that the sequence of w_s is non-decreasing for $\leq_{lex}$, and the third one by the fact that Ω is the pointwise limit of the w_s), the first n bits of $w_{f(n)}$ coincide with those of Ω. Hence for any such n:

$$K(\Omega\restriction_n) \leq K(w_{f(n)}, n) \leq K(n) + O(1) \leq O(\log n)$$

Since $K(\Omega\restriction_n) \geq n - O(1)$ for all n, there are only finitely many n that can satisfy the above inequality, hence finitely many n such that $f(n) \geq T(n)$. Hence, T dominates f. ■

2.2.2 Computable randomness and Schnorr randomness vs Kolmogorov complexity

As we just saw, Martin-Löf random sequences can be nicely and simply characterized in terms of the Kolmogorov complexity of their initial segments (which more or less need to have close-to-maximal complexity). Is it possible to give such a characterization for computable randomness or Schnorr randomness? Since these notions are weaker than Martin-Löf randomness, one could expect a characterization involving a weaker condition on the complexity of the initial segments, something like "α is computably random iff $K(\alpha) \geq n - f(n)$" for some function f. It turns out that the situation is radically different. Indeed, the concepts of computable and

Schnorr randomness are somewhat orthogonal to Kolmogorov complexity, as there exists some computably random sequences of very small Kolmogorov complexity and some sequences of high complexity that are not Schnorr random. The following result of Merkle[1] illustrates the first part of this statement:

Theorem 2.2.18 (Merkle [44]). *There exists a computably random sequence α such that, for every computable order h, $K(\alpha\restriction_n |n) \leq h(n) + O(1)$. Moreover, there exists such a sequence α that is $\mathbf{0}'$-computable.*

This is quite a contrast to Theorem 2.2.5! Let us have a look at the proof.

Proof. The idea of the proof is the following. In order to make a computably random sequence, we will play the role of the bank, and we need to defeat an infinite number of players: those that have a total computable betting strategy. Of course, we cannot (effectively) find out without extra information who these players are since the set

$$\{n : \text{ the } n\text{-th partial computable function is total}\}$$

is not computable, but we will see how to deal with this.

We already know how to defeat one given player, by the diagonalization technique presented in Propostion 1.4.10. This also tells us how to defeat a finite number of players: if $d_0, \dots, d_n$ are martingales, then for any n-uple $c_0, \dots, c_n$ of positive real numbers, $d = c_0 d_0 + \dots + c_n d_n$ is a martingale such that $\mathrm{Succ}(d_i) \subseteq \mathrm{Succ}(d)$ for all $i \leq n$. Hence, diagonalizing against d, we can defeat all the d_i at the same time. Here, the situation is slightly more complicated since we have to defeat infinitely many players. The idea of the construction is the following. We start by enumerating all normed partial computable martingales $(d_n)_{n\in\mathbb{N}}$ (i.e. we enumerate partial computable functions d such that $d(\epsilon) = 1$ and if $d(w0)$ and $d(w1)$ are defined, then $d(w)$ is defined and is the average of these two quantities). Then, we take a sequence $0 = N_0 < N_1 < N_2 < \dots$ of natural numbers (to be specified later). We split $\mathbb{N}$ into intervals $I_k = [N_k, N_{k+1})$. During the stages $n \in I_k$ of the construction of α, we will only diagonalize against the weighted sum $\sum_{i\in P_k} 2^{-i-N_i} d_i$ where P_k is the set of $i \leq k$ such that the martingale d_i is defined on all strings of length N_{k+1}. The factor 2^{-i-N_i} ensures that by the time d_i starts being taken into account in the diagonalization, it has not made much money (namely less than 2^{-i}). The diagonalization thus ensures that none of the d_i succeeds (we will give the details in a moment). How do we perform the construction of α while keeping the Kolmogorov complexity of its prefixes small? We take a sequence $0 < N_0 \leq N_1 \leq N_2 \dots$ that grows very fast (faster than any computable function) but such that the complexity of N_k does not exceed k (for this we will use Proposition 2.1.10). Hence, for all stages $n < N_k$, to compute α_n we only need to know which of the first k players are still playing (this represent an amount of information of k bits) and their corresponding

[1] Some similar results appeared in [33]

factor 2^{-N_i-i} (which represents an amount $K(N_0)+K(N_1)+...+K(N_k)$ of information). We conclude by arguing that k and $K(N_k)$ are very small relatively to N_k.

Here are the formal details of the proof. Let $\{d_k : n \in \mathbb{N}\}$ be an enumeration of partial computable normed martingales . It is rather easy to see that these functions can be effectively enumerated. Let Tot $= \{n \in \mathbb{N} : d_k \text{ is total}\}$. Let $B' : \mathbb{N} \to \mathbb{N}$ be the function defined by $B'(k) = \max\{n \in \mathbb{N} : K(n) \leq k\}$ which for the same reason as the function B defined in Proposition 2.1.10 dominates any computable function and is $\mathbf{0}'$-computable (see proof of Theorem 2.1.12). Set $N_k = B'(k)+1$ for all k.

The diagonalization between stage N_k and stage $N_{k+1}-1$ is performed against a linear combination of the $\{d_i : i \in P_k\}$ where

$$P_k = \left\{i \in \mathbb{N} : \ \big(i \leq k\big) \wedge \big(\forall w \ s.t. \ |w| \leq N_{k+1} \ \ d_i(w)\downarrow\big)\right\}$$

Notice that for all i, i belongs to Tot if and only if it belongs to almost all P_k.

Construction. We construct the sequence α by induction where after step k of the induction $\alpha{\restriction}_{N_k}$ is constructed. Suppose we have already defined $\alpha{\restriction}_{N_k}$. To construct $\alpha{\restriction}_{N_{k+1}}$, pick a $w \in 2^{<\omega}$ of length $N_{k+1}-N_k$ such that:

$$\sum_{i\in P_k} 2^{-N_i-i} d_i(\alpha{\restriction}_{N_k} \ w) < 2^{-(k+1)} + \sum_{i\in P_k} 2^{-N_i-i} d_i(\alpha{\restriction}_{N_k})$$

Such a w must exist by the fairness condition of martingales. Notice that if N_k, N_{k+1} and P_k are known, w can be found effectively since for all $i \in P_k$ the restriction of d_i to the strings of length at most N_{k+1} is total computable. Set $\alpha{\restriction}_{N_{k+1}} = \alpha{\restriction}_{N_k} \ w$.

Verification. By definition of the P_k, we have for all k: $P_{k+1} \subseteq P_k \cup \{k+1\}$. Hence, for all k:

$$\begin{aligned}
\sum_{i\in P_{k+1}} 2^{-N_i-i} d_i(\alpha{\restriction}_{N_{k+1}}) &\leq 2^{-N_{k+1}-k-1} d_{k+1}(\alpha{\restriction}_{N_{k+1}}) + \sum_{i\in P_k} 2^{-n_i-i} d_i(\alpha{\restriction}_{N_{k+1}}) \\
&\leq 2^{-N_{k+1}-k-1} 2^{N_{k+1}} + \sum_{i\in P_k} 2^{-N_i-i} d_i(\alpha{\restriction}_{N_{k+1}}) \\
&\leq 2^{-(k+1)} + 2^{-(k+1)} + \sum_{i\in P_k} 2^{-N_i-i} d_i(\alpha{\restriction}_{N_k}) \\
&\leq 2^{-k} + \sum_{i\in P_k} 2^{-N_i-i} d_i(\alpha{\restriction}_{N_k})
\end{aligned}$$

This proves that the sequence $\left(\sum_{i\in P_k} 2^{-N_i-i} d_i(\alpha{\restriction}_{N_k})\right)_{k\in\mathbb{N}}$ is bounded.

Suppose that α is not computably random. This means that there exists a normed martingale d_j ($j \in$ Tot) such that $\lim_n d_j(\alpha{\restriction}_n) = +\infty$ hence in particular

$\lim_k d_j(\alpha\restriction_{N_k}) = +\infty$. Since $j \in \mathrm{Tot}$, j belongs to almost all P_k thus for almost all k:

$$\sum_{i \in P_k} 2^{-N_i - i} d_i(\alpha\restriction_{N_k}) \geq 2^{-N_j - j} d_j(\alpha\restriction_{N_k})$$

and the right-hand side of this inequality tends to $+\infty$. This is a contradiction because $\sum_{i \in P_k} 2^{-N_i - i} d_i(\alpha\restriction_{N_k})$ is bounded. Hence α is computably random.

It remains to evaluate the Kolmogorov complexity of the initial segments of α. Let $n \in \mathbb{N}$. Let k be such that $n \in (N_{k-1}, N_k]$, that is $k = B'^{-1}(n)$. To compute $\alpha\restriction_n$, as we said above, we only need to know $P_0, \ldots, P_{k-1}$ and $N_0, \ldots, N_k$. Thus:

$$K(\alpha\restriction_n | n) \leq \sum_{i<k} K(P_i) + \sum_{i \leq k} K(N_i) + O(1)$$

By construction of the N_i and P_i, $K(P_i) = O(i)$ and $K(N_i) = O(i)$ for all i. Hence

$$K(\alpha\restriction_n) = O(k^2) = O\Big((B'^{-1}(n))^2\Big)$$

Since B' grows faster that any computable function, B'^{-1} is an order which grows slower that any computable one, hence this is also the case for $n \mapsto (B'^{-1}(n))^2$.

Finally, notice that the sequence α is $\mathbf{0}'$-computable. This is because both the sequence $(N_k)_{k \in \mathbb{N}}$ and $(P_k)_{k \in \mathbb{N}}$ are $\mathbf{0}'$-computable, and the construction is computable when these two sequences are given as an oracle. ■

Since a sequence is Martin-Löf random if and only if $K(\alpha\restriction_n) \geq n - O(1)$ (Theorem 2.2.5), we get as an immediate corollary:

Corollary 2.2.19. *Computable randomness is weaker than Martin-Löf randomness.*

The theorem we just proved tells us how low the Kolmogorov complexity of the initial segments of a computably random (or Schnorr random reals) can be, and this allowed us to separate Martin-Löf randomness from computable randomness and Schnorr randomness. We now turn our attention to the dual question: can a sequence $\alpha \in 2^\omega$ have segments of high Kolmogorov complexity and yet be not computably random, or not Schnorr random? Similarly, can we find a *sufficient* condition on the Kolmogorov complexity of the initial segments for a sequence to be computably or Schnorr random? The following proposition provides a first answer for Schnorr randomness.

Proposition 2.2.20. *(i) For every computable order h, there exists a non-Schnorr random sequence β such that $K(\beta\restriction_n) \geq n - h(n) - O(1)$ for all n.*
(ii) If β is not Schnorr random, there exists a computable order g such that $K(\beta\restriction_n) \leq n - g(n) + O(1)$ for infinitely many n.

Proof. (i) Let h be a given order. The idea is simple: we are going to splice zeros in a sparse way into some Martin-Löf random sequence. Set $f = h^{-1}$. f is computable since h is. We take a Martin-Löf random sequence α (by the Levin-Schnorr theorem it satisfies $K(\alpha\restriction_n) \geq n - O(1)$) and we insert zeros into α at places $f(0) < f(1) < ...$ and we call β the resulting sequence:

$$\beta = \alpha_{(0)}\alpha_{(1)}...\alpha_{(f(0)-1)}\,0\,\alpha_{(f(0))}\alpha_{(f(0)+1)}...\alpha_{(f(1)-1)}\,0\,\alpha_{(f(1))}...$$

Since f is computable, the inserted zeros carry no information. Hence, for all n, if we call k_n the number of zeros among the first n bits of β that come from the insertion (we easily see that $k_n = f^{-1}(n - k_n) \leq f^{-1}(n)$), we have $K(\beta\restriction_n) = K(\alpha\restriction_{n-k_n})$. But since α is Martin-Löf random, $K(\alpha\restriction_{n-k_n}) \geq n - k_n - O(1)$. Thus:

$$K(\beta\restriction_n) \geq n - k_n \geq n - f^{-1}(n) \geq n - h(n) + O(1)$$

And of course β is not Schnorr random as one can predict with certainty all the inserted bits. hence it is easy to construct a computable martingale d such that $d(\alpha) \geq 2^{h^{-1}(n)}$ for all n, and $n \mapsto 2^{h^{-1}(n)}$ is an order as h^{-1} is.
(ii) We postpone the proof of this part, as it is a direct corollary of Theorem 2.3.26 below.

■

The second part of the above proposition gives us a sufficient condition for Schnorr randomness: let h be an order that is dominated by all computable orders; if $K(\alpha\restriction_n) \geq n - h(n) - O(1)$ for all n, then α is Schnorr random. We now prove that there is no such sufficient condition for computable randomness. In fact, not even for Church stochasticity!

Theorem 2.2.21. *Let h be any order (non necessarily computable). There exists a sequence $\alpha \in 2^\omega$ which is not Church stochastic (a fortiori not computably random) such that $K(\alpha\restriction_n) \geq n - h(n) - O(1)$ for all n.*

The idea of the proof is the following. We take a very fast growing function $F : \mathbb{N} \to \mathbb{N}$ and, like in the previous proof, we insert zeros into a Martin-Löf random sequence α (which we will chose later) at positions $F(0), F(1), \ldots$. If F grows fast enough, this will ensure that the resulting sequence β has high Kolmgorov complexity, as the set of inserted zeros will be sparse. To make β not Church stochastic, we define a selection rule that selects the bits in positions $F(i)$. However, F might not be computable, hence the selection rule might not know when to perform the selections. To overcome this difficulty, we pick a random sequence α which contains enough information to compute F. This is made possible by the Kučera-Gács theorem:

Theorem 2.2.22 (Kučera [31], Gács [20])**.** *For every sequence $\alpha \in 2^\omega$, there exists $\beta \in$ **MLR** such that $\alpha \leq_T \beta$.*

We do not provide the proof of this theorem, referring the reader to the original papers, or to [46] for an elegant proof using a martingale argument.

Back to our construction, given our fast growing function F, we can take a random α which computes F and insert zeros in α at positions $F(i)$. However, this still not enough. Indeed, we want to have a monotonic computable selection rule σ that selects the bits in positions $F(i)$. If a zero is inserted at position n, the selection rule will only know $\alpha\restriction_{[0,n-1]}$ before making its decision to select the n-th bit or not. And this might not be enough as the number of bits of α which are necessary to figure out that $n \in range(F)$ might exceed n. Hence, instead of inserting a zero at position $n = F(i)$, we insert it at a later position $n' = F'(i)$ ensuring that reading $\alpha\restriction_{n'}$ gives the selection rule enough information and time to figure out that $n' \in range(F')$.

Proof of Theorem 2.2.21. The detailed proof goes as follows. First, let us introduce a useful piece of notation. If Z is a subset of $\mathbb{N}$, and α, β are two elements of 2^ω, we call Z-join of α and β, and denote by $\alpha \oplus_Z \beta$, the element of 2^ω we get by merging α and β, placing the bits of β in positions i's such that $i \in Z$. Formally,

$$(\alpha \oplus_Z \beta)_{(i)} = \begin{cases} \alpha_{(|\bar{Z} \cap \{0..i-1\}|)} & \text{if } i \notin Z \\ \beta_{(|Z \cap \{0..i-1\}|)} & \text{if } i \in Z \end{cases}$$

If $Z = 2\mathbb{N}+1$, we have $\alpha \oplus_Z \beta = \alpha_{(0)}\beta_{(0)}\alpha_{(1)}\beta_{(1)}...$, and we abbreviate $\alpha \oplus_Z \beta$ by $\alpha \oplus \beta$. If $F : \mathbb{N} \to \mathbb{N}$ is a non-decreasing function, we abbreviate $\alpha \oplus_{F(\mathbb{N})} \beta$ by $\alpha \oplus_F \beta$.

Let h be an order. Let $F = h^{-1}$. By the Kučera-Gács theorem, let $\alpha \in \mathbf{MLR}$ such that $F \leq_T \alpha$. We call Φ the Turing reduction such that $F = \Phi^\alpha$, i.e., for all n, $\Phi^\alpha(n) = F(n)$. For all $n \in \mathbb{N}$, set

$$F'(n) = F(n) + n + \min\left\{t \in \mathbb{N} : \ (\forall k \leq n)\ \Phi^{\alpha\restriction_t}(k)[t] \downarrow = F(k)\right\}$$

(where $\Phi^{\alpha\restriction_t}(k)[t]$ is the result of the computation of $\Phi(n)$ using only the first t bits of α and during at most t steps of computation, which may be undefined). Notice that F' is increasing, $F' > F$ and $F' \leq_T \alpha$.

We set $\beta = \alpha \oplus_{F'} 0^\omega$ i.e. we insert zeros in α at positions $F'(0) < F'(1) < F'(2) < \ldots$. We now define a total computable (monotonic) selection rule σ which, when running on β, selects the bits in positions $F'(0), F'(1), \ldots$ and only these bits.

On an input $w \in 2^{<\omega}$, σ does the following:

1. First, it computes the finite string u corresponding to w from which the bits previously selected by σ have been deleted (in other words, $u = \overline{\sigma}[w]$ where $\overline{\sigma}$ is the dual of σ which scans whenever σ selects and vice-versa).

2. It computes the set
$$A = \{k \leq |u| : \ \Phi^u(k)[|u|] \downarrow\}$$
and then finds the largest integer m such that $[0, m] \subseteq A$.

3. For all $n \leq m$, it computes

$$t_n = \min\{t \in \mathbb{N} : \ (\forall k \leq n)\ \Phi^{u\restriction_t}(k)[t] \downarrow\}$$

and then

$$B = \{\Phi^u(n)[|u|] + n + t_n : \ n \leq m\}$$

4. Finally, if $|w| \in B$, $\sigma(w) = \texttt{select}$; otherwise, $\sigma(w) = \texttt{scan}$.

Let us now verify that σ is as desired. First, notice that σ is computable (this is clear) and is total as the above algorithm only involves bounded computations. Let us now prove that if σ runs on $\alpha \oplus_{F'} 0^\omega$, it only selects the bits of positions $F'(0) < F'(1) < F'(2) < \dots$ (which are all zeros hence β is not Church stochastic).

By induction, suppose that after reading $\beta\restriction_n$, σ has exactly selected the bits in positions $F'(0), \dots, F'(i-1)$ where i is the largest integer such that $F'(i-1) < n$. Then, the above step 1 produces $u = \alpha\restriction_{(n-i)}$. The step 2 produces a set A of integers k for which $\alpha\restriction_{(n-i)}$ and $n-i$ steps of computation are enough to compute $F(k)$. Thus, by definition of F', the set B constructed at step 3 can only contain elements in the range of F'. Since σ only selects the next bit if the position of this bit belongs to B, we have proven that σ will never make a bad selection. It remains to prove that it never misses a selection of a bit whose position belongs to the range of F'. Suppose $n = F'(i)$, then by definition of F':

$$n - i = F'(i) - i \geq \min\left\{t \in \mathbb{N} : \ (\forall k \leq i)\ \Phi^{\alpha\restriction_t}(k)[t] \downarrow = F(k)\right\}$$

Thus, at step 2 of the algorithm, all $j \leq i$ are in A and at step 3, $F'(i) \in B$. Hence, the bit of position n will indeed be selected.

We have proven that when running on β, σ selects an infinite subsequence of zeros, hence β is not Church stochastic. To complete the proof, it remains to evaluate the Kolmogorov complexity of the prefixes of β. In the above argument, for $n \in \mathbb{N}$ and i the largest integer such that $F'(i-1) < n$, the string $\beta\restriction_n$ is made of $\alpha\restriction_{(n-i)}$ in which zeros have been inserted in positions $F'(0), \dots, F'(i-1)$. One can retrieve $\alpha\restriction_{(n-i)}$ from $\beta\restriction_n$, simply by running σ on $\beta\restriction_n$ and deleting the bits selected by σ, hence $K(\alpha\restriction_{(n-i)}) \leq K(\beta\restriction_n)$. Since α is Martin-Löf random, we have

$$K(\beta\restriction_n) \geq K(\alpha\restriction_{(n-i)}) \geq n - i - O(1) \geq n - F'^{-1}(n) - O(1)$$

(the last inequality coming from the definition of i). Remember that $F = h^{-1}$ and F' dominates F, hence F'^{-1} is dominated by F^{-1} and $F^{-1} \leq h + O(1)$. Hence, $K(\beta\restriction_n) \geq n - h(n) - O(1)$ for all n, which completes the proof.

■

Corollary 2.2.23. *Schnorr randomness does not imply Church stochasticity. Since computable randomness implies both Schnorr randomness and Church chasticity, this in particular means that Schnorr randomness is weaker than computable randomness.*

Proof. Let h be an order that is dominated by any computable order (such as B^{-1} where B is the function defined in Proposition 2.1.10. By Theorem 2.2.21, there exists $\beta \in 2^\omega$ such that is not Church stochastic and such that $K(\beta\restriction_n) \geq n - h(n) - O(1)$ for all n. By Proposition 2.2.20, this second property guarantees that β is Schnorr random. ■

Adapting the proof of Theorem 2.2.21, we can prove that there exists a *left-c.e.* sequence α which has high complexity and yet is not Church stochastic:

Proposition 2.2.24. *There exists a left-c.e. sequence $\alpha \in 2^\omega$ which is not Church stochastic and such that for every computable order h, $K(\alpha\restriction_n) \geq n - h(n) - O(1)$ (implying in particular that α is Schnorr random).*

This improves a result of Merkle et al. [47] who proved this theorem for a stronger notion of stochasticity (Mises-Wald-Church stochasticity, which is defined like Church stochasticity, but allowing *partial* computable selection rules).

Proof. The idea is to insert zeros in Ω, which we know is random and can compute a function which dominates any computable one. Indeed, given a computable sequence $(w_s)_{n\in\mathbb{N}}$ that is non-decreasing for $\leq_{lex}$ and pointwise converges to Ω, the function T defined in Proposition 2.2.17 (computable from Ω) dominates every computable function. Define for all n:

$$F(n) = n + 1 + T(n) = n + 1 + \min\{s : \ w_s >_{lex} \Omega\restriction_n\}$$

and take $\beta = \Omega \oplus_F 0^\omega$. We claim that this α has the desired properties.
First, we need to check that α is left-c.e. For all s, set $\Omega^s = w_s 0^\omega$ and

$$F^s(n) = n + 1 + \min\{t : \ w_t >_{lex} \Omega^s\restriction_n\} \ \text{ for all } n$$

The F^s form a uniformly computable sequence of non-decreasing functions, which satisfy $F^s \leq F^{s+1}$ for all s, and pointwise converge to the above function F. Hence, α is the pointwise limit of the sequence $\alpha^s = \Omega^s \oplus_{F^s} 0^\omega$. This sequence is clearly uniformly computable and pointwise converges to α. To prove that α is left-c.e., we need to prove that $(\alpha^s)_{s\in\mathbb{N}}$ is non-decreasing for the lexicographic order. This is a consequence of the following simple lemma:

Lemma 2.2.25. *(i) Let F be some non-decreasing function, and β, β' two elements of 2^ω such that $\beta \leq_{lex} \beta'$. Then, $\beta \oplus_F 0^\omega \leq_{lex} \beta' \oplus_F 0^\omega$*
(ii) Let F and G be two non-decreasing functions such that $F \leq G$. For all $\beta \in 2^\omega$, $\beta \oplus_F 0^\omega \leq_{lex} \beta \oplus_G 0^\omega$

Subproof. Part (i) is straightforward. For part (ii), observe that $F \leq G$ means that one goes from $\beta \oplus_F 0^\omega$ to $\beta \oplus_G 0^\omega$ by moving the inserted zeros to the right. And since for any strings u, v, w we have $u0vw \leq_{lex} uv0w$, the result follows. □

Hence, in our construction, we have for all s:

$$\alpha^s = \Omega^s \oplus_{F^s} 0^\omega \leq_{lex} \Omega^{s+1} \oplus_{F^s} 0^\omega \leq_{lex} \Omega^{s+1} \oplus_{F^{s+1}} 0^\omega = \alpha^{s+1}$$

Hence, α is left-c.e. The proof that α is not Church stochastic is somewhat similar to the proof of Theorem 2.2.21. Let σ be the selection rule which on an input u does the following:

1. First, it computes the finite string u' corresponding to u from which the bits previously selected by σ have been deleted
2. Then, it computes w_s, where $s = |u'|$ and finds the shortest common prefix v of u' and w_s
3. For all $n \leq |v|$, it computes

$$t_n = min\{t : \ w_t >_{lex} v\restriction_n\}$$

and then computes the set

$$B = \{n + 1 + t_n : \ n < |v|\}$$

4. If $|u| \in B$, then $\sigma(u) = \texttt{select}$; otherwise $\sigma(u) = \texttt{scan}$.

It is clear that σ is total. It remains to prove that, when running on $\alpha = \Omega \oplus_F 0^\omega$, σ selects exactly the bits in positions $F(0) < F(1) < \ldots$. Let $n \in \mathbb{N}$, and i be the largest integer such that $F(i-1) < n$. By induction, suppose that after running σ on $\alpha\restriction_n$, the selected bits are exactly those in positions $F(0), \ldots, F(i-1)$. In the above algorithm, if $u = \alpha\restriction_n$, $u' = \Omega\restriction_{n-i}$, hence $v \sqsubseteq \Omega$, hence by definition of F, the set B constructed at step 3 only contains elements in the $range(F)$, hence σ never selects a bit in a position which is not in $range(F)$. On the other hand, if $n = F(i)$, $u' = \Omega\restriction_{F(i)-i}$ hence at step 2, $w_s = w_{F(i)-i}$. By definition of F, $F(i) - i > T(i)$ hence

$$\Omega\restriction_i \leq_{lex} w_{T(i)} \leq_{lex} w_s \leq_{lex} \Omega$$

(the first inequality comes from the definition of T). Hence, w_s coincides with Ω on at least the first i bits, which means that v computed at step 2 has at least length i, hence at step 3, the computed set B contains $F(i)$, and hence at step 4 the bit of position $F(i)$ is selected.

Finally, like in the proof of Theorem 2.2.21, we can prove that $K(\alpha\restriction_n) \geq n - F^{-1}(n) - O(1)$ for all n. Since F dominates every computable function (by definition of F and because T does, see Proposition 2.2.17), the function F^{-1} is an order that is dominated by any computable one. ■

Combining Proposition 2.2.24 with Proposition 2.2.20, we get:

Corollary 2.2.26. *There exists a left-c.e. sequence α which is Schnorr random but not computably random.*

2.2.3 Effective Hausdorff dimension vs Kolmogorov complexity

In this section, we present the elegant result of Mayordomo [43] who, elaborating on the work of Staiger, Ryabko and others, proved that constructive Hausdorff dimension *can* be characterized in terms of Kolmogorov complexity.

Theorem 2.2.27 (Mayordomo [43]). *For every sequence $\alpha \in 2^\omega$:*

$$\mathrm{cdim}(\alpha) = \liminf_{n\to+\infty} \frac{K(\alpha\restriction_n)}{n}$$

Remark 2.2.28. *Since K and C only differ by a logarithmic factor, this theorem is also true with C is place of K.*

Proof. We first prove that $\mathrm{cdim}(\alpha) \leq \liminf \frac{K(\alpha\restriction_n)}{n}$. Let s be a rational such that $\liminf \frac{K(\alpha\restriction_n)}{n} < s$. For all k, there are infinitely many n such that $K(\alpha\restriction_n) \leq sn-k$. Let for all k

$$A_k = \{w \in 2^{<\omega} : \ K(w) \leq s|w| - k\}$$

A_k is a c.e. set of strings uniformly in k. Moreover, by Kraft-Chaitin's theorem:

$$\sum_{w\in A_k} 2^{-s|w|+k} \leq \sum_{w\in A_k} 2^{-K(w)} \leq 1$$

Hence

$$\sum_{w\in A_k} 2^{-s|w|} \leq 2^{-k}$$

Thus, $(A_k)_{k\in\mathbb{N}}$ is a constructive s-test that covers α. This means that $\mathrm{cdim}(\alpha) \leq s$. Conversely, let s be a rational number such that $s > \mathrm{cdim}(\alpha)$. By definition of $\mathrm{cdim}(\alpha)$, let $(A_n)_{n\in\mathbb{N}}$ be an s-test that covers α that is for all n $\alpha \in [A_n]$ and $\sum_{w\in A_n} 2^{-s|w|} \leq 2^{-n}$. We build a Kraft-Chaitin set L as follows. Each time a string w is enumerated in some A_n with $n > 0$, we enumerate $(w, s|w|)$ into L. This makes L a Kraft-Chaitin set since

$$\sum_{(w,k)\in L} 2^{-k} = \sum_{w\in\bigcup A_n} 2^{-s|w|} \leq \sum_{n>0}\sum_{w\in A_n} 2^{-s|w|} \leq \sum_{n>0} 2^{-n} \leq 1$$

Hence by Kraft-Chaitin's theorem, if $w \in \bigcup_{n>0} A_n$, $K(w) \leq s|w| + O(1)$. Since α has infinitely many prefixes in $\bigcup_{n>0} A_n$, this proves $\liminf \frac{K(\alpha\restriction_n)}{n} \leq s$. ∎

We have seen that every Martin-Löf random sequence has constructive dimension 1. Comparing the Levin-Schnorr theorem (Theorem 2.2.5) for Martin-Löf randomness and Mayordomo's theorem for constructive dimension, we see that the Kolmogorov complexity point of view also explains well why this is true. The converse of this result is not true: a sequence of constructive dimension 1 needs not be Martin-Löf random. Indeed, not even Church stochastic.

Proposition 2.2.29. *(i) A sequence of constructive dimension 1 needs not be Church stochastic nor Schnorr random (a fortiori, computably random).*
(ii) A computably random sequence (a fortiori: Church stochastic, Schnorr random) may have constructive dimension 0.

Proof. (*i*) By Theorem 2.2.21, there exists a sequence α which is not Church stochastic and such that $K(\alpha\restriction_n) \geq n - \log n - O(1)$ for all n (hence by the above theorem, it has constructive dimension 1). Also, by and Proposition 2.2.20, there exists a non-Schnorr sequence β such that $K(\beta\restriction_n) \geq n - \log n - O(1)$ for all n.
(*ii*) By Theorem 2.2.18, there exists a sequence α which is computably random and satisfies $K(\alpha\restriction_{|n}) \leq \log n + O(1)$ for all n. Hence, $K(\alpha\restriction_n) = O(\log n)$, which means by the above theorem that α has constructive dimension 0. ■

2.2.4 Stochasticity vs Kolmogorov complexity

We just saw in Theorem 2.2.18 that some computably random sequences are of very low Kolmogorov complexity (i.e. their prefixes are all of low Kolmogorov complexity). As computable randomness implies Church stochasticity, this implies that Church stochasticity is also somewhat orthogonal to Kolmogorov complexity: high Kolmogorov complexity does not imply Church stochasticity and Church stochasticity does not imply high Kolmogorov complexity. Quite surprisingly, introducing non-monotonicity in the definition of stochasticity completely changes this situation: a Kolmogorov-Loveland stochastic sequence needs to have high Kolmogorov complexity. Kolmogorov had conjectured that there existed a Kolmogorov-Loveland stochastic sequence α such that $K(\alpha\restriction_n) = O(\log n)$. This conjecture was refuted by Muchnik et al. [50] who proved that any Kolmogorov-Loveland stochastic sequence α satisfies $K(\alpha\restriction_n) \geq n - o(n)$ for infinitely many n. Elaborating on Muchnik et al.'s techniques, Merkle et al. were able to prove that in fact:

Theorem 2.2.30 (Merkle et al. [47])**.** *For any Kolmogorov-Loveland stochastic sequence α: $K(\alpha\restriction_n) \geq n - o(n)$ for all n (or equivalently, $\mathrm{cdim}(\alpha) = 1$).*

In other words, if $K(\alpha\restriction_n) \leq rn$ for some $r < 1$ and infinitely many n, then one can select from α (in a non-monotonic way) a subsequence β such that $\mathrm{Bias}(\beta) > 0$. However, Merkle et al.'s paper does not give any relation between r and the maximal $\mathrm{Bias}(\beta)$ one can obtain by computable non-monotonic selection. The relation between the randomness deficiency and the maximal bias of selected subsequences was studied for finite binary sequences by Asarin [3], Durand and Vereshchagin [16]. In what follows, we will study this problem for *infinite* binary sequences. Combining the techniques we developed in Chapter 1 to characterize Church stochasticity by martingales with Merkle et al.'s approach, we will prove the following improvement of Theorem 2.2.30:

Theorem 2.2.31. *Let $\alpha \in 2^\omega$ and s such that $\mathrm{cdim}(\alpha) \leq s$. There exists a non-monotonic selection rule σ, computable with oracle s such that, setting $\delta = \mathrm{Bias}(\sigma[\alpha])$, we have $\mathcal{H}\left(\frac{1}{2} + \delta\right) \leq s$.*

Both papers [47] and [16] use the same main three techniques, which are already present in [50]:

1. Splitting technique. Any sequence (finite or infinite) which has a linear randomness deficiency can be split into a finite number of subsequences such that at least two of the subsequences have a linear randomness deficiency relatively to the other ones.

2. Competing strategies. Given two finite sequences u and v with known randomness deficiencies (say, respectively $K(u) = |u| - d_1$ and $K(v) = |v| - d_2$), one can construct (the construction depending only on d_1 and d_2, not on (u, v)) two strategies (the concept of strategy is formalized below) S_1 and S_2 such that: S_1 reads v and bets on u, S_2 reads u and bets on v, and either S_1 multiply its initial capital by a least 2^{d_1} or S_2 multiply its initial capital by at least 2^{d_2}. Hence, a good way to predict the bits of a string w with some random deficiency is to use the above technique 1 to split w into pieces such that two of them have some randomness deficiency and apply technique 2.

3. Converting a strategy into a selection rule. If a betting strategy wins on a sequence (finite or infinite) an amount of money which is expontential in the number of bets, one can construct from this strategy a selection rule which selects a biased subsequence.

In turn, we will use these three techniques. Our improvement of Theorem 2.2.30 mainly comes from the precise quantitative results we have given in Chapter 1 regarding the conversion of an exponentially successful strategy (martingale) into a selection rule (see Theorem 1.4.16, which can easily be adapted to the non-monotonic case). The rest of this section will be devoted to the proof of Theorem 2.2.31.

The splitting argument yields the following result:

Proposition 2.2.32. *Let $\alpha \in 2^\omega$ and s be such that $\mathrm{cdim}(\alpha) \leq s$. There exists a computable co-infinite $Z \subseteq \mathbb{N}$ such that, writing $\alpha = (\beta \oplus \beta') \oplus_Z \gamma$, we have:*

$$\mathrm{cdim}^\gamma(\beta) \leq s \quad \textit{and} \quad \mathrm{cdim}^\gamma(\beta') \leq s$$

(where cdim^γ is the dimension relative to the oracle γ).

Proof. First, we prove the following inequality (which was initially proven by Merkle et al. [47]):

$$\forall \beta', \beta'' \in 2^\omega \qquad \mathrm{cdim}(\beta' \oplus \beta'') \geq \frac{1}{2}\mathrm{cdim}(\beta') + \frac{1}{2}\mathrm{cdim}^{\beta'}(\beta'') \tag{2.2}$$

This equation follows from the sequence of inequalities:

$$\begin{aligned}
\mathrm{cdim}(\beta \oplus \beta'') &= \liminf_n \frac{K\big((\beta' \oplus \beta'')\restriction_{2n}\big)}{2n} \\
&= \liminf_n \frac{K(\beta'\restriction_n) + K(\beta''\restriction_n | \beta'\restriction_n) + O(\log n)}{2n} \\
&\geq \liminf_n \frac{K(\beta'\restriction_n) + K^{\beta'}(\beta''\restriction_n) + O(\log n)}{2n} \\
&\geq \frac{1}{2}\mathrm{cdim}(\beta') + \frac{1}{2}\mathrm{cdim}^{\beta'}(\beta'')
\end{aligned}$$

The first line comes from Theorem 2.2.27, the second from the symmetry of information for K (Theorem 2.1.21), the third from the easy fact that $K(w|\beta'\restriction_n) \leq K^{\beta'}(w) + O(\log n)$ for all w, and the last one from Theorem 2.2.27 again (and its relativization to β').

To obtain Proposition 2.2.32 from the inequality (2.2), let $\alpha \in 2^\omega$ and s such that $\mathrm{cdim}(\alpha) \leq s$. Let us set $t = \inf\{\mathrm{cdim}^\gamma(\beta) : \exists\ \mathrm{Z} \text{ recursive s.t. } \alpha = \beta \oplus_Z \gamma\}$. It is clear that $t \leq s$ (since s is attained for $Z = \emptyset$). We distinguish two cases:

Case 1: $t = s$. We then write $\alpha = \alpha' \oplus \alpha''$. By (2.2), we have:

$$\begin{aligned}
s = \mathrm{cdim}(\alpha) \geq \tfrac{1}{2}\mathrm{cdim}(\alpha') + \tfrac{1}{2}\mathrm{cdim}^{\alpha'}(\alpha'') \\
s = \mathrm{cdim}(\alpha) \geq \tfrac{1}{2}\mathrm{cdim}(\alpha'') + \tfrac{1}{2}\mathrm{cdim}^{\alpha''}(\alpha')
\end{aligned}$$

But by definition of t: $\mathrm{cdim}^{\alpha'}(\alpha'') \geq t = s$, and $\mathrm{cdim}^{\alpha''}(\alpha') \geq t = s$. Thus, $\mathrm{cdim}(\alpha') \leq s$ and $\mathrm{cdim}(\alpha'') \leq s$. We then get the desired result writing $\alpha = (\alpha' \oplus \alpha'') \oplus_\emptyset 0^\omega$.

Case 2: $t < s$. In this case, let $\beta, \gamma \in 2^\omega$ and Z recursive be such that $\alpha = \beta \oplus_Z \gamma$ and $\mathrm{cdim}^\gamma(\beta) \leq \frac{s+t}{2}$. Then write $\beta = \beta' \oplus \beta''$, and let X_1, X_2, Y_1, Y_2 be the recursive subsets of $\mathbb{N}$ such that $\alpha = \beta' \oplus_{X_1} (\beta'' \oplus_{X_2} \gamma)$ and $\alpha = \beta'' \oplus_{Y_1} (\beta' \oplus_{Y_2} \gamma)$. Relativizing (2.2) to the oracle γ, we get:

$$\begin{aligned}
\tfrac{s+t}{2} \geq \mathrm{cdim}^\gamma(\beta) \geq \tfrac{1}{2}\mathrm{cdim}^\gamma(\beta') + \tfrac{1}{2}\mathrm{cdim}^{(\beta' \oplus_{Y_2} \gamma)}(\beta'') \\
\tfrac{s+t}{2} \geq \mathrm{cdim}^\gamma(\beta) \geq \tfrac{1}{2}\mathrm{cdim}^\gamma(\beta'') + \tfrac{1}{2}\mathrm{cdim}^{(\beta'' \oplus_{X_2} \gamma)}(\beta')
\end{aligned}$$

By definition of t, $\mathrm{cdim}^{(\beta' \oplus_{Y_2} \gamma)}(\beta'') \geq t$ and $\mathrm{cdim}^{(\beta'' \oplus_{X_2} \gamma)}(\beta') \geq t$. Hence, $\mathrm{cdim}^\gamma(\beta') \leq s$ and $\mathrm{cdim}^\gamma(\beta'') \leq s$, as desired.

■

Once we have split our sequence α into two sequences of dimension at most $\dim(\alpha)$ (we forget about the oracle γ for now), we prove the existence of a strategy that succeeds exponentially fast on α:

Proposition 2.2.33. *Let $\beta', \beta'' \in 2^\omega$ and s such that $\mathrm{cdim}(\beta') \leq s$ and $\mathrm{cdim}(\beta'') \leq s$. There exists a strategy S, computable with oracle s, such that for all $t > s$, $\limsup \frac{V_n(\beta' \oplus \beta'', S)}{2^{(1-t)n}} = +\infty$.*

Proof. Up to relativizing everything to the oracle s, we can assume that s is computable. Let then $(s_k)_{k\in\mathbb{N}}$ be a computable decrasing sequence of rational numbers such that $s \leq s_k \leq s + 1/k$. Set for all k:

$$A_k = \{w \in 2^{<\omega} : \ K(w) \leq s_k|w| - 3k\}$$

(which is a c.e. subset of $2^{<\omega}$, uniformly in k). By Theorem 2.2.27, $\dim(\beta') \leq s$ precisely means that β' belongs to all $[A_k]$ (and the same for β''). Similarly to the proof of Theorem 1.5.15, the martingale

$$\sum_k \sum_{u\in A_k} 2^{(1-s_k)|u|+2k} d_u$$

t-succeeds on both β' and β'' for all $t > s$. However, this martingale is only left-c.e, so we need to come up with a more *effective* way to make money on $\beta' \oplus \beta''$. For all k, let t'_k (resp. t''_k) be the first stage of the enumeration of A_k such that $A_k[t'_k]$ contains a prefix of β' (resp. β''). Clearly, the sequence of t'_k (resp. t''_k) is computable with oracle β' (resp. β''). The key idea (introduced by Muchnik et al. [50]) is that either $t'_k \leq t''_k$ infinitely often, or $t''_k \leq t'_k$ infinitely often. Suppose for the rest of the proof that the first case holds. Then, with oracle β'', we get a sequence of time bounds t''_k such that β' has a prefix in $A_k[t''_k]$ for infinitely many k. Thus, the martingale

$$\sum_k \sum_{u\in A_k[t''_k]} 2^{(1-s_k)|u|+2k} d_u$$

is β''-computable and t-succeeds on β' for all $t > s$. Actually, in order to be completely rigorous, we will need a slight variation of this martingale. Let

$$d_1 = \sum_k \sum_{u\in A_k[t''_k]} 2^{(1-s_k)|u|+2k} d_u^{[k]}$$

where $d_u^{[k]}$ is the martingale which, like d_u, bets all its money on $u_{(i)}$ at stage i, but waits for the k-th stage to start playing. Formally:

$$d_u^{[n]}(w) = \begin{cases} 2^{-|u|} & \text{if } |w| \leq n \\ 2^{-|u|} d_u(w)/d_u(w\restriction_n) & \text{if } |w| \geq n, \text{ with convention } 0/0 = 0 \end{cases}$$

In particular, $d_u^{[n]}(u) \geq 2^{-n}$ for all n, u.

Lemma 2.2.34. *The above martingale d_1 is computable with oracle β'', it t-succeeds on β' for all $t > s$.*

Subproof. For all w and all k:

$$\begin{aligned} \sum_{u\in A_k[t''_k]} 2^{(1-s_k)|u|+2k} d_u^{[k]}(w) &\leq \sum_{u\in A_k} 2^{(1-s_k)|u|+2k}\, 2^{-|u|+|w|} \\ &\leq \sum_{u\in A_k} 2^{(-s_k)|u|+2k}\, 2^{|w|} \\ &\leq \sum_{u\in A_k} 2^{-K(u)-k} 2^{|w|} \\ &\leq 2^{-k}\, 2^{|w|} \end{aligned}$$

Hence, $d_1(w)$ is a series of terms which are computable with oracle β'' and decrease exponentially. Hence, $d_1(w)$ is computable with oracle β'' (uniformly in w). Moreover, by assumption, β' has a prefix in infinitely many of the $A_k[t''_k]$. If $w = \beta'\restriction_n$ is in $A_k[t''_k]$, we have:

$$d_1(w) \geq 2^{(1-s_k)|w|+2k} d_w^{[k]}(w) \geq 2^{(1-s_k)|w|+2k}\, 2^{-k} = 2^{(1-s_k)|w|+k}$$

Since this happens for infinitely many k, and since the s_k then to s, this proves that d_1 t-succeeds on β' for all $t > s$. □

We are now ready to construct the strategy S_1 satifying the conclusions of Proposition 2.2.33. S_1 plays against $\beta' \oplus \beta''$. It bets on the bits of β' according to the martingale d_1, and scans the bits of β'' when necessary to get the oracle information to compute d_1. More precisely, if S_1 has already applied d_1 on $\beta'\restriction_n$, it scans enough bits of β'' to compute $t''_0, \ldots, t''_{n+1}$. Then, it computes

$$d_1(\beta\restriction_n 0) - d_1(\beta\restriction_n) = \sum_{k=0}^{n+1} \sum_{u \in A_k[t''_k]} 2^{(1-s_k)|u|+2k} \left(d_u^{[k]}(\beta\restriction_n 0) - d_u^{[k]}(\beta\restriction_n) \right)$$

The first sum is bounded all the martingales of type $d_u^{[k]}$ with $k > n+1$ involved in the definition of d_1 satisfy $d_u^{[k]}(\beta\restriction_n 0) = d_u^{[k]}(\beta\restriction_n)$. Then, S_1 bets the fraction

$$\rho = \frac{d_1(\beta\restriction_n 0) - d_1(\beta\restriction_n)}{d_1(\beta\restriction_n)}$$

of its current capital that the value of the bit $\beta'_{(n)}$ is 0.

This way, we have for all n: $V_n(\beta' \oplus \beta'', S_1) = d_1(\beta'\restriction_n)$. Since d_1 t-succeeds on β' for all $t > s$, we have for all $t > s$:

$$\limsup \frac{V_n(\beta' \oplus \beta'', S_1)}{2^{(1-t)n}} = +\infty$$

And in the symmetric case where $t''_k \leq t'_k$ for infinitely many k, we use the strategy S_2 which uses the bits of β' to compute and apply on β'' the martingale

$$d_2 = \sum_k \sum_{u \in A_k[t'_k]} 2^{(1-s_k)|u|+2k} d_u^{[k]}$$

and for the same reason, we then have:

$$\limsup \frac{V_n(\beta' \oplus \beta'', S_2)}{2^{(1-t)n}} = +\infty$$

■

The last part of the proof is the following proposition:

Proposition 2.2.35. *If there exists a computable strategy S such that*

$$\limsup_n \frac{V_n(\alpha, S)}{2^{(1-s)n}} = +\infty$$

then there exists a non-monotonic selection rule σ, computable with oracle s, such that $\mathrm{Bias}(\sigma[\alpha]) \geq \delta$ *where δ is such that $\mathcal{H}\left(\frac{1}{2} + \delta\right) = s$.*

Proof. This is a straightforward adaptation of Theorem 1.4.16 to the non-monotonic case. Given a strategy s satifying the above proposition, let σ be the selection rule which visits the same bits as S, scanning whenever S scans and on the bits on which S bets, σ uses the compactness argument presented in the proof of Theorem 1.4.16, which allows it to select a subsequence of bias at least δ (where δ is such that $s = \mathcal{H}\left(\frac{1}{2} + \delta\right)$). ∎

Our toolbox is now complete, and we can prove Theorem 2.2.31:

Proof of Theorem 2.2.31. Let α be a sequence of constructive dimension at most s, and let $\delta \geq 0$ be such that $\mathcal{H}\left(\frac{1}{2} + \delta\right) = s$. Up to relativizing everything to s, suppose that s is computable. By Proposition 2.2.32, there exists a computable co-infinite $Z \subseteq \mathbb{N}$ such that, writing $\alpha = (\beta' \oplus \beta'') \oplus_Z \gamma$, we have:

$$\mathrm{cdim}^{(\gamma)}(\beta') \leq s \quad \text{and} \quad \mathrm{cdim}^{(\gamma)}(\beta'') \leq s$$

Relativizing Proposition 2.2.33 to the oracle γ, there exist two strategies S_1 and S_2 computable with oracle γ such that

$$\text{either } \limsup_n \frac{V_n(\beta' \oplus \beta'', S_1)}{2^{(1-s)n}} = +\infty \quad \text{or} \quad \limsup_n \frac{V_n(\beta' \oplus \beta'', S_2)}{2^{(1-s)n}} = +\infty$$

One can transform S_1 (resp. S_2) into a computable strategy S_1' (resp. S_2') which plays on $\alpha = (\beta' \oplus \beta'') \oplus_Z \gamma$ by following S_1 (resp. S_2) on the bits of positions $n \notin Z$ and scanning bits in positions $n \in Z$ whenever some bits of γ are needed to compute the action of S_1 (resp. S_2). Since no bet is made while scanning the bits of γ, this implies that $V_n(\alpha, S_1') = V_n(\beta' \oplus \beta'', S_1)$ and $V_n(\alpha, S_2') = V_n(\beta' \oplus \beta'', S_2)$ for all n. Hence:

$$\text{either } \limsup_n \frac{V_n(\alpha, S_1')}{2^{(1-s)n}} = +\infty \quad \text{or} \quad \limsup_n \frac{V_n(\alpha, S_2')}{2^{(1-s)n}} = +\infty$$

Hence, either S_1' or S_2' has the desired property. ∎

One can interpret Theorem 2.2.31 in terms of effective dimension:

Proposition 2.2.36. *For all sequence $\alpha \in 2^\omega$, there exists a non-monotonic selection rule, computable with oracle* $\mathrm{cdim}(\alpha)$*, which selects from α an infinite sequence β such that* $\dim_{\mathrm{comp}}(\beta) \leq \mathrm{cdim}(\alpha)$

Proof. Let $s = \mathrm{cdim}(\alpha)$ and $\delta \geq 0$ such that $\mathcal{H}\left(\frac{1}{2} + \delta\right) = s$. By Theorem 2.2.31, there exists a selection rule σ, computable with oracle s, which selects an infinite subsequence with bias at least δ. By Proposition 1.5.18, we have $\dim_{\mathrm{comp}}(\beta) \leq \mathcal{H}\left(\frac{1}{2} + \delta\right) = s$. ∎

2.3 Computable upper bounds of Kolmogorov complexity: a unifying concept

2.3.1 Motivation, definitions

We saw in the previous section that Kolmogorov complexity, whether plain or prefix, is not computable. This can be considered problematic for two reasons. First, one can question the use of Kolmogorov complexity as the central notion of *effective* randomness. Second, if we care about possible practical applications of Kolmogorov complexity, its non-computability is definitely a major obstacle. Can we overcome it? The Kolmogorov complexity of a string can be seen as the length of its shortest compressed form, although this shortest compressed form cannot be found effectively. What we can do is to give up on the hope to find the best compressed form, only hoping to find a "good one". After all, there exists rather good data compressors (gzip, bzip...) that we use all the time in our daily lives. Given one of those, one can get an approximation of C: for a string w run a compressor on it, and compute the length of the compressed output w'. This is only an *upper bound* of $C(w)$, but at least it is computable! This approach was followed by Cilibrasi and Vitanyi in their work on data clustering via a distance based on Kolmogorov complexity (see for example Cilibrasi and Vitanyi [14]).

In this section, we take a new look at algorithmic randomness where, instead of Kolmogorov complexity, we use *computable upper bounds* of Kolmogorov complexity as a primitive notion. One can expect that any computable upper bound approximates Kolmogorov complexity very poorly. As we will see, this is indeed the case. However, quite surprisingly, we will show that many notions of randomness can be characterized using computable upper bounds of Kolmogorov complexity, even those for which Kolmogorov complexity alone does not work (e.g. Schnorr and weak randomness). We start by the basic definition.

Definition 2.3.1. *A* COMPUTABLE UPPER BOUND (OR C.U.B) OF C *is a total computable function* $\widehat{C} : 2^{<\omega} \to \mathbb{N} \cup \{+\infty\}$ *such that* $C(w) \leq \widehat{C}(w) + O(1)$ *for all* $w \in 2^{<\omega}$. *A* COMPUTABLE UPPER BOUND OF K *is a total computable function* $\widehat{K} : 2^{<\omega} \to \mathbb{N} \cup \{+\infty\}$ *such that* $K(w) \leq \widehat{K}(w) + O(1)$ *for all* $w \in 2^{<\omega}$. *We denote by* $\mathfrak{C}$ *(resp.* $\mathfrak{K}$*) the class of computable upper bounds of* C *(resp. of* K*).*

Let us first notice that the Kraft-Chaitin theorem provides a simple characterization of c.u.b. for K.

Proposition 2.3.2. *A computable function* $f : 2^{<\omega} \to \mathbb{N}$ *is a c.u.b. for* K *if and only if*

$$\sum_{w \in 2^{<\omega}} 2^{-f(w)} < +\infty$$

Proof. If f is a c.u.b. for K then is particular $f \geq K - O(1)$ hence

$$\sum_{w \in 2^{<\omega}} 2^{-f(w)} \leq 2^{O(1)} \sum_{w \in 2^{<\omega}} 2^{-K(w)} < +\infty$$

Conversely, if $\sum_w 2^{-f(w)} < +\infty$, we can apply Corollary 2.1.24 to the set $\{(w, f(w)) : w \in 2^{<\omega}\}$ (which is computable hence in particular c.e.) and we get $K(w) \leq f(w) + O(1)$ for all $w \in 2^{<\omega}$. ■

We will use this proposition extensively in the sequel.

Like we said above, there is no computable upper bound for Kolmogorov complexity which is a good approximation for all strings.

Proposition 2.3.3. *Let $h : \mathbb{N} \to \mathbb{N}$ be a total computable increasing function. There exists no computable upper bound $\widehat{C} \in \mathfrak{C}$ of C such that $\widehat{C}(w) \leq h(C(w)) + O(1)$. The same statement holds with K in place of C.*

Proof. This proof is also inspired by Berry's paradox. Since $C \leq \widehat{C} + O(1)$, for each $k \in \mathbb{N}$, there are only finitely many strings w such that $\widehat{C}(w) \leq k$. Now, define a machine $M : 2^{<\omega} \to 2^{<\omega}$ as follows: for all $p \in 2^{<\omega}$ set $M(p)$ to be the first string w in the length-lexicographic order such that $\widehat{C}(w) \geq h(2|p|) + |p|$. Then, for all p, we have $C(M(p)) \leq |p| + O(1)$ (by Proposition 2.1.5), which implies that $\widehat{C}(M(p)) \leq h(C(M(p))) + O(1)$ is not true. The proof for K in place of C is identical. ■

What if we only want a computable upper bound which is good only for infinitely many strings? For plain complexity, this is very easy to do. It suffices to take $\widehat{C}(w) = |w|$, which works because most strings have a plain complexity which is close to their length. For prefix complexity this is less obvious since most strings of a given size n have prefix complexity close to $n + K(n)$. And if we want to get close to this, we need to find a good upper bound for $K(n)$, and we are back to square 1. Such an upper bound for K can nonetheless be constructed:

Proposition 2.3.4 (Solovay)**.** *There exists $\widehat{K} \in \mathfrak{K}$ such that $\widehat{K}(w) \leq K(w) + O(1)$ for infinitely many $w \in 2^{<\omega}$.*

Proof. Let S be the set

$$\{\langle p, w, t\rangle : \mathbb{V}(p) \text{ outputs } w \text{ in exactly } t \text{ steps}\}$$

where $\langle ., ., .\rangle$ is any computable encoding of $2^{<\omega} \times 2^{<\omega} \times \mathbb{N}$ into $2^{<\omega}$. The set S is clearly computable. For all $\langle p, w, t\rangle \in S$, we set $\widehat{K}(\langle p, w, t\rangle) = |p|$ and for all other strings $\widehat{K}$ takes the value $+\infty$. $\widehat{K}$ is computable since S is. Moreover, $\widehat{K}$ is a c.u.b. for K: if $\mathbb{V}(p)$ outputs w in exactly t steps of computation, then the knowledge of p is sufficient to retrieve w and t, thus $K(\langle p, w, t\rangle) \leq K(p) + O(1) \leq |p| + O(1)$ (the inequality $K(p) \leq |p| + O(1)$ holds because $p \in dom(\mathbb{V})$). Now, for all $w \in 2^{<\omega}$, let p_w be a shortest program of $\mathbb{V}$ which outputs w, and let t_w be the computation time of $\mathbb{V}(p_w)$. We have $K(p_w) \geq |p_w| - O(1)$. Indeed, from the shortest program q of $\mathbb{V}$ which outputs p_w, one can compute p_w and then w, thus $|p_w| = K(w) \leq |q| + O(1) = K(p_w) + O(1)$. Hence, we have for all w:

$$K(\langle p_w, w, t_w\rangle) \geq K(p_w) + O(1) \geq |p_w| + O(1) \geq \widehat{K}(\langle p_w, w, t_w\rangle) + O(1)$$

hence $\widehat{K}$ is as desired.

■

Another drawback of computable upper bounds of Kolmogorov complexity is that there is no optimal one. Indeed, let us define a preorder $\preceq$ on $\mathfrak{C}$ by

$$\forall\, \widehat{C}_1, \widehat{C}_2 \in \mathfrak{C} \qquad \left(\widehat{C}_1 \preceq \widehat{C}_2\right) \Leftrightarrow \left(\forall w \in 2^\omega\ \widehat{C}_1(w) \leq \widehat{C}_2(w) + O(1)\right)$$

and similarly an order $\preceq$ on $\mathfrak{K}$. A best computable upper bound of C (resp. of K) would be a minimal element in $(\mathfrak{C}, \preceq)$ (resp. $(\mathfrak{K}, \preceq)$) Such a minimal element does not exist:

Proposition 2.3.5. *$(\mathfrak{C}, \preceq)$ and $(\mathfrak{K}, \preceq)$ are lower semilattices with no minimal element.*

Here the term semilattice is a slight abuse of terminology since $\preceq$ is only a preorder: to make the statement completely rigorous one technically needs to consider the quotients of $\mathfrak{C}$ and $\mathfrak{K}$ by the equivalence relation induced by the conjunction of $\preceq$ and $\succeq$.

Proof. The fact that any two elements $\widehat{C}_1$ and $\widehat{C}_2$ have a greatest lower bound is easy: this greatest lower bound is $\min(\widehat{C}_1, \widehat{C}_2)$. It is computable, majorizes C up to an additive constant as $\widehat{C}_1$ and $\widehat{C}_2$ do, and is obviously greater than any function that is majorized by both $\widehat{C}_1$ and $\widehat{C}_2$. The proof that there is no minimal element is very similar to the one of Proposition 2.3.3. Let $\widehat{C}_1 \in \mathfrak{C}$. Let M be the total machine such that $M(p)$ is the shortest string w of length greater than $|p|$ satisfying $\widehat{C}_1(w) \geq 2|p|$. Then, $C_M(M(p)) \leq |p| \leq \widehat{C}_1(M(p))/2$ for all p. Now, notice that C_M is an upper bound for C (this is by the optimality theorem). We can also see that C_M is computable: for all p, $|M(p)| \geq |p|$ by definition. Hence, given a string w, one can compute $C_M(w)$ by computing $C_M(p)$ for all $|p| \leq |w|$ (which is possible because M is total computable) and returning the length of the shortest p such that $M(p) = w$ (or $+\infty$ if there is no such p). Let now $\widehat{C}_2 \in \mathfrak{C}$ be a lower bound for $\widehat{C}_1$ and C_M. We have $\widehat{C}_2 \prec \widehat{C}_1$ and the inequality in in fact strict since for all p: $\widehat{C}_2(M(p)) \leq C_M(M(p)) \leq \widehat{C}_1(M(p))/2$. ■

Since there is no optimal computable upper bound for Kolmogorov complexity, and since there is no good reason to study a particular one, the class of computable upper bounds will be considered in the sequel as a whole. That is, we will mainly make statements like: "for all computable upper bounds for C etc etc.". Notice that this approach does not allow us to give a reasonable definition of randomness for finite strings as for every string w there exists a computable upper bound $\widehat{C}$ such that $\widehat{C}(w) = 0$. Despite all this, computable upper bounds are enough to characterize various notions of randomness for infinite sequences, as we shall now see.

2.3.2 Some particular computable upper bounds

Before we move on to the discussion on how effective randomness can be described via computable upper bounds of Kolmogorov complexity, let us briefly review a few examples of how we can obtain computable upper bounds for Kolmogorov complexity.

Trimming the enumeration from above

As we saw earlier with (Proposition 2.1.11), the Kolmogorov complexity $C(w)$ of a string w is the limit of a nonincreasing sequence $(C(w)[t])_{t\in\mathbb{N}}$ where $C(w)[t]$ is uniformly computable in (t, w). To get a computable upper bound of C, it suffices to bound this enumeration from above by a computable function. Fix a total computable function $f : 2^{<\omega} \to \mathbb{N}$, and define $C^{[f]}$ for all $w \in 2^{<\omega}$ by

$$C^{[f]}(w) = C(w)[f(w)]$$

By definition, $C^{[f]}$ is total computable and is an upper bound for C since $C(w)[t] \geq C(w)$ for all (t, w). We define $K^{[f]}$ for all total computable functions f in the same fashion.

It turns out that this is a good way to build computable upper bounds for Kolmogorov complexity. Indeed, for any computable upper bound $\widehat{C}$, one can find a computable function f such that $C^{[f]}$ approximates C better than $\widehat{C}$ (and the same is true for prefix-complexity).

Proposition 2.3.6. *For every $\widehat{C} \in \mathfrak{C}$, there exists a total computable function $f : 2^{<\omega} \to \mathbb{N}$ such that $C^{[f]} \leq \widehat{C} + O(1)$. In other words, the class $\{C^{[f]} : f \text{ computable}\}$ is downward dense in $(\mathfrak{C}, \preceq)$. Similarly, the class $\{K^{[f]} : f \text{ computable}\}$ is downward dense in $(\mathfrak{K}, \preceq)$.*

Proof. Given some $\widehat{C} \in \mathfrak{C}$, and let $c > 0$ be such that $C \leq \widehat{C} + c$. Then, take

$$f : w \mapsto \min\{t : \ C(w)[t] \leq \widehat{C}(w) + c\}$$

It is clear that f is total computable and by definition, $C^{[f]}(w) \leq \widehat{C}(w) + c$ for all w. The proof is identical for prefix complexity. ∎

Approximation via compression

The approach of Cilibrasi and Vitanyi in [14] was to approximate Kolmogorov complexity by compression. As we want to study this approach from a theoretical point of view, we need to give a formal definition of a compressor. The intuitive understanding of a compressor is a procedure that maps a word to a code for that word, where the mapping is one-to-one and hence in principle invertible. For compressors that are to be applied in practice, in addition one will surely require that coding and decoding are efficient and that redundant sources will be mapped to reasonably short codes; however, these latter requirements will not be considered

here. We consider a most general notion of compressor where one simply requires that the compressor is a computable one-to-one function.

Definition 2.3.7. *A* COMPRESSOR *is a partial computable function* $\Gamma : 2^{<\omega} \to 2^{<\omega}$ *which is one-to-one and has computable domain. A compressor is said to be prefix-free if its range is prefix-free.*

Now, fix a compressor Γ. We claim that the function $C^\Gamma : 2^{<\omega} \to \mathbb{N} \cup \{+\infty\}$ defined by

$$C^\Gamma(w) = |\Gamma(w)|$$

(with the convention $C^\Gamma(w) = +\infty$ if $w \notin dom(\Gamma)$) is a c.u.b. for C. Indeed, it is computable since Γ is partial computable with computable domain. Moreover, let M be the machine that performs the decompression algorithm i.e. on an input p, M computes all the values $\Gamma(w)$ in parallel until it finds a u such that $\Gamma(u) = p$, and then outputs u. We have for all $w \in dom(\Gamma)$: $M(\Gamma(w)) = w$, hence $C_M(w) \leq |\Gamma(w)| = C^\Gamma(w)$. And by the optimality theorem, $C(w) \leq C_M(w)+O(1) \leq C^\Gamma(w)+O(1)$. And since we set $C^\Gamma(w) = +\infty$ for all $w \notin dom(\Gamma)$, this proves that $C \leq C^\Gamma + O(1)$. The same holds with K in place of C.

Proposition 2.3.8. *The class* $\{C^\Gamma : \Gamma$ *compressor*$\}$ *is downward dense in* $(\mathfrak{C}, \preceq)$. *The class* $\{C^\Gamma : \Gamma$ *prefix-free compressor*$\}$ *is downward dense in* $(\mathfrak{K}, \preceq)$.

Proof. Let $\widehat{C}$ be a c.u.b. for C and let $c > 0$ be such that $C(w) \leq \widehat{C}(w) + c$ for all $w \in 2^{<\omega}$. We build a compressor Γ as follows. Given $w \in 2^{<\omega}$, we run $\mathbb{U}$ in parallel on all the programs p such that $|p| \leq \widehat{C} + c$, until we find one (call it q) such that $\mathbb{U}(q) = w$ (such a q exists by definition of $\widehat{C}$ and c). Set $\Gamma(w) = q$. It is clear that Γ is total computable. Moreover, by definition, for all w: $|\Gamma(w)| \leq \widehat{C} + c$. The proof is the same for the prefix complexity case (with $\mathbb{V}$ in place of $\mathbb{U}$), and the inclusion $range(\Gamma) \subseteq dom(\mathbb{V})$ (which follows from the construction) ensures the prefix-freeness of $range(\Gamma)$. ■

Time-bounded Kolmogorov complexity

Another way to produce upper bounds for Kolmogorov complexity is to put a bound on the running times of programs.

Definition 2.3.9. *Let* $t : \mathbb{N} \to \mathbb{N}$ *be a function. The* TIME-BOUNDED PLAIN COMPLEXITY WITH BOUND t *is defined by:*

$$C^t(w) = \inf\{|p| : \mathbb{U}(p)[t(|w|)] \downarrow = w\}$$

and the TIME-BOUNDED PREFIX COMPLEXITY WITH BOUND t *by:*

$$K^t(w) = \inf\{|p| : \mathbb{V}(p)[t(|w|)] \downarrow = w\}$$

Remark 2.3.10. *For a fixed computable time bound t, the function C^t (resp. K^t) depends on the particular choice of $\mathbb{U}$, more than up to an additively bounded term. However, the statements we will make about time-bounded Kolmogorov complexity will* not *depend on $\mathbb{U}$ and $\mathbb{V}$.*

It is easy to see that for any computable time bound t, C^t and K^t are computable. Indeed, given a string w, one can compute $C(w)$ by computing $\mathbb{U}(p)$ during $t(|w|)$ steps for all programs p of length at most $t(|w|)$ and return the length of the shortest such p which outputs w within this time frame. The reason we can consider only programs of size at most $t(|w|)$ is that during $t(|w|)$ steps of computation, $\mathbb{U}$ can read at most $t(|w|)$ hence if it halts on a program q of length $l > t(|w|)$, then it also halts (returning the same value) on a prefix p of q of length at most $t(|w|)$ (corresponding to the bits of q read by $\mathbb{U}$ to compute $\mathbb{U}(q)$). The same proof works for K^t.

Proposition 2.3.11. *The class $\{C^t : t \text{ computable}\}$ is downward dense in $(\mathfrak{C}, \preceq)$. The class $\{K^t : t \text{ computable}\}$ is downward dense in $(\mathfrak{K}, \preceq)$.*

Proof. Let $\widehat{C} \in \mathfrak{C}$, and c a constant such that $C \leq \widehat{C} + c$. Let $n \in \mathbb{N}$. For each of the strings w_i $(0 \leq i < 2^n - 1)$ there exists a program p_i of length at most $\widehat{C}(w_i) + c$ such that $\mathbb{U}(p_i) = w_i$. $\widehat{C}$ being computable, the p_i can be found effectively. It then suffices to take $t(n)$ to be the maximum of the computation times of the $\mathbb{U}(p_i)$; we then have $C^t \leq \widehat{C} + c$. The same proof works for the K^t. ■

Decidable machines

Recall the definition of Kolmgorov complexity for a machine M:

$$C_M(w) = \min\{|p| : M(p) = w\}$$

The Kolmogorov complexity is defined as being $C_{\mathbb{U}}$ for some fixed optimal machine $\mathbb{U}$. The non-computability of C can be seen as a consequence of the undecidability of the halting problem for $\mathbb{U}$. In the above definition, suppose that the machine M has the particular following property:

Definition 2.3.12. *A machine M is said to be* DECIDABLE *if $dom(M)$ is a computable subset of $2^{<\omega}$.*

Suppose also that M is surjective i.e. there is no string w for which $C_M(w) = +\infty$. Then, the function C_M is computable. Indeed, given $w \in 2^{<\omega}$, one can sequentially compute all the $M(p)$ for $p \in dom(M)$ in order (which is possible because $dom(M)$ is computable) and we output the length of first p such that $M(p) = w$ (we know that we will eventually find one since M is surjective). This is exactly $C_M(w)$. And since by the optimality theorem, $C \leq C_M + O(1)$, C_M is a c.u.b. for C. And if M is decidable surjective and has prefix-free domain, K_M is a c.u.b. for K.

Remark 2.3.13. *Notice that the surjectivity of decidable machines is not a strong assumption, as we can easily turn a decidable machine into a surjective one without changing C_M by more than an additively bounded term. Indeed, given a decidable machine M, consider the machine M' defined by $M'(0p) = M(p)$ and $M'(1p) = p$ for all $p \in 2^{<\omega}$. Then, M' is decidable, surjective, and $C_{M'} \leq C_M + 1$*

Unlike the other classes of upper bounds presented above, Kolmogorov complexity with respect to surjective decidable machines is not downward dense in $(\mathfrak{C}, \preceq)$ or $(\mathfrak{K}, \preceq)$.

Proposition 2.3.14. *The class $\{C_M : M$ surjective decidable$\}$ is* not *downward dense in $(\mathfrak{C}, \preceq)$. The class $\{K_M : M$ surjective prefix-free decidable$\}$ is* not *downward dense in $(\mathfrak{K}, \preceq)$.*

Proof. We prove this for prefix complexity, using the particular c.u.b. $\widehat{K}$ for K constructed in the proof of Proposition 2.3.4. Recall that it is defined by $\widehat{K}(\langle p, w, t\rangle) = |p|$ if $\mathbb{V}(p)$ outputs w after exactly t steps of computation (and $\widehat{K}$ takes the value $+\infty$ on all other strings). Recall also the definition of the function B' (defined page 67): for all $k \in \mathbb{N}$, $B'(k)$ is the biggest integer n such that $K(n) \leq k$. Finally, recall that B' dominates every computable function (Proposition 2.1.10 adapted to K). Suppose for the sake of contradiction that there exists a prefix-free decidable machine M such that $K_M(w) \leq \widehat{K}(w) + c$ for some constant c and all w. For all n, set

$$l(n) = \max\{|M(p)| : \ p \in dom(M) \ \wedge \ |p| \leq n\}$$

It is clear that l is computable since $dom(M)$ is decidable, and by definition: if for some string w we have $|w| > l(n)$, then $K_M(w) > n$.

For all k, let p_k be a shortest program such that $\mathbb{V}(p_k) = B'(k)$ (notice that by definition of b' this implies $|p_k| \leq k$) and t_k be the computation time of $\mathbb{V}(p_k)$. We then have:

$$\begin{aligned} K_M(\langle p_k, B'(k), t_k\rangle) &\leq \widehat{K}(\langle p_k, B'(k), t_k\rangle) + c \\ &\leq |p_k| + c \\ &\leq k + c \end{aligned}$$

Hence, by definition of l:

$$\left|\langle p_k, B'(k), t_k\rangle\right| \leq l(k + c)$$

for all k. Hence, the function $k \mapsto l(k + c)$ dominates $k \mapsto |\langle p_k, B'(k), t_k\rangle|$, which is a contradiction since B' dominates every computable function. This finishes the proof for prefix complexity, and the proof for plain complexity is identical. ∎

However, surjective decidable machines satisfy a slightly weaker version of downward density:

Proposition 2.3.15. *For every $\widehat{C} \in \mathfrak{C}$ (resp. $\widehat{K} \in \mathfrak{K}$), and every computable order h, there exists a surjective decidable machine M (resp. a surjective prefix-free decidable machine M) such that $C_M(w) \leq \max(h(|w|), \widehat{C}(w)) + O(1)$ for all w (resp. $K_M(w) \leq \max(h(|w|), \widehat{K}(w)) + O(1)$ for all w).*

Proof. Let $\widehat{C} \in \mathfrak{C}$ and h a computable order. By Proposition 2.3.8, there exists a compressor Γ such that $|\Gamma(w)| \leq \widehat{C}(w) + O(1)$ for all w. We first transform Γ into a compressor Γ' which never compresses a string of length n by a string of length smaller than $h(n)$. To do so, it suffices to add some useless digits when Γ compresses a string too well. More precisely, set

$$\Gamma'(w) = \begin{cases} 0\Gamma(w) & \text{if } |\Gamma(w)| \geq h(|w|) \\ 1^{h(|w|)-\Gamma(w)}0\Gamma(w) & \text{if } |\Gamma(w)| < h(|w|) \end{cases}$$

Γ' is computable and is one-to-one. Hence, Γ' is a compressor. It moreover satisfies $\Gamma'(w)| = 1 + \max(h(|w|), \Gamma(w))$. We now build a decidable machine M which computes the inverse of Γ'. On an input $0q$, M computes $\Gamma(w)$ for all w in the domain of Γ such that $h(|w|) \leq |q|$ (there are finitely many such w), and if it finds among them a u such that $\Gamma(u) = q$, it outputs u (otherwise $M(0q)$ stays undefined). On an input 1^k0r, M computes $m = |r| + k$, and then computes $\Gamma(w)$ for all w in the domain of Γ such that $h(|w|) = m$ (there are finitely many such w), and if it finds among them a v such that $\Gamma(v) = r$, it outputs v (otherwise $M(!^k0r)$ stays undefined). It is easy to see that the domain of M' is decidable, and that $M(\Gamma'(w)) = w$ for all $w \in 2^{<\omega}$. This implies

$$C_M(w) \leq |\Gamma'(w)| \leq 1 + \max(h(|w|), |\Gamma(w)|) \leq \max(h(|w|), \widehat{C}(w)) + O(1)$$

It then suffices to turn M into a surjective decidable machine according to the technique described in Remark 2.3.13. This finishes the proof for plain complexity. The proof for prefix complexity is the same; just notice that if Γ is a prefix-free compressor, then Γ' is too. ■

2.3.3 Randomness via computable upper bounds

We now discuss how we can use computable upper bounds of Kolmogorov complexity to characterize various notions of randomness. We will be able to give such a characterization for Martin-Löf randomness, $\mathbf{0}'$-Martin-Löf randomness, Schnorr randomness, weak randomness, and computable dimension.

Martin-Löf randomness

The first natural attempt to define randomness via computable upper bounds of Kolmogorov complexity is to adapt the Levin-Schnorr characterization of Martin-Löf randomness in terms of Kolmogorov complexity: we can call "random" a sequence α such that for every c.u.b. $\widehat{K}$ of K we have $\widehat{K}(\alpha\restriction_n) \geq n - O(1)$. Quite surprisingly, this is equivalent to Martin-Löf randomness:

Theorem 2.3.16. *A sequence $\alpha \in 2^\omega$ is Martin-Löf random if and only if for all $\widehat{K} \in \mathfrak{K}$: $\widehat{K}(\alpha\restriction_n) \geq n - O(1)$*

Proof. One direction is easy: if α is Martin-Löf random, then by the Levin-Schnorr theorem we have $K(\alpha\restriction_n) \geq n - O(1)$ which a fortiori is true for any upper bound $\widehat{K}$ in place of K.

Suppose conversely that α is not Martin-Löf random. We will need the following lemma stating that every c.e open set, or even computable sequence of c.e. open sets, can be generated by a computable set of cylinders (instead of a c.e. set of cylinders):

Lemma 2.3.17. *For every computable family of c.e. open sets $(\mathcal{U}_n)_{n\in\mathbb{N}}$ of 2^ω, there exists a computable sequence of computable subsets $(A_n)_{n\in\mathbb{N}}$ of $2^{<\omega}$ such that for all n, $\mathcal{U}_n = [A_n]$. One can moreover assume that the A_n are prefix-free.*

Subproof. Given an effectively open subset $\mathcal{U}$ of 2^ω, generated by a c.e. subset A of $2^{<\omega}$ (that is, $\mathcal{U} = \bigcup_{w\in A}[w]$). We construct a c.e. subset A' of $2^{<\omega}$ as follows. Fix an enumeration of A. If at stage t a word u is enumerated into A, we enumerate all the extensions u'_i of u of length $t + |u|$ (for $0 \leq i \leq 2^t - 1$), which in particular implies $[u] = \bigcup_i [u'_i]$. This way, we have $[A'] = [A] = \mathcal{U}$. Moreover, A' is in fact a computable subset of $2^{<\omega}$. Indeed, to check if some word v belongs to A', it suffices to check, for every prefix u of v whether u is enumerated in A at stage $|v| - |u|$. This can be done effectively, and v is in A' if and only if such a prefix is found. It is easy to see that this construction can be done uniformly given an index for A, hence the result (the fact that the A_n can be taken to be prefix-free is obvious, see Remark 1.2.6) . □

To finish the proof of 2.3.16, since α is not Martin-Löf random, there exists a Martin-Löf test $(\mathcal{U}_n)_{n\in\mathbb{N}}$ such that $\alpha \in \bigcap_n \mathcal{U}_n$. By the above lemma, one can assume that the $\mathcal{U}_n$ are generated by a computable family $(A_n)_{n\in\mathbb{N}}$ of computable prefix-free subsets of $2^{<\omega}$. Let $\widehat{K}$ be defined as follows. For all $w \in 2^{<\omega}$, set $\widehat{K}(w) = |w| - k$ where k is the largest $l \leq |w|$ such that $w \in A_{2l}$, and set $\widehat{K}(w) = +\infty$ if there exists no such l. We then have:

$$\begin{aligned}
\sum_{w\in 2^{<\omega}} 2^{-\widehat{K}(w)} &= \sum_{k\in\mathbb{N}} \sum_{w\in A_{2k}} 2^{-|w|+k} \\
&\leq \sum_{k\in\mathbb{N}} 2^k \sum_{w\in A_{2k}} 2^{-|w|} \\
&\leq \sum_{k\in\mathbb{N}} 2^k \lambda(\mathcal{U}_{2k}) \\
&\leq \sum_{k\in\mathbb{N}} 2^k\, 2^{-2k} \\
&< +\infty
\end{aligned}$$

which proves that $\widehat{K}$ is indeed a c.u.b. for K (Proposition 2.3.2). And since α is in $[A_{2k}]$ for all k, this means that for all k, there exists n such that $\alpha\restriction_n \in A_{2k}$ and hence $\widehat{K}(\alpha\restriction_n) \leq n - k$.

■

It should be noted that in the above proof, we only use a *particular* upper bound $\widehat{K}$ which works for all $\alpha \notin \mathbf{MLR}$. Thus, we have in fact proven:

Proposition 2.3.18. *There exists $K^\star \in \mathfrak{K}$ such that, for any $\alpha \in 2^\omega$, the following are equivalent:*
(a) α is Martin-Löf random
(b) $K^\star(\alpha\restriction_n) \geq n - O(1)$

Remark 2.3.19. *There are in fact infinitely many functions $K^\star$ which make this equivalence true. They form an ideal in $(\mathfrak{K}, \preceq)$: if $K^{\star\star} \preceq K^\star$ is another element of $\mathfrak{K}$, the condition*

$$(b')\ K^{\star\star}(\alpha\restriction_n) \geq n - O(1)$$

is clearly implied by (a) and implies (b), and thus is equivalent to both.

Proposition 2.3.18 will play the key role in the proof of the Miller-Yu theorem. Remember that we postponed the part $(c) \Rightarrow (a)$ in the proof of Theorem 2.2.6. We now provide the proof of this result, which we slightly reformulate using the characterization of $\mathfrak{K}$ stated in Proposition 2.3.2.

Proposition 2.3.20. *If a sequence $\alpha \in 2^\omega$ is such that*

$$\forall \widehat{K} \in \mathfrak{K} \quad C(\alpha\restriction_n) \geq n - \widehat{K}(n) - O(1)$$

then α is Martin-Löf random.

Proof. Suppose α is not Martin-Löf random. For all $n, c \in \mathbb{N}$ define

$$A_n^c = \{u \in 2^{<\omega} : |u| = n \wedge K^\star(u) \leq |u| - 2c\}$$

and $a_n^c = \#A_n^c$. Notice that $a_n^c \leq 2^{n-2c}$ by a simple counting argument. Moreover, since $K^\star \geq K$: $\sum_n a_n\, 2^{-n}\, 2^{2c} \leq 1$, and hence $\sum_c \sum_n a_n^c\, 2^{-n}\, 2^c \leq 1$. Set for all n, $b_n = \sum_c a_n^c 2^c$. Since $a_n^c \leq 2^{n-2c}$ for all (n, c), b_n is computable (as the sum of an exponentially decreasing computable series) uniformly in n. One has $\sum_n b_n 2^{-n} \leq 1$. Thus, by an elementary inversion of summations argument, $\sum_n (b_1 + ... + b_n) 2^{-n} < +\infty$. Let then G be the (computable) function defined by $G(n) = n - \log(b_1 + ... + b_n)$, which, by the previous inequality, satisfies $\sum 2^{-G(n)} < +\infty$.

Now, let w be of length n and such that $K^\star(w) \leq |w| - 2c$. In other words, w belongs to A_n^c, hence can be described by c together with its position in the enumeration of $\bigcup_n A_n^c$ where for all n, the elements of A_n^c are enumerated before those of A_{n+1}^c (this can be done since the A_n^c's are uniformly computable). Hence:

$$\begin{aligned} C(w) &\leq 2\log c + \log(a_1^c + a_2^c ... + a_n^c) + O(1) \\ &\leq 2\log c + \log(2^{-c} b_1 + 2^{-c} b_2 + ... + 2^{-c} b_n) + O(1) \quad \text{(definition of } b_n\text{)} \\ &\leq 2\log c - c + n - G(n) + O(1) \quad \text{(definition of } G\text{)} \end{aligned}$$

By Proposition 2.3.18, and definition of the A^c_n, as α is not Martin-Löf random, then for arbitrarily large c, for infinitely many n, $\alpha\restriction_n \in A^c_n$ which by the above discussion implies $C(\alpha\restriction_n) \leq 2\log c - c + n - G(n) + O(1)$. This finishes the proof. ■

It should be noticed that in this proof, we also implicitly use a computable upper bound for C. Indeed, given w of length n, let $\widehat{C}(w)$ be the minimum value of $2\log c - c + n - G(n) + O(1)$ over the c such that $w \in A^c_n$ (this is effectively computable). The above proof tells us that if α is not Martin-Löf random, then $\widehat{C}(\alpha\restriction_n) \leq n - G(n) - O(1)$ fails to hold. Moreover, this $\widehat{C}$ works for any $\alpha \notin \mathbf{MLR}$. Calling $C^\star$ this c.u.b. and $K^\star$ the above function G, we can sum up the different characterizations of Martin-Löf randomness *à la* Miller-Yu:

Theorem 2.3.21. *There exist $C^\star \in \mathfrak{C}$ and $K^\star \in \mathfrak{K}$ such that the following are equivalent for all $\alpha \in 2^\omega$:*
(a) α is Martin-Löf random
(b) $C(\alpha\restriction_n) \geq n - K(n) - O(1)$
(c) $C^\star(\alpha\restriction_n) \geq n - K^\star(n) - O(1)$

In fact, there are nine statements of that sort that we can make, using all the possible combinations of $C/C^\star/(\text{for all } \widehat{C} \in \mathfrak{C})$ and $K/K^\star/(\text{for all } \widehat{K} \in \mathfrak{K})$. Since the above (b) is the strongest and (c) the weakest of such statements and since they are both equivalent to being Martin-Löf random, all the nine statements express Martin-Löf randomness. For example, it is true that α is Martin-Löf random if and only if $(\forall \widehat{C} \in \mathfrak{C})\ (\forall \widehat{K} \in \mathfrak{K})\ \ \widehat{C}(\alpha\restriction_n) \geq n - \widehat{K}(n) - O(1)$.

Notice also that, like in Remark 2.3.19, there are infinitely many functions $C^\star$ and $K^\star$ which make the equivalence $(a) \Leftrightarrow (b) \Leftrightarrow (c)$ of Theorem 2.3.21 true, and they form an ideal in $(\mathfrak{C} \times \mathfrak{K}, \preceq \times \preceq)$. Indeed, for all $(C^{\star\star}, K^{\star\star}) \in \mathfrak{C} \times \mathfrak{K}$ such that $C^{\star\star} \preceq C^\star$ and $K^{\star\star} \preceq K^\star$, the condition

$$(c')\ C^{\star\star}(\alpha\restriction_n) \geq n - K^{\star\star}(n) - O(1)$$

is implied by (b) and implies (c), hence is equivalent to both. This justifies our choice to use the same name $K^\star$ both in Proposition 2.3.18 and Theorem 2.3.21. We can indeed assume that these functions are the same: if $K^\star_1$ works for Proposition 2.3.18 and $K^\star_2$ works for Theorem 2.3.21, then $K^\star_0 = \min(K^\star_1, K^\star_2)$ works for both.

Quite surprisingly, even Kolmogorov randomness, which we now know to be equivalent to $\mathbf{0}'$-Martin-Löf randomness, can be characterized by computable upper bounds.

Theorem 2.3.22. *There exists $C^\star \in \mathfrak{C}$ such that the following are equivalent for every sequence $\alpha \in 2^\omega$:*
(a) α is $\mathbf{0}'$-Martin-Löf random
(b) $(\forall \widehat{C} \in \mathfrak{C})\ \ \widehat{C}(\alpha\restriction_n) \geq n - O(1)$ for infinitely many n
(c) $C^\star(\alpha\restriction_n) \geq n - O(1)$ for infinitely many n

Proof. $(a) \Rightarrow (b)$ follows directly from Theorem 2.2.11. $(b) \Rightarrow (c)$ is trivial. To see that $(c) \Rightarrow (a)$, it suffices to notice that the proof of the the part $(\Rightarrow)$ of Theorem 2.2.11, we implicitly use a computable upper bound for C. Indeed, we use the function $C^\star$ such that $C^\star(w)$ is the minimum over $k \leq |w|$ of $|w| - k + 2\log k$ for those k such that $[w] \subseteq [A_k(|w|)]$ (we can set by convention $C^\star(w) = +\infty$ if there exists no such k). $C^\star$ is computable as $[A_k(|w|)]$ is a clopen set uniformly computable in k, w.

■

Remark 2.3.23. *Here again, the functions $C^\star$ making the equivalence of Theorem 2.3.22 true form an ideal in $(\mathfrak{C}, \preceq)$.*

We mentionned earlier (page 63) the very recent result of Miller who showed that a sequence α is $\mathbf{0}'$-Martin-Löf random if and only if $K(\alpha\restriction_n) \geq n + K(n) - O(1)$ for infinitely many n. The following theorem is an analogue of Miller's result in terms of computable upper bounds (and is not a direct consequence of Miller's theorem).

Theorem 2.3.24. *There exists $K^* \in \mathfrak{K}$ such that the following are equivalent for all $\alpha \in 2^\omega$:*
(a) α is $\mathbf{0}'$-Martin-Löf random
(b) $(\forall \widehat{K} \in \mathfrak{K})$ $\widehat{K}(\alpha\restriction_n) \geq n + K^*(n) - O(1)$ for infinitely many n
(c) $K^*(\alpha\restriction_n) \geq n + K^*(n) - O(1)$ for infinitely many n

Proof. Let $\widehat{K}_0$ be a computable upper bound of K such that $\widehat{K}_0(w) \leq K(w) + O(1)$ for infinitely many strings w (whose existence is asserted by Proposition 2.3.4). For all $n \in \mathbb{N}$, we effectively find a t_n such that:

- for all $w \in 2^{<\omega}$ of length n, $K(w)[t_n] \leq \widehat{K}_0(w)$
- for all $w \in 2^{<\omega}$ of length n, $K(w)[t_n] \leq C^\star(w) + \widehat{K}(w) + e$ where $C^\star$ is the c.u.b. for C of Theorem 2.3.22 and e is a constant such that $K(w) \leq C(w) + K(|w|) + e$ for all $w \in 2^{<\omega}$ (whose existence is asserted by Proposition 2.1.27).
- $(\forall k \leq n) \quad \#\{w \in 2^{<\omega} : \ |w| = n \wedge K(w)[t_n] \leq n + K(n)[t_n] - k\} \leq 2^{n-k+e'}$ where e' the constant hidden in the $O(2^{n-k})$ in Proposition 2.1.27(ii).

Such a t_n exists because $K(w)[t]$ tends to $K(w)$ as t tends to infinity, and if $K(w)[t] = K(w)$, the three conditions are satisfied: the first one because $K \leq \widehat{K}$, the second one because $K(w) \leq C(w) + K(|w|) + c \leq C^\star(w) + \widehat{K}(w) + c$ for all $w \in 2^{<\omega}$, and the third one by Proposition 2.1.27). The fact that $n \mapsto t_n$ is computable comes from the computability of $K(.)[.]$ and $\widehat{K}$.

Finally, set for all $w \in 2^{<\omega}$ of length n:

$$K^*(w) = K(w)[t_n]$$

Let us now check that K^* works.

$(a) \Rightarrow (b)$. Suppose that (b) fails, i.e. there exists some $\widehat{K} \in \mathfrak{K}$ such that $\widehat{K}(\alpha\restriction_n) - n - K^*(n)$ tends to $-\infty$. For all integers n, c, set

$$A^c_n = \{w \in 2^{<\omega} : \ |w| = n \wedge \widehat{K}(w) \leq n + K^*(n) - c - e'\}$$

As usual, we have $\alpha \in \bigcap_c \liminf_n([A^c_n])$ and the A^c_n are uniformly computable in c, n (since $\widehat{K}$ and K^* are computable. We claim that for any c, there are infinitely many n such that $\lambda([A^c_n]) \leq 2^{-c}$. This is because by construction, we have $K^* \leq \widehat{K}_0$ and for infinitely many n, $\widehat{K}_0(n) \leq K(n)$. For all those n, we have $K^*(n) \leq K(n)$ and hence

$$\begin{array}{rcl} \#(A^c_n) & = & \#\{w \in 2^{<\omega} : \ |w| = n \wedge \widehat{K}(w) \leq n + K^*(n) - c - e'\} \\ & \leq & \#\{w \in 2^{<\omega} : \ |w| = n \wedge K(w) \leq n + K(n) - c - e'\} \\ & \leq & 2^{n-c} \end{array}$$

Hence, we can apply Proposition 2.2.12 to the A^c_n, which tells us that α is not $\mathbf{0}'$-Martin-Löf random.

$(b) \Rightarrow (c)$ is trivial.

$(c) \Rightarrow (a)$. Suppose that α is not $\mathbf{0}'$-Martin-Löf random. Then by Theorem 2.3.22 the quantity $C^\star(\alpha\restriction_n) - n$ tends to $-\infty$. By construction, for all w, $K^*(w) \leq C^\star(w) + K^*(|w|) + c$, hence $K^*(\alpha\restriction_n) - n - K^*(n) \leq C(\alpha\restriction_n) - n + c$, which tends to $-\infty$. This proves that (c) fails to hold.

■

Remark 2.3.25. *The set of $K^* \in \mathfrak{K}$ making the equivalence in Theorem 2.3.24 work is downward dense in $(\mathfrak{K}, \preceq)$. Indeed, given some $\widehat{K} \in \mathfrak{K}$, it suffices to replace in the above proof "$K(w)[t_n] \leq \widehat{K}_0(w)$" by "$K(w)[t_n] \leq \min(\widehat{K}_0(w), \widehat{K}(w))$" in the definition of t_n. We do not know whether the set of such K^* form an ideal.*

Schnorr randomness and weak randomness

Theorem 2.3.26. *The following are equivalent for every sequence $\alpha \in 2^\omega$:*
(a) α is Schnorr random
(b) For every $\widehat{K} \in \mathfrak{K}$ and every computable order h: $\widehat{K}(\alpha\restriction_n) \geq n - h(n) - O(1)$

Proof. $(a) \Rightarrow (b)$ Suppose that (b) does not hold, i.e. there exists a computable order h and a c.u.b. $\widehat{K}$ of K such that $\widehat{K}(\alpha\restriction_n) \leq n - h(n)$ for infinitely many n. For all k, set

$$D_k = \{w \in 2^{<\omega} : h(|w|) = k \wedge \widehat{K}(w) \leq |w| - k\}$$

D_k is uniformly computable since $\widehat{K}$ and h; it is finite since h is an order (the number of values n for which $h(n) = k$ is finite for all k). Moreover, by definition, $\alpha \in [D_k]$ for all k. Finally, we have $\lambda([D_k]) \leq 2^{-k}$ by Lemma 2.2.4. Hence, using Lemma 1.5.9, we have proven that α is not Schnorr random.

$(b) \Rightarrow (a)$ Suppose conversely that α is not Schnorr random. Using Lemma 1.5.9 again, there exists a uniformly computable sequence $(D_k)_{k\in\mathbb{N}}$ of finite subsets of $2^{<\omega}$ such that $\lambda([D_k]) \leq 2^{-k}$ for all k, and $\alpha \in [D_k]$ for infinitely many k. Using the (now) standard trick of taking all the extensions of a string of a certain length to generate the same open cylinder, we can assume that for all k, the elements of D_k have the same length $f(k)$ and we can also assume that f is increasing. Now, define $\widehat{K}$ by $\widehat{K}(w) = |w| - k/2$ where k is the largest $l \leq |w|$ such that $w \in D_l$, and $\widehat{K}(w) = +\infty$ if there exists no such l. Clearly $\widehat{K}$ is computable and by the same kind of argument as in the proof of Theorem 2.3.16, we have $\sum_w 2^{-\widehat{K}(w)} < +\infty$, hence $\widehat{K}$ is a c.u.b. for $\widehat{K}$. By assumption, for infinitely many k, $\alpha \in [D_k]$ i.e. there exists an n such that $\alpha\restriction_n \in D_k$. For such k, n, we have $\widehat{K}(\alpha\restriction_n) \leq n - k/2$. But we also know that $n = f(k)$ or equivalently $k = f^{-1}(n)$. In other words, $\widehat{K}(\alpha\restriction_n) \leq n - f^{-1}(n)/2$ for infinitely many n. Since f is increasing $f^{-1}/2$ is an order, which completes the proof.

■

A corollary of this is that the Kolmogorov complexity of the prefix of length n of a not Schnorr random sequence gets further and further from n with computable speed.

Corollary 2.3.27. *If α is not Schnorr random, there exists a computable order h such that $K(\alpha\restriction_n) \leq n - h(n)$ for infinitely many n.*

Proof. This is exactly the part $(b) \Rightarrow (a)$ in the above theorem, together with the fact that $K \leq \widehat{K}$. ■

We saw above that Schnorr randomness and weak randomness had dual characterizations in terms of martingales (Theorem 1.5.10 and Theorem 1.5.12). The same kind of duality holds for their characterizations in terms of c.u.b. of K:

Theorem 2.3.28. *The following are equivalent for every sequence $\alpha \in 2^\omega$:*
(a) α is weakly random
(b) For every $\widehat{K} \in \mathfrak{K}$ and every computable order h: $\widehat{K}(\alpha\restriction_n) \geq n - h(n) - O(1)$ for infinitely many n.

Proof. The proof is almost exactly the same as the previous one: just use Lemma 1.5.13 in place of Lemma 1.5.9 and replace all the "infinitely many" by "for all". ■

And as a corollary of this theorem, we get:

Corollary 2.3.29. *If α is not weakly random, there exists a computable order h such that $K(\alpha\restriction_n) \leq n - h(n)$ for all n.*

Computable dimension

Finally, we present a characterization of computable dimension in terms of computable upper bounds.

Theorem 2.3.30. *For every sequence $\alpha \in 2^\omega$:*

$$\dim_{\mathrm{comp}}(\alpha) = \inf_{\widehat{K} \in \mathfrak{K}} \liminf_{n \to +\infty} \frac{\widehat{K}(\alpha\restriction_n)}{n}$$

Proof. First suppose that $\dim_{\mathrm{comp}}(\alpha) < s$ for some $s \in \mathbb{Q}$. By definition, there exists a computable s-test $(A_n)_{n\in\mathbb{N}}$ such that $\alpha \in [A_n]$ for all n. We set for all string w: $\widehat{K}(w) = s|w| - k/2$ where k is the largest $l \leq |w|$ such that $w \in A_l$ and $\widehat{K}(w) = +\infty$ if there exists no such l. $\widehat{K}$ is computable and

$$\begin{aligned}
\sum_{w \in 2^{<\omega}} 2^{-\widehat{K}(w)} &= \sum_{k\in\mathbb{N}} \sum_{w \in A_k} 2^{-s|w|+k/2} \\
&\leq \sum_{k\in\mathbb{N}} 2^{k/2} \sum_{w\in A_k} 2^{-s|w|} \\
&\leq \sum_{k\in\mathbb{N}} 2^{k/2}\, 2^{-k} \quad \text{by definition of a } s\text{-test} \\
&< +\infty
\end{aligned}$$

Hence $\widehat{K}$ is a c.u.b. for K, and since $\alpha \in [A_k]$ for all k, we have

$$\forall k\, \exists n\ \widehat{K}(\alpha\restriction_n) \leq sn - k/2 \qquad \text{and thus} \qquad \liminf_{n\to+\infty} \frac{\widehat{K}(\alpha\restriction_n)}{n} \leq s$$

Conversely, suppose that for some $\widehat{K}$ and some $s \in \mathbb{Q}$ one has

$$\liminf_{n\to+\infty} \frac{\widehat{K}(\alpha\restriction_n)}{n} < s$$

We set for all $k \in \mathbb{N}$:

$$A_k = \{w \in 2^{<\omega} :\ \widehat{K}(w) \leq s|w| - k\}$$

Clearly, $\alpha \in [A_k]$ for all k. We claim that the A_k form a computable s-test. First, they are uniformly computable since $\widehat{K}$ is. Second, we have for all k:

$$\sum_{w\in A_k} 2^{-s|w|} \leq \sum_{w\in A_k} 2^{-\widehat{K}(w)-k} \leq 2^{-k} \sum_{w\in A_k} 2^{-K(w)} \leq 2^{-k}$$

This proves that $\dim_{\mathrm{comp}}(\alpha) \leq s$. ∎

Randomness via particular upper bounds

All the results of this section can be rephrased in terms of the particular upper bounds we presented above (via compression, computable time bounds, decidable machines, etc.), thanks to their "downward density" property (see Proposition 2.3.8, Proposition 2.3.11, Proposition 2.3.15). Here are a few examples of such reformulations:

Lemma 2.3.31. *A sequence $\alpha \in 2^\omega$ is Martin-Löf random if and only if for every (surjective) prefix-free decidable machine M, $K_M(\alpha\restriction_n) \geq n - O(1)$*

Proof. If α is Martin-Löf random, then $K(\alpha\restriction_n) \geq n - O(1)$, hence a fortiori $K_M(\alpha\restriction_n) \geq n - O(1)$. Conversely, if α is not Martin-Löf random, then by Theorem 2.3.16, there exists $\widehat{K} \in \mathfrak{K}$ such that $\widehat{K}(\alpha\restriction_n) - n$ takes on arbitrarily large negative values. By Proposition 2.3.15, there exists a surjective decidable machine M such that $K_M(w) \leq \max(|w|/2, \widehat{K}(w)) + O(1)$ for all n. Then, $K_M(\alpha\restriction_n) - n \leq \max(-n/2, \widehat{K}(\alpha\restriction_n) - n) + O(1)$. Hence, $K_M(\alpha\restriction_n) - n$ takes on arbitrarily large negative values. □

Lemma 2.3.32. *There exists a compressor Γ and a prefix-free compressor Ξ such that a sequence α is Martin-Löf random if and only if $C^\Gamma(\alpha\restriction_n) \geq n - K^\Xi(n) - O(1)$.*

Proof. By Theorem 2.3.21, there exist $C^\star \in \mathfrak{C}$ and $K^\star \in \mathfrak{K}$ such that α is Martin-Löf random if and only if $C^\star(\alpha\restriction_n) \geq n - K^\star(n) - O(1)$. By Proposition 2.3.8, let Γ be a compressor such that $C^\Gamma \leq C^\star + O(1)$ and Ξ a prefix-free compressor such that $K^\Xi \leq K^\star + O(1)$. We can conclude by the remark made in page 91: any pair $\widehat{C}, \widehat{K} \in \mathfrak{C} \times \mathfrak{K}$ such that $\widehat{C} \leq C^\star + O(1)$ and $\widehat{K} \leq K^\star + O(1)$ also works for the equivalence in Theorem 2.3.21. □

Lemma 2.3.33. *A sequence α is Schnorr random if and only if for every computable time bound t and every computable order h, $K^t(\alpha\restriction_n) \geq n - h(n) - O(1)$.*

Proof. If α is Schnorr random, then by Theorem 2.3.26, for all $\widehat{k} \in \mathfrak{K}$, for al computable orders h, $\widehat{K}(\alpha\restriction_n) \geq n - h(n) - O(1)$. This is in particular true for the K^t, with t computable. Conversely, if α is not Schnorr random, there exists $\widehat{K} \in \mathfrak{K}$ and a computable order h such that $\widehat{K}(\alpha\restriction_n) - n + h(n)$ takes on arbitrarily large negative values. Take a computable time bound t such that $K^t \leq \widehat{K} + O(1)$ (which is possible according to Proposition 2.3.11). Then, $K^t(\alpha\restriction_n) - n + h(n)$ takes on arbitrarily large negative values. □

The time-bounded version of Theorem 2.3.22 was already stated in Nies et al. [51], and the time-bounded version of Theorem 2.3.30 appeared in Hitchcock [23].

Chapter 3

Randomness for computable measures

In the previous chapters, we have defined several notions of effective randomness for Lebesgue measure. Most of them can be extended to arbitrary computable Borel probability measures, except for the stochasticity notions since they rely on a frequency approach which does not make sense for some measures.

In probability theory, two measures are said to be equivalent if they have the same nullsets. From the effective randomness viewpoint, we can effectivize this idea by saying that two measures are "effectively equivalent" if they have the same random sequences. Hence, every notion of randomness induces an equivalence relation on computable probability measures. This chapter studies the different implications between these equivalence relations. We first focus on a particular set of measures, the so-called generalized Bernoulli measures. For this class, Kakutani's theorem gives a criterion for classical equivalence. We extend Kakutani's theorem to all effective equivalence relations (part of this task was already done in the work of Vovk and Muchnik), and we use it to separate Kolmogorov-Loveland stochasticity from other randomness notions.

The rest of the chapter is devoted to the classification of effective equivalence relations in the case of arbitrary computable probability measure.

3.1 Extending notions of randomness to computable measures

In this chapter, we want to extend some of the notions of randomness we have seen so far (Martin-Löf randomness, computable randomness, Schnorr randomness and weak randomness) to a larger class of computable measures. Since their definitions all involve Borel sets (G_δ for the first three, open and closed sets for weak randomness), we will only consider Borel probability measures. The following theorem, which is a particular case of Caratheodory's extension theorem (see for example [4]), asserts that any such measure can be constructed by specifying only the measure of each cylinder.

Theorem 3.1.1. *Let $m : 2^{<\omega} \to \mathbb{R}_+$ be a function such that $m(\epsilon) = 1$ and for all $w \in 2^{<\omega}$, $m(w) = m(w0) + m(w1)$. There exists a unique Borel probability measure μ on 2^ω such that for all $w \in 2^{<\omega}$, $\mu([w]) = m(w)$.*

In the sequel, we will often identify a measure μ with the function $w \mapsto \mu([w])$ and we will abbreviate $\mu([w])$ by $\mu(w)$

A property which we used repeatedly in the previous chapters for Lebesgue measure is that for a Lebesgue-measurable set $\mathcal{X} \subseteq 2^\omega$, $\lambda(\mathcal{X})$ is the infimum, over the open sets $\mathcal{U}$ containing $\mathcal{X}$, of $\lambda(\mathcal{U})$. Equivalently, $\lambda(\mathcal{X})$ is the supremum, over the closed sets $\mathcal{C}$ contained in $\mathcal{X}$, of $\lambda(\mathcal{U})$. Any measure having this property is called REGULAR. A classical theorem of measure theory asserts that on any locally compact Hausdorff space that has a countable basis, all Borel measures are regular (see for example Cohn [15]). Since 2^ω is metric compact and does have a countable basis (the cylinders $[w]$ for all $w \in 2^{<\omega}$ for example):

Theorem 3.1.2 (Regularity of Borel measures)**.** *Every Borel measure μ on 2^ω is regular, i.e. for all $\mathcal{X} \subseteq 2^\omega$:*

$$\mu(\mathcal{X}) = \inf \{\mu(\mathcal{U}) : \mathcal{U} \text{ open} \wedge \mathcal{X} \subseteq \mathcal{U}\}$$

Since this is the only type of measures we will consider in this chapter, we will abbreviate "Borel probability measure on the Cantor space" by "measure".

The central notions studied in this chapter are *equivalence* and *consistency*.

Definition 3.1.3. *Let μ and ν be two measures. They are said to be* EQUIVALENT*, which we write $\mu \sim \nu$, if they have the same nullsets, i.e. for all $\mathcal{X} \subseteq 2^\omega$, $\mu(\mathcal{X}) = 0 \Leftrightarrow \nu(\mathcal{X}) = 0$. They are said to be* consistent *if there is no set $\mathcal{X} \subseteq 2^\omega$ such that $\mu(\mathcal{X}) = 0$ and $\nu(\mathcal{X}) = 1$.*

The following lemma will be a useful criterion for equivalence and consistency:

Lemma 3.1.4. *Let μ and μ' be two measures and c a positive constant. If for all $\varepsilon > 0$ there exists a set $\mathcal{X} \subseteq 2^\omega$ such that $\mu(\mathcal{X}) < \varepsilon$ and $\nu(\mathcal{X}) > c - \varepsilon$, then there exists a set $\mathcal{Y} \subseteq 2^\omega$ such that $\mu(\mathcal{Y}) = 0$ and $\nu(\mathcal{Y}) \geq c$.*

Proof. Under this assumption, for all n there exists a set $\mathcal{X}_n$ such that $\mu(\mathcal{X}_n) < 2^{-n}$ and $\mu'(\mathcal{X}_n) > c - 2^{-n}$ Consider the set

$$\mathcal{Y} = \limsup_n(\mathcal{X}_n) = \bigcap_{n\in\mathbb{N}} \bigcup_{k>n} \mathcal{X}_k$$

($\mathcal{Y}$ is the set of sequences α that belong to infinitely many $\mathcal{X}_n$). For all n, we have:

$$\mu\left(\bigcup_{k>n} \mathcal{X}_k\right) \leq \sum_{k>n} \mu(\mathcal{X}_k) \leq \sum_{k>n} 2^{-k} \leq 2^{-n}$$

and thus $\mu(\mathcal{Y}) = 0$. On the other hand, for all n:

$$\nu\left(\bigcup_{k>n} \mathcal{X}_k\right) \geq \nu(\mathcal{X}_{n'}) \geq c - 2^{-n'} \qquad \text{for all } n' > n$$

Hence $\nu\left(\bigcup_{k>n} \mathcal{X}_k\right) \geq c$. Since the sequence $\left(\bigcup_{k>n} \mathcal{X}_k\right)$ is nonincreasing in k, the measure of their intersection $\mathcal{Y}$ is the limit of their measures, which are all greater or equal to c. Thus, $\nu(\mathcal{Y}) \geq c$. □

Definition 3.1.5. *A* COMPUTABLE MEASURE *on the Cantor space 2^ω is a measure μ such that the function $w \mapsto \mu(w)$ is computable.*

With this definition, we can canonically extend the notions of randomness relying on tests. A μ-MARTIN-LÖF TEST (resp. a μ-SCHNORR TEST) will be a computable sequence $(\mathcal{U}_n)_{n\in\mathbb{N}}$ of c.e. open sets such that $\mu(\mathcal{U}_n) \leq 2^{-n}$ for all n (resp. $\mu(\mathcal{U}_n) = 2^{-n}$ for all n). To extend computable randomness, we first need to generalize the notion of martingale.

Definition 3.1.6. *Let μ be a measure. A* μ-MARTINGALE *is a function $d : 2^{<\omega} \to \mathbb{R}_+$ such that for all $w \in 2^{<\omega}$:*

$$d(w)\mu(w) = d(w0)\mu(w0) + d(w1)\mu(w1)$$

Like in the Lebesgue case, a μ-martingale represents the capital of a player who bets money on the values of the bits of α, where the relation between $d(w)$, $d(w0)$, and $d(w1)$ ensures that the game is fair if α is chosen at random w.r.t. the probability measure μ. In this game, during the n-th move, having read the $n-1$ first bits $w = \alpha\restriction_{n-1}$ of α, the player bets a fraction $\rho \in [0, 1]$ of his current capital on a value for the bit $\alpha_{(n)}$. Suppose for example that he guesses that the value of α_n will be 1. If his bet is wrong, he loses his stake, i.e. $d(w0) = (1-\rho)d(w)$. If his guess is correct, he wins his stake multiplied by a fairness factor: $d(w1) = \left(1 + \rho\frac{\mu(w0)}{\mu(w1)}\right) d(w)$. Hence, in general there is an assymetry between the potential gain and the potential loss at each move (this was not the case for Lebesgue measure) which we will need to take into account when manipulating μ-martingales. In particular, we need to adapt the notation Bet/Stake of page 13. We do this as follows:

Let d be a μ-martingale for some measure μ. For any $u \in 2^{<\omega}$, let

$$\text{Stake}^+(d, u) = d(u) - \min(d(u0), d(u1)) \qquad \text{and} \qquad \text{Bet}^+(d, u) = \frac{\text{Stake}^+(d, u)}{d(u)}$$

where we let $\text{Bet}^+(d, u) = 0$ in case $d(u) = 0$. Furthermore, let

$$\text{Guess}(d, u) = \begin{cases} 0 & \text{if } d(u0) \geq d(u1) \\ 1 & \text{if } d(u0) < d(u1) \end{cases}$$

Guess tells us on which value d bets, Stake$^+$ is the amount of money he bets on this value (as opposed to the Stake function, we have Stake$^+(d,u) \geq 0$ for all u) and Bet$^+$ is the fraction of the capital Stake$^+$ represents.

Like in the case of Lebesgue measure, we say that a μ-martingale d succeeds on a sequence α if $\limsup\ d(\alpha\restriction_n) = +\infty$.

Remark 3.1.7. *Ville's inequality (Theorem 1.5.1) holds for any measure μ: the μ-measure of the set of sequences α such that $\sup_n\ d(\alpha\restriction_n) \geq c$ is bounded by $1/c$ (the proof is a straightforward adaptation of the uniform case). Hence the success set of a μ-martingale has μ-measure 0.*

We can now call μ-COMPUTABLY RANDOM as sequence $\alpha \in 2^\omega$ such that no computable μ-martingale succeeds on α.

Remark 3.1.8. *It is easy to check that the characterizations of Martin-Löf, Schnorr and weak randomness still hold for arbitrary computable measures: a sequence α is μ-Martin-Löf random iff no left-c.e. μ-martingale succeeds on α; α is μ-Schnorr random (resp. μ-weakly random) iff there exists no computable μ-martingale d and no computable order h such that $d(\alpha\restriction_n) \geq h(n)$ for infinitely many n (resp. for all n).*

Proposition 3.1.9. *If μ and ν are two measures, the function $d : 2^{<\omega} \to \mathbb{R}_+$ defined by*

$$d(w) = \frac{\mu(w)}{\nu(w)}$$

is a normed ν-martingale.
Conversely, for any normed ν-martingale d, there exists a measure μ such that the above equation is satisfied for all w.

Remark 3.1.10. *In the above proposition, it may be the case that $\nu(w) = 0$ and $\mu(w) > 0$. In that case, we set by convention $d(w) = \mu(w)/\nu(w) = +\infty$, meaning that $d = \mu/\nu$ immediately succeeds. If $\mu(w) = 0$ and $\nu(w) = 0$ we set $d(w) = \mu(w)/\nu(w) = 1$.*

Proof. The first part is just a simple computation. For the second part, it suffices to check that given a measure ν and a normed ν-martingale d, the product $\mu = d\,\nu$ is a measure (this also is an easy computation). ■

The ν-martingale μ/ν has a natural intuitive meaning which is a generalization of our discussion on page 16. If we are betting on the values of a sequence α. Your opponent (the bank/casino) thinks α has been chosen at random according to the probability ν. However, we have some inside information which tells us that α was in fact chosen at random with respect to some other measure μ. The ν-martingale $d = \mu/\nu$ represents the best betting strategy we can use to take advantage of this information, in the sense that if α is indeed chosen at random according to μ, then d majorizes on α any other ν-martingale (up to a multiplicative constant):

Lemma 3.1.11. *Let μ and ν be two measures, d the ν-martingale $d = \mu/\nu$, and d' be some other ν-martingale. We have, with μ-probability 1:*

$$d'(\alpha\restriction_n) = O\Big(d(\alpha\restriction_n)\Big)$$

Proof. Let μ, ν, d and d' be as above. The "up to a multiplicative constant" in the statement allows us to assume that d' is normed. Hence, by Proposition 3.1.9, there exists a measure ξ such that $d' : \xi/\nu$. For all n, we have:

$$d'(\alpha\restriction_n) = \frac{\xi(\alpha\restriction_n)}{\nu(\alpha\restriction_n)} = \frac{\xi(\alpha\restriction_n)}{\mu(\alpha\restriction_n)}\frac{\mu(\alpha\restriction_n)}{\nu(\alpha\restriction_n)} = \left(\frac{\xi(\alpha\restriction_n)}{\mu(\alpha\restriction_n)}\right) d(\alpha\restriction_n)$$

By Proposition 3.1.9, ξ/μ is a μ-martingale. Hence, by Ville's inequality, with μ-probability 1, one has $\xi(\alpha\restriction_n)/\mu(\alpha\restriction_n) = O(1)$. ■

This correspondence between measures and martingales will be omnipresent in this chapter, as we will frequently see things from a game-theoretic viewpoint.

3.2 Generalized Bernoulli measures

We begin our study by considering a particular class of measures: generalized Bernoulli measures. This is perhaps the most natural class of Lebesgue-like measures: while Lebesgue measure on 2^ω corresponds to a random trial where all the bits are chosen independently with probability distribution $(1/2, 1/2)$, a generalized Bernoulli measure corresponds to a trial where all the bits are chosen independently, but possibly with different probability distributions (the term "generalized" distinguishing them from Bernoulli measures, for which all the bits are chosen independently and all have the same probability distribution $(p, 1-p)$ for some p).

3.2.1 Definition

The formal definition of generalized Bernoulli measures is the following.

Definition 3.2.1. *The* GENERALIZED BERNOULLI MEASURE *of parameter $(p_i)_{i\in\mathbb{N}}$ is the measure μ such that for all $w \in 2^{<\omega}$:*

$$\mu(w) = \prod_{\substack{i<|w|\\ w_i=0}} p_i \prod_{\substack{i<|w|\\ w_i=1}} (1-p_i)$$

This measure μ is said to be STRONGLY POSITIVE *if there exists some $\eta > 0$ such that $p_i \in [\eta, 1-\eta]$ for all i.*

Remark 3.2.2. *It is easy to see that a generalized Bernoulli measure of parameter $(p_i)_{i\in\mathbb{N}}$ is computable if and only if the p_i form a computable sequence of computable real numbers.*

3.2.2 Kakutani's theorem

Theorem 3.2.3 (Kakutani [28]). *Let μ and ν be two strongly positive generalized Bernoulli measures of respective parameters (p_i) and (q_i).*
If $\sum_i (p_i - q_i)^2 < +\infty$, then μ and ν are equivalent. If $\sum_i (p_i - q_i)^2 = +\infty$, then μ and ν are inconsistent.

Proof. We only prove the second part of the theorem for now since we will get the first part as a corollary of stronger results later on (see page 112). We follow the proof given in Muchnik et al. [50]. Suppose μ and μ' are two generalized Bernoulli measures of respective parameters $(p_i)_{i\in\mathbb{N}}$ and $(q_i)_{i\in\mathbb{N}}$. By the strong positivity hypothesis, let ε be such that $\varepsilon \leq p_i \leq 1-\varepsilon$ and $\varepsilon \leq q_i \leq 1-\varepsilon$

Let ν be the the generalized Bernoulli measure of parameter $(\frac{p_i+q_i}{2})_{i\in\mathbb{N}}$. We have for all i:

$$\left(\frac{p_i+q_i}{2}\right)^2 = p_i q_i \left(1 + \frac{(p_i-q_i)^2}{4p_iq_i}\right) \geq p_i q_i \left(1 + \frac{(p_i-q_i)^2}{4\varepsilon^2}\right)$$

and

$$\begin{aligned}\left(\frac{(1-p_i)+(1-q_i)}{2}\right)^2 &= (1-p_i)(1-q_i)\left(1 + \frac{(p_i-q_i)^2}{4(1-p_i)(1-q_i)}\right)\\ &\geq (1-p_i)(1-q_i)\left(1 + \frac{(p_i-q_i)^2}{4\varepsilon^2}\right)\end{aligned}$$

Thus, for all $w \in 2^{<\omega}$ of length n:

$$\nu(w)^2 \geq \mu(w)\mu'(w) \prod_{i=0}^{n-1} \left(1 + \frac{(p_i-q_i)^2}{4\varepsilon^2}\right)$$

The limit when n tends to $+\infty$ of the product term in the above equation is $+\infty$ (because $\sum_i (p_i-q_i)^2 = +\infty$), hence for any fixed $M > 0$ one can find N such that $\nu(w)^2 \geq M\mu(w)\mu'(w)$ for all w of length N. Let A be the set of strings w of length N such that $\mu(w) \geq \mu'(w)$. For any such string w, we have $\nu(w)^2 \geq M\mu'(w)^2$ i.e. $\nu(w) \geq \sqrt{M}\mu'(w)$. Thus, $\mu'([A]) \leq \nu([A])/\sqrt{M} \leq 1/\sqrt{M}$. On the other hand, for all strings w of length N that is not in A, we have $\nu(w)^2 \geq M\mu(w)^2$ hence by the same argument we have $\mu(2^\omega \setminus [A]) \leq 1/\sqrt{M}$. Since we can find such an $[A]$ for arbitrarily large M, applying Lemma 3.1.4 (with $c = 1$), this proves that μ and μ' are inconsistent. ∎

3.2.3 Constructive versions of Kakutani's theorem

An effective version of Kakutani's theorem was obtained by V. Vovk.

Theorem 3.2.4 (Vovk [60]). *Let μ and ν be computable strongly positive generalized Bernoulli measures of respective parameters $(p_i)_{i\in\mathbb{N}}$ and $(q_i)_{i\in\mathbb{N}}$.*
If $\sum_i (p_i - q_i)^2 < +\infty$, then $\mu\mathbf{MLR} = \nu\mathbf{MLR}$. If $\sum_i (p_i - q_i)^2 = +\infty$, then $\mu\mathbf{MLR} \cap \nu\mathbf{MLR} = \emptyset$.

One of the main results of this chapter is that Vovk's theorem can be extended to computable randomness, to Schnorr randomness, and (almost) to weak randomness:

Theorem 3.2.5. *Let μ and ν be computable strongly positive generalized Bernoulli measures of parameter $(p_i)_{i\in\mathbb{N}}$ and $(q_i)_{i\in\mathbb{N}}$, respectively.*

(i) The following are equivalent.

(a) $\sum_i (p_i - q_i)^2 < +\infty$
(b) μ *and* ν *are consistent*
(c) $\mu \sim \nu$
(d) $\mu\mathbf{MLR} = \nu\mathbf{MLR}$
(e) $\mu\mathbf{CR} = \nu\mathbf{CR}$
(f) $\mu\mathbf{SR} = \nu\mathbf{SR}$
(g) $\mu\mathbf{WR} = \nu\mathbf{WR}$

(ii) The following are equivalent.

(a) $\sum_i (p_i - q_i)^2 = +\infty$
(b) μ *and* ν *are inconsistent*
(c) μ *and* ν *are not equivalent*
(d) $\mu\mathbf{MLR} \cap \nu\mathbf{MLR} = \emptyset$
(e) $\mu\mathbf{CR} \cap \nu\mathbf{CR} = \emptyset$
(f) $\mu\mathbf{SR} \cap \nu\mathbf{SR} = \emptyset$
(g) $\mu\mathbf{SR} \cap \nu\mathbf{WR} = \mu\mathbf{WR} \cap \nu\mathbf{SR} = \emptyset$

Remark 3.2.6. *The item (g) of part (ii) cannot be strengthened to $\mu\mathbf{WR} \cap \nu\mathbf{WR}$. In fact, given any two computable strongly positive generalized Bernoulli measures μ and ν, we* always *have $\mu\mathbf{WR} \cap \nu\mathbf{WR} \neq \emptyset$. This is because the class* **WG** *of generic sequences is included in both (the argument we used for Proposition 1.8.3 works for any computable strongly positive generalized Bernoulli measure).*

Some of the implications of this theorem hold for any pair (μ, ν) of computable measures. These will be proven in the next section (Theorem 3.3.1 and Theorem 3.3.2). For now, we will prove the implications of this theorem that are specific to the Bernoulli case, waiting for the next section to give the full proof (page 112).

Proposition 3.2.7. *Let μ and ν be two computable strongly positive generalized Bernoulli measures of parameter $(p_i)_{i\in\mathbb{N}}$ and $(q_i)_{i\in\mathbb{N}}$, respectively, such that $\sum_i (p_i - q_i)^2 < +\infty$. Let d be a computable μ-martingale and α be a sequence such that $\limsup d(\alpha\restriction_n) = +\infty$. Then there exists a (not necessarily normed) computable ν-martingale d' such that for all n it holds that $\ln d(\alpha\restriction_n) \leq d'(\alpha\restriction_n) + O(1)$.*

Proof. We first consider the ν-martingale that satisfies

$$\mathrm{Guess}(d', u) = \mathrm{Guess}(d, u) \qquad \text{and} \qquad \mathrm{Stake}^+(d', u) = \mathrm{Bet}^+(d, u)\ .$$

for all $u \in 2^{<\omega}$. This means than whenever d bets a fraction ρ of its capital on a value of a bit, d' bets an amount ρ on that value (e.g. if d bets 10% of its capital, d' bets 0.1 units of money).

For the moment, assume that d' may incur debts, i.e., d' is allowed to take negative values. We will see later on how to modify d' to make it positive while still successful on α.

For ease of notation, let $\rho_n = \mathrm{Bet}^+(d, \alpha\restriction_n)$. For each n, there are three cases:

	$d(\alpha\restriction_{n+1})/d(\alpha\restriction_n)$	$d'(\alpha\restriction_{n+1}) - d'(\alpha\restriction_n)$
$\mathrm{Guess}(d, \alpha\restriction_n) \neq \alpha_n$	$1 - \rho_n$	$-\rho_n$
$\mathrm{Guess}(d, \alpha\restriction_n) = \alpha_n = 0$	$1 + \rho_n \frac{1-p_n}{p_n}$	$\rho_n \frac{1-q_n}{q_n}$
$\mathrm{Guess}(d, \alpha\restriction_n) = \alpha_n = 1$	$1 + \rho_n \frac{p_n}{1-p_n}$	$\rho_n \frac{q_n}{1-q_n}$

By letting x_n be equal to $-\rho_n$, or $\rho_n \frac{p_n}{1-p_n}$, or $\rho_n \frac{1-p_n}{p_n}$ in the three different cases, respectively, the entries in the table above can be rewritten as follows.

	$d(\alpha\restriction_{n+1})/d(\alpha\restriction_n)$	$d'(\alpha\restriction_{n+1}) - d'(\alpha\restriction_n)$
$\mathrm{Guess}(d, \alpha\restriction_n) \neq \alpha_n$	$1 + x_n$	x_n
$\mathrm{Guess}(d, \alpha\restriction_n) = \alpha_n = 0$	$1 + x_n$	$x_n\big(1 + \frac{q_n - p_n}{p_n(1-q_n)}\big)$
$\mathrm{Guess}(d, \alpha\restriction_n) = \alpha_n = 1$	$1 + x_n$	$x_n\big(1 + \frac{p_n - q_n}{q_n(1-p_n)}\big)$

By induction it follows for all n, that

$$d(\alpha\restriction_n) = \prod_{k=0}^{n-1}(1 + x_k) \; .$$

By strong positivity, choose $\eta > 0$ such that all p_i and q_i are contained in the interval $[\eta, 1-\eta]$ and let $m = \lceil \eta^{-1} \rceil$. Then by definition all x_n are in the interval $[-1, m]$. Let c be a constant such that for all $x \in [-1, m]$, $\ln(1 + x) \leq x - c\,x^2$.

$$\ln d(\alpha\restriction_n) = \ln \prod_{k=0}^{n-1}(1 + x_k) = \sum_{k=0}^{n-1} \ln(1 + x_k) \leq \sum_{k=0}^{n-1} x_k - \sum_{k=0}^{n-1} c\,x_k^2 \; . \tag{3.1}$$

Concerning the martingale d', for all n and for all three cases discussed above, we have

$$x_n - m^2 |x_n||p_n - q_n| \leq d'(\alpha\restriction_{n+1}) - d'(\alpha\restriction_n) \; ,$$

hence, by induction, it follows for all n that

$$\sum_{k=0}^{n-1} x_k - \sum_{k=0}^{n-1} m^2 |x_k||p_k - q_k| \leq d'(\alpha\restriction_n) \; . \tag{3.2}$$

Next let $c' = \sqrt{\sum_{i=0}^{\infty}(p_i - q_i)^2}$ and choose a constant c'' such that for all $t \geq 0$,

$$m^2 c' \sqrt{t} \leq ct + c'' \; .$$

Then we obtain by the Cauchy-Schwarz inequality

$$\begin{aligned}
\sum_{k=0}^{n-1} m^2 |x_k||p_k - q_k| &\leq m^2 \sqrt{\sum_{k=0}^{n-1}(p_k - q_k)^2} \sqrt{\sum_{k=0}^{n-1} x_k^2} \\
&= m^2 c' \sqrt{\sum_{k=0}^{n-1} x_k^2} \\
&\leq c \left(\sum_{k=0}^{n-1} x_k^2 \right) + c''
\end{aligned}$$

Together with (3.1) and (3.2), this yields

$$\ln d(\alpha\restriction_n) \leq d'(\alpha\restriction_n) + c'' .$$

Recall that up to now, we have assumed that d' is a normed ν-martingale of a special type that is allowed to incur debts. Furthermore, by assumption, we have

$$\lim_n d(\alpha\restriction_n) = +\infty , \quad \text{hence} \quad \lim_n d'(\alpha\restriction_n) = +\infty ,$$

and thus there is a natural number M such that $-M < d'(\alpha\restriction_n)$ for all n. Hence, $d'' = d' + \max(c'', M+1)$ is a ν-martingale that for all n satisfies $1 \leq d''(\alpha\restriction_n)$ and $\ln d(\alpha\restriction_n) \leq d''(\alpha\restriction_n)$. Finally let d''' be the ν-martingale that $\mathrm{Guess}(d''', u) = \mathrm{Guess}(d'', u)$ for all u, and $\mathrm{Stake}^+(d''', u) = \mathrm{Stake}^+(d'', u)$ if $d''(u) \geq 1$, and $\mathrm{Stake}^+(d''', u) = 0$ otherwise. The ν-martingale d''' is then positive on every $u \in \mathbb{N}$ and by definition, for all n:

$$d'''(\alpha\restriction_n) = d''(\alpha\restriction_n) \geq \ln d(\alpha\restriction_n)$$

The ν-martingale d''' is as desired. ■

Corollary 3.2.8. *Let μ and ν be two computable strongly positive generalized Bernoulli measures of parameter $(p_i)_{i\in\mathbb{N}}$ and $(q_i)_{i\in\mathbb{N}}$, respectively, such that $\sum_i (p_i - q_i)^2 < +\infty$. Then $\mu\mathbf{CR} = \nu\mathbf{CR}$, $\mu\mathbf{SR} = \nu\mathbf{SR}$ and $\mu\mathbf{WR} = \nu\mathbf{WR}$.*

Proof. Let us first prove this for computable randomness. Let $\alpha \notin \mu\mathbf{CR}$. By definition, there exists a computable μ-martingale d such that $\limsup_n d(\alpha\restriction_n) = +\infty$. By Remark 1.4.8 (which can be easily adapted to the setting of arbitrary computable measures), we can assume that $\lim_n d(\alpha\restriction_n) = +\infty$. By Proposition 3.2.7, there exists a computable ν-martingale d' such that $d'(\alpha\restriction_n) \geq \ln d(\alpha\restriction_n)$ which in particular implies $\lim_n d'(\alpha\restriction_n)$. Hence, $\alpha \notin \nu\mathbf{CR}$. By symmetry, we have proven that $\mu\mathbf{CR} = \nu\mathbf{CR}$. The argument is almost the same for Schnorr randomness: if $\alpha \notin \mu\mathbf{SR}$, as we said in Remark 3.1.8 there exists a computable μ-martingale d and a computable order h such that $d(\alpha\restriction_n) \geq h(n)$ for infnitely many n. By Proposition 3.2.7, there exists a computable ν-martingale d' such that $d'(\alpha\restriction_n) \geq \ln d(\alpha\restriction_n)$ for all n, which in particular implies $d'(\alpha\restriction_n) \geq \ln h(n)$ for infinitely many n. And since $n \mapsto \ln h(n)$ is a computable order, this proves $\alpha \notin \nu\mathbf{SR}$. The proof for weak randomness is the same as for Schnorr randomness, just replacing "infinitely many n" by "for all n". ■

3.2.4 Applications to stochasticity

The effective version of Kakutani's theorem turns out very useful in the study of stochasticity. The central theorem in this direction is the following.

Theorem 3.2.9 (Shen [54], after van Lambalgen [57])**.** *Let μ be a computable strongly positive generalized Bernoulli measure of parameter $(p_i)_{i\in\mathbb{N}}$, such that the p_i tend to $1/2$. Then $\mu\mathbf{MLR} \subseteq \mathbf{KLStoch}$.*

To prove this theorem, we will use the following result of classical probability theory (for a proof, see for example Alon and Spencer [1]).

Lemma 3.2.10 (Azuma's inequality)**.** *Let $\mathbf{X}_0, \ldots, \mathbf{X}_{n-1}$ be random variables taking their values in $\mathbb{R}$ such that for all k, $|\mathbf{X}_{k+1} - \mathbf{X}_k| \leq c$ and $\mathrm{E}(\mathbf{X}_{k+1}|\mathbf{X}_0, \ldots, , \mathbf{X}_k) = \mathbf{X}_k$. For all $n \in \mathbb{N}$ and $m > 0$, we have the following inequality:*

$$\Pr\Big[|\mathbf{X}_n - \mathbf{X}_0| > m\Big] < 2 \exp\left(\frac{-m^2}{2nc^2}\right)$$

Proof of Theorem 3.2.9. Let μ be a computable generalized Bernoulli measure of parameter $(p_i)_{i\in\mathbb{N}}$ where the p_i tend to $1/2$. Let σ be a computable non-monotonic selection rule. We pick a sequence α at random with respect to the probability measure μ and we run σ on α. Let $i_0, i_1, \ldots$ be the positions of the bits selected from α by σ. The i_k are random variables of parameter α which take their values in $\mathbb{N} \cup \{\perp\}$, where $i_k = \perp$ means that the subsequence selected from α by σ has length less than k.

We define a sequence of random variables $(\mathbf{X}_n^\sigma)_{n\in\mathbb{N}}$ by setting $\mathbf{X}_0^\sigma = 0$ and for all n:

$$\mathbf{X}_{n+1}^\sigma = \begin{cases} \#0(\alpha_{i_0} \ldots \alpha_{i_n}) - \sum_{k=0}^n p_{i_k} & \text{if } i_n \neq \perp \\ \mathbf{X}_n^\sigma \text{ otherwise} \end{cases}$$

By definition of μ, the sequence $(\mathbf{X}_n^\sigma)_{n\in\mathbb{N}}$ satisfies the hypotheses of Azuma's inequality, with $c = 1$. This allows us to prove:

Lemma 3.2.11. *If α is μ-Martin-Löf random we have*

$$\lim_{n\to+\infty} \frac{\mathbf{X}_n^\sigma}{n} = 0$$

Subproof. Suppose this is not the case, i.e. there exists some rational $\delta > 0$ such that $|\mathbf{X}_n^\sigma(\alpha)| \geq \delta n$ for infinitely many n. In other words, for infinitely many n, α belongs to $\mathcal{U}_n$ where we define for all n:

$$\mathcal{U}_n = \{\beta \in 2^\omega : \ |\mathbf{X}_n^\sigma(\beta)| \geq \delta n\}$$

By Azuma's inequality (with $c = 1$), for all n, $\mu(\mathcal{U}_n) \leq 2\exp\left(\frac{-\delta^2 n}{2}\right)$ This implies that $\sum_n \mu(\mathcal{U}_n) < +\infty$. Moreover, it is easy to see that the $\mathcal{U}_n$ are open sets, and that they are computably enumerable (uniformly in n) as σ is computable. Thus, by Theorem 1.3.4 (which can easily be adapted to any computable measure), no μ-Martin-Löf random sequence can belong to infinitely many $\mathcal{U}_n$, contradicting the fact that α is μ-Martin-Löf random. □

To conclude the proof of Theorem 3.2.9, let α by μ-Martin-Löf random. Let σ be a computable non-monotonic selection rule that selects an infinite subsequence from α. We have by the above lemma:

$$\lim_{n\to+\infty} \frac{\mathbf{X}_n^\sigma(\alpha)}{n} = 0$$

i.e.

$$\lim_{n\to+\infty}\left(\frac{\#0(\alpha_{i_0}\dots\alpha_{i_n})}{n}-\frac{\sum_{k=0}^{n}p_{i_k}}{n}\right)=0 \tag{3.3}$$

But since the p_i tend to $1/2$:

$$\lim_{n\to+\infty}\frac{\sum_{k=0}^{n}p_{i_k}}{n}=\frac{1}{2} \tag{3.4}$$

By (3.3) and (3.4), we have:

$$\lim_{n\to+\infty}\frac{\#0(\alpha_{i_0}\dots\alpha_{i_n})}{n}=\frac{1}{2}$$

which precisely means that the subsequence selected from α by σ satisfies the Law of Large Numbers. This being true for all computable non-monotonic selection rules that select an infnite subsequence from α, we have proven that α is Kolmogorov-Loveland stochastic.

■

As it turns out, Shen's theorem even allows us to separate λ-weak randomness from Kolmogorov-Loveland stochasticity. Indeed, Merkle et al. [47] constructed a particular computable generalized Bernoulli measure of parameter $(p_i)_{i\in\mathbb{N}}$ such that the p_i tend to $1/2$ but $\sum_i(p_i-\frac{1}{2})^2=+\infty$, whose Martin-Löf random elements are not λ-weakly random (but are Kolmogorov-Loveland stochastic by Shen's theorem). In fact, thanks to Theorem 3.2.5, we can strengthen this result: any generalized Bernoulli measure of this type works!

Corollary 3.2.12. *Let μ be a computable generalized Bernoulli measure of parameter $(p_i)_{i\in\mathbb{N}}$ such that the p_i tend to $1/2$ and $\sum_i(p_i-1/2)^2=+\infty$. Then $\mu\mathbf{MLR}\subseteq\mathbf{KLStoch}$ and $\mu\mathbf{MLR}\cap\lambda\mathbf{WR}=\emptyset$. This in particular means that KL-stochasticity does not imply λ-weak randomness.*

Proof. We have just proven that $\mu\mathbf{MLR}\subseteq\mathbf{KLStoch}$. To see that $\mu\mathbf{MLR}\cap\lambda\mathbf{WR}=\emptyset$, apply Theorem 3.2.5 (part (ii), implication $(a)\Rightarrow(g)$) to μ and $\nu=\lambda$ to prove that $\mu\mathbf{SR}\cap\lambda\mathbf{WR}=\emptyset$. A fortiori, $\mu\mathbf{MLR}\cap\lambda\mathbf{WR}=\emptyset$. ■

Using Azuma's inequality, we can also prove that the bound $\mathcal{H}\left(\frac{1}{2}+\delta\right)$ in Theorem 2.2.31 is optimal:

Theorem 3.2.13. *Let α be a μ-Martin-Löf random sequence, where μ is the generalized Bernoulli measure of constant parameter $\frac{1}{2}+\delta$ where δ is a positive rational. Then:*
(i) $\mathrm{cdim}(\alpha)\leq\mathcal{H}\left(\frac{1}{2}+\delta\right)$
(ii) Every subsequence β selected from α by a computable non-monotonic selection rule satisfies $\mathrm{Bias}(\beta)\leq\delta$.

Proof. Let α be a μ-Martin-Löf random sequence. To prove (i) we argue that the number of zeros in $\alpha\restriction_n$ is roughly $(1/2+\delta)n$ for all n, and that a string of length n with $(1/2+\delta)n$ zeros has a Kolmogorov complexity of at most $\mathcal{H}\left(\frac{1}{2}+\delta\right)n$. We start with the first assertion:

Lemma 3.2.14. $\#0(\alpha\restriction_n) = \left(\frac{1}{2}+\delta\right)n + o(n)$

Subproof. Let $(\mathbf{Y}_n)_{n\in\mathbb{N}}$ be the random variables defined, for a β chosen at random according to μ, by $\mathbf{Y}_n = \#0(\beta\restriction_n) - (1/2+\delta)n$. The $\mathbf{Y}_n$ satisfy the conditions of Azuma's inequality (with measure μ understood, and $c=1$). Hence for all n, the open set

$$\mathcal{U}_n = \left\{\beta \in 2^\omega : \ |\mathbf{Y}_n(\beta)| \geq n^{2/3}\right\}$$

(which is c.e. uniformly in n) has μ-measure at most $2\exp\left(\frac{-n^{1/3}}{2}\right)$. Thus, $\sum_n \mu(\mathcal{U}_n) < +\infty$. Hence by Theorem 1.3.4, α belongs to only finitely many $\mathcal{U}_n$, which means:

$$\#0(\alpha\restriction_n) = \left(\frac{1}{2}+\delta\right)n + O(n^{2/3})$$

□

Lemma 3.2.15. *Let w be a string of length n, and $s = \frac{\#0(w)}{n}$. The following inequality holds: $C(w) \leq \mathcal{H}(s)\,n + O(\log n)$.*

Subproof. w can be described by its position in the set

$$A = \{u \in 2^{<\omega} : \ |u| = n \wedge \#0(u) = sn\}$$

i.e. we have by Lemma 2.1.15:

$$\begin{aligned} C(w) &\leq \log(\#A) + \log(C(A)) \\ &\leq \log(\#A) + O(log(n)) \\ &\leq \log\binom{n}{sn} + O(\log n) \\ &\leq \mathcal{H}(s)\,n + O(\log n) \end{aligned}$$

(the last inequality follows from Stirling's formula). □

By the two above lemmas, we have $C(\alpha\restriction_n) \leq \mathcal{H}\left(\frac{1}{2}+\delta\right)n + o(n)$ for all n. By Theorem 2.2.27 (and Remark 2.2.28), it follows that $\mathrm{cdim}(\alpha) \leq \mathcal{H}\left(\frac{1}{2}+\delta\right)$.

The part (ii) of the proof is very similar to the proof of Theorem 3.2.9. Fix a computable non-monotonic selection rule σ. We pick a sequence β at random with respect to the probability measure μ and we run σ on β. Let $i_0, i_1, \ldots$ be the positions of the bits selected from β by σ. The i_k are random variables of parameter β which take their values in $\mathbb{N}\cup\{\perp\}$. We define a sequence of random variables $(\mathbf{X}^\sigma_n)_{n\in\mathbb{N}}$ by setting $\mathbf{X}^\sigma_0 = 0$ and for all n:

$$\mathbf{X}^\sigma_{n+1} = \begin{cases} \#0(\beta_{(i_0)}\ldots\beta_{(i_n)}) - \left(\frac{1}{2}+\delta\right)n & \text{if } i_n \neq \perp \\ \mathbf{X}^\sigma_n \text{ otherwise} \end{cases}$$

By definition of μ, the sequence $(\mathbf{X}^\sigma_n)_{n\in\mathbb{N}}$ satisfies the hypotheses of Azuma's inequality. Hence, the open set

$$\mathcal{V}_n = \{\beta \in 2^\omega : \ |\mathbf{X}^\sigma_n(\beta)| \geq n^{2/3}\}$$

(which is c.e. uniformly in n) has μ-measure at most $2\exp\left(\frac{-n^{1/3}}{2}\right)$. Thus, $\sum_n \mu(\mathcal{V}_n) < +\infty$. Hence by Theorem 1.3.4, our μ-Martin-Löf random sequence α belongs to only finitely many $\mathcal{U}_n$, which means

$$\left|\#0(\alpha_{(i_0)}\dots\alpha_{(i_n)}) - \left(\frac{1}{2}+\delta\right)n\right| = O\left(n^{2/3}\right)$$

In other words, if the selected subsequence $\gamma = \alpha_{(i_0)}\alpha_{(i_1)}\dots$ is infinite, it satisfies $\mathrm{Bias}(\gamma) = \delta$. This being true for every computable non-monotonic selection rule σ, we are done. ■

3.3 Equivalence and consistency for arbitrary measures

We now turn to the general case of arbitrary computable measures. Given a notion of randomness, two measures are "effectively consistent" if their classes of random sequences are not disjoint, and "effectively equivalent" if they have the same random sequences. We first classify the different effective consistency relations, for which the picture is quite simple. As we will see, things get much more complicated for effective equivalence relations.

3.3.1 Consistency

Effective consistency relations induced by Martin-Löf randomness, Schnorr randomness and computable randomness are equivalent to classical consistency:

Theorem 3.3.1. *Let μ and ν be computable measures. The following are equivalent:*
(a) μ and ν are inconsistent
(b) $\mu\mathbf{MLR} \cap \nu\mathbf{MLR} = \emptyset$
(c) $\mu\mathbf{CR} \cap \nu\mathbf{CR} = \emptyset$
(d) $\mu\mathbf{SR} \cap \nu\mathbf{SR} = \emptyset$
(e) $\mu\mathbf{WR} \cap \nu\mathbf{SR} = \emptyset$

Proof. Since for any given measure μ one has

$$\mu\mathbf{MLR} \subseteq \mu\mathbf{CR} \subseteq \mu\mathbf{SR} \subseteq \mu\mathbf{WR}$$

the implications $(e) \Rightarrow (d) \Rightarrow (c) \Rightarrow (b)$ are immediate. Moreover, the class of μ-Martin-Löf sequence having measure 1 for every μ, the implication $(b) \Rightarrow (a)$ follows directly. It remains to prove the implication $(a) \Rightarrow (e)$. Let μ and ν be two inconsistent measures i.e. there exists a set $\mathcal{X}$ such that $\mu(\mathcal{X}) = 0$ and $\nu(\mathcal{X}) = 1$. Fix $\varepsilon > 0$. By regularity of μ, for all n, there exists an open set $\mathcal{U}$ that contains $\mathcal{X}$ and such that $\mu(\mathcal{U}) < \varepsilon$ (and of course $\nu(\mathcal{U}) = 1$ since $\mathcal{U}$ contains $\mathcal{X}$). But since the cylinders form a basis for the topology, $\mathcal{U}$ is a countable union of cylinders: $\mathcal{U} = \bigcup_{i=0}^{+\infty}[w_i]$. We have

$$1 = \nu\left(\mathcal{U}_n\right) = \nu\left(\bigcup_{i=0}^{+\infty}[w_i]\right) = \lim_{N\to+\infty}\nu\left(\bigcup_{i=0}^{N}[w_i]\right)$$

Hence, taking N large enough, the set $\mathcal{V} = \bigcup_{i=0}^{N} [w_i]$ satisfies $\mu(\mathcal{V}) < \varepsilon$ (since $\mathcal{V} \subseteq \mathcal{U}$) and $\nu(\mathcal{V}) > 1 - \varepsilon$. What we have proven is that for all $\varepsilon > 0$ there exists a finite union $\mathcal{V}$ of cylinders that has μ-measure smaller than ε and ν-measure greater than $1 - \varepsilon$.

Now, consider the procedure which given n enumerates all the finite subsets A of $2^{<\omega}$ and returns the first one such that $\mu([A]) < 2^{-n}$ and $\nu([A]) > 1 - 2^{-n}$. This procedure will eventually find such an A by the above discussion. Moreover, this procedure is computable as the finite subsets of $2^{<\omega}$ can be computably enumerated and μ and ν are computable measures. Hence, there exists a uniformly computable sequence $(A_n)_{n\in\mathbb{N}}$ of finite subsets of $2^{<\omega}$ such that $\mu([A_n]) < 2^{-n}$ and $\nu([A_n]) > 1 - 2^{-n}$ for all n. Now, consider the set

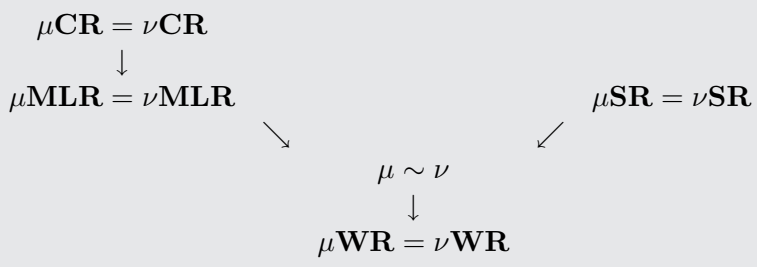
$$\mathcal{Y} = \limsup_n([A_n]) = \bigcap_{n\in\mathbb{N}} \bigcup_{k>n} [A_k]$$

For any given n, the set $\mathcal{W}_n = \bigcup_{k>n}[A_k]$ is a c.e. open set, uniformly in n. Moreover, one has $\mu(\mathcal{W}_n) \leq \sum_{k>n} 2^{-k} \leq 2^{-n}$. Moreover, $\mu(\mathcal{W}_n)$ is computable (uniformly in n) as the measure of $[A_k]$ decreases exponentially in k. Hence, the sequence $(\mathcal{W}_n)_{n\in\mathbb{N}}$ is a μ-Schnorr test, which implies $\mu\mathbf{SR} \cap \mathcal{Y} = \emptyset$. On the other hand, for all n, $\nu(\mathcal{W}_n) = 1$ as $\nu(\mathcal{W}_n) \geq \nu([A_{n'}]) \geq 1 - 2^{-n'}$ for all $n' > n$. Being a c.e. open set of ν-measure 1, $\mathcal{W}_n$ must contain all ν-weakly random sequences. This being true for all n, we have $\nu\mathbf{WR} \subseteq \mathcal{Y}$. Together with $\mu\mathbf{SR} \cap \mathcal{Y} = \emptyset$, this implies $\mu\mathbf{SR} \cap \nu\mathbf{WR} = \emptyset$. ■

3.3.2 A classification of equivalence relations

The effective equivalence relations induced by Martin-Löf randomness, Schnorr randomness, computable randomness and weak randomness can be classified as follows.

Theorem 3.3.2. *For all computable probability measures μ and ν, the following implications hold. Except for the transitive closure of the implications shown, no other implication is true in general.*



Remark 3.3.3. *The implication structure between the different equivalence relations stated in this theorem is surprising in so far as it does not reflect the implications that hold between the underlying notions of randomness (see page 40).*

The rest of this chapter will be devoted to the proof of this theorem. In this section, we prove that all the above implications hold, while the next section will

deal with the construction of counter-examples for all the other possible implications.

One of the implications of Theorem 3.3.2 was already present in the work of Muchnik et al.:

Proposition 3.3.4 (Muchnik et al. [50]). *Let μ and ν be measures. If $\mu\mathbf{CR} = \nu\mathbf{CR}$, then $\mu\mathbf{MLR} = \nu\mathbf{MLR}$.*

Proof. Let μ and ν be two measures that have the same computably random sequences. Let α be a sequence that is not μ-Martin-Löf random. We shall prove that α is not ν-Martin-Löf random either. We can assume that α is μ-computably random, for if it was not, then by the hypothesis it would not be ν-computably random, hence not ν-Martin-Löf random and we would be done. Since α is not μ-Martin-Löf random, by Theorem 1.5.8 (which can be adapted in a straightforward way to any computable measure), there exists a normed left-c.e. μ-martingale d such that $\limsup_n d(\alpha\restriction_n) = +\infty$. By Proposition 3.1.9, there exists a measure ξ such that $d = \xi/\mu$. Notice that this makes ξ left-c.e. For all n, we have:

$$d(\alpha\restriction_n) = \frac{\xi(\alpha\restriction_n)}{\mu(\alpha\restriction_n)} = \frac{\xi(\alpha\restriction_n)}{\nu(\alpha\restriction_n)}\frac{\nu(\alpha\restriction_n)}{\mu(\alpha\restriction_n)}$$

The term $\nu(\alpha\restriction_n)/\mu(\alpha\restriction_n)$ is upper-bounded by a constant. This is because ν/μ is a computable μ-martingale (Proposition 3.1.9), and we have assumed that α is μ-computably random. Hence the divergence of $\limsup_n d(\alpha\restriction_n)$ comes from the term $\xi(\alpha\restriction_n)/\nu(\alpha\restriction_n)$ in the above equation i.e. one must have

$$\limsup_n \frac{\xi(\alpha\restriction_n)}{\nu(\alpha\restriction_n)} = +\infty$$

But ξ/ν is a ν-martingale (Proposition 3.1.9), and it is left-c.e. as ξ is left-c.e. and ν is computable. This means that a left-c.e. ν-martingale succeeds on α, hence α is not ν-Martin-Löf random (Theorem 1.5.8). ■

Proposition 3.3.5. *Let μ and ν be two computable measures.*
(i) If $\mu\mathbf{MLR} = \nu\mathbf{MLR}$, then $\mu \sim \nu$.
(ii) If $\mu\mathbf{SR} = \nu\mathbf{SR}$, then $\mu \sim \nu$.

Proof. The proof is very similar to that of Theorem 3.3.1, therefore we only sketch this one. We prove (i) and (ii) at the same time. Suppose that μ and ν are not equivalent. By symmetry we may assume that there exists a set $\mathcal{X}$ such that $\mu(\mathcal{X}) = 0$ and $\nu(\mathcal{X}) > 0$. Let q be a rational number such that $\nu(\mathcal{X}) > q > 0$. Given n, it is possible to find effectively (and uniformly in n) a finite subset A_n of $2^{<\omega}$ such that $\mu([A_n]) < 2^{-n}$ and $\nu([A_n]) > q - 2^{-n}$. Then, the set

$$\mathcal{Y} = \limsup_n([A_n]) = \bigcap_{n\in\mathbb{N}} \bigcup_{k>n} [A_k]$$

satisfies $\mu(\mathcal{Y}) = 0$ and $\nu(\mathcal{Y}) \geq q$. The set $\bigcup_{k>n}[A_k]$ is a c.e. open set (uniformly in n) whose μ-measure is computable (uniformly in n) and smaller than 2^{-n}. Hence, the family $\left(\bigcup_{k>n}[A_k]\right)_{n\in\mathbb{N}}$ is a μ-Schnorr test whose intersection is $\mathcal{Y}$, hence $\mu\mathbf{SR} \cap \mathcal{Y} = \emptyset$ (and a fortiori $\mu\mathbf{MLR} \cap \mathcal{Y} = \emptyset$). But since $\nu(\mathcal{Y}) > 0$ its intersection with $\nu\mathbf{MLR}$ (which has ν-measure 1) is non-empty. Any sequence $\alpha \in 2^\omega$ that belong to this intersection witnesses the fact that $\mu\mathbf{MLR} \neq \nu\mathbf{MLR}$ and $\mu\mathbf{SR} \neq \nu\mathbf{SR}$. ■

Time is now ripe to give the full proof of Theorem 3.2.5:

Proof of Theorem 3.2.5. (i). By Corollary 3.2.8, we have $(a) \Rightarrow (e)$ and $(a) \Rightarrow (f)$. By Theorem 3.3.2: $(e) \Rightarrow (d)$, $(d) \Rightarrow (c)$, $(f) \Rightarrow (c)$, and $(c) \Rightarrow (g)$. The implication $(g) \Rightarrow (b)$ comes from Theorem 3.3.1 (part $(a) \Rightarrow (e)$). Finally, $(b) \Rightarrow (a)$ is exactly the part of Kakutani's theorem that we have proven.

(ii) The equivalence of (b), (d), (e), (f) and (g) holds by Theorem 3.3.1. $(a) \Rightarrow (b)$ is the part of Kakutani's theorem that we have proven, $(b) \Rightarrow (c)$ is trivial, and $(c) \Rightarrow (a)$ is exactly the implication $(a) \Rightarrow (c)$ of part (i). ■

It turns out that Theorem 3.2.5 (which we have now fully proven) implies Kakutani's theorem:

Proof of Theorem 3.2.3. Given two strongly positive generalized Bernoulli measures μ and ν, it suffices to take an oracle $\beta \in 2^\omega$ that computes their respective parameters and relativize Theorem 3.2.5 to β. ■

3.3.3 Counter-examples

We now turn to the more delicate task of showing that all other implications between the equivalence relations we study do not hold. When constructing corresponding counter-examples, the sets introduced in the following definition will play a crucial role.

Definition 3.3.6. *For computable measures μ and ν and for $k \in \mathbb{R}_+$, let*

$$\mathcal{L}^k_{\mu/\nu} = \left\{\alpha \in 2^\omega : \sup_n \frac{\mu(\alpha\restriction_n)}{\nu(\alpha\restriction_n)} \geq k\right\} \qquad \text{and} \qquad \mathcal{L}^\infty_{\mu/\nu} = \bigcap_{k\in\mathbb{N}} \mathcal{L}^k_{\mu/\nu}$$

Since for all μ, ν the ratio μ/ν is a ν- martingale, by Ville's inequality, one has:

$$\nu(\mathcal{L}^k_{\mu/\nu}) \leq 1/k \qquad \text{and} \qquad \nu(\mathcal{L}^\infty_{\mu/\nu}) = 0$$

If μ and ν are computable, the second part of the above statement can be made even more precise:

Lemma 3.3.7. *Let μ and ν be computable measures. We have $\mathcal{L}^\infty_{\mu/\nu} \cap \nu\mathbf{CR} = \emptyset$ (a fortiori, $\mathcal{L}^\infty_{\mu/\nu} \cap \nu\mathbf{MLR} = \emptyset$ and $\nu(\mathcal{L}^\infty_{\mu/\nu}) = 0$).*

Proof. This is because μ/ν is a computable ν-martingale. □

The equivalence relations between μ and ν that we study in this chapter are closely related to the set $\mathcal{L}^{\infty}_{\mu/\nu}$.

Proposition 3.3.8. *For every pair μ and ν of computable measures the following equivalences hold.*

(i) $\mu \sim \nu$ *if and only if* $\mu(\mathcal{L}^{\infty}_{\mu/\nu}) = \nu(\mathcal{L}^{\infty}_{\nu/\mu}) = 0$

(ii) $\mu\mathbf{MLR} = \nu\mathbf{MLR}$ *if and only if* $\mathcal{L}^{\infty}_{\mu/\nu} \cap \mu\mathbf{MLR} = \mathcal{L}^{\infty}_{\nu/\mu} \cap \nu\mathbf{MLR} = \emptyset$

(iii) $\mu\mathbf{CR} = \nu\mathbf{CR}$ *if and only if* $\mathcal{L}^{\infty}_{\mu/\nu} \cap \mu\mathbf{CR} = \mathcal{L}^{\infty}_{\nu/\mu} \cap \nu\mathbf{CR} = \emptyset$

Proof. For all three equivalences the "only if" direction is immediate from Lemma 3.3.7. Let us now prove the "if" directions. By symmetry for assertion (i) it suffices to demonstrate that every set that is ν-null is also μ-null, and similarly for the two other assertions it suffices to demonstrate $\mu\mathbf{MLR} \subseteq \nu\mathbf{MLR}$ and $\mu\mathbf{CR} \subseteq \nu\mathbf{CR}$, respectively.

Lemma 3.3.9. *For every open set $\mathcal{U}$ and all measures μ and ν it holds that*

$$\mu\left(\mathcal{U} \cap \overline{\mathcal{L}^{k}_{\mu/\nu}}\right) \leq k\, \nu(\mathcal{U})\ .$$

Subproof. By additivity of measures, it suffices to prove this for cylinders. Let $w \in 2^{<\omega}$, and let us prove that $\mu\left([w] \cap \overline{\mathcal{L}^{k}_{\mu/\nu}}\right) \leq k\, \nu([w])$. If $[w] \cap \overline{\mathcal{L}^{k}_{\mu/\nu}} = \emptyset$, we are done. Otherwise, let $\alpha \in [w] \cap \overline{\mathcal{L}^{k}_{\mu/\nu}}$. By definition, this means that w is a prefix of α and for all n, $\mu(\alpha\restriction_n)/\nu(\alpha\restriction_n) \leq k$ which in particular implies $\mu(w) \leq k\, \nu(w)$. A fortiori, $\mu\left([w] \cap \overline{\mathcal{L}^{k}_{\mu/\nu}}\right) \leq k\, \nu([w])$. □

We return to the proof of Proposition 3.3.8:

(i) Suppose $\mu(\mathcal{L}^{\infty}_{\mu/\nu}) = \nu(\mathcal{L}^{\infty}_{\nu/\mu}) = 0$, and let $\mathcal{X}$ be a set such that $\nu(\mathcal{X}) = 0$. For given $k \in \mathbb{N}$, by regularity, let $\mathcal{U}$ be an open set that contains $\mathcal{X}$ and such that $\nu(\mathcal{U}) \leq 1/k^2$. Then we have

$$\begin{aligned}
\mu(\mathcal{X}) &\leq \mu(\mathcal{U}) \\
&\leq \mu\left(\mathcal{U} \cap \mathcal{L}^{k}_{\mu/\nu}\right) + \mu(\mathcal{U} \cap \overline{\mathcal{L}^{k}_{\mu/\nu}}) \\
&\leq \mu(\mathcal{L}^{k}_{\mu/\nu}) + k\, \nu(\mathcal{U}) \quad \text{(by Lemma 3.3.9)} \\
&\leq \mu(\mathcal{L}^{k}_{\mu/\nu}) + 1/k
\end{aligned}$$

This is true for all k, we can let k tend to $+\infty$, and we get $\mu(\mathcal{X}) \leq \mu(\mathcal{L}^{\infty}_{\mu/\nu})$. And by assumption, $\mu(\mathcal{L}^{\infty}_{\mu/\nu}) = 0$.

(ii) Suppose $\mathcal{L}^{\infty}_{\mu/\nu} \cap \mu\mathbf{MLR} = \mathcal{L}^{\infty}_{\nu/\mu} \cap \nu\mathbf{MLR} = \emptyset$. Let $\alpha \notin \nu\mathbf{MLR}$. In case α is a member of $\mathcal{L}^{\infty}_{\mu/\nu}$, by assumption $\alpha \notin \mu\mathbf{MLR}$ holds and we are done. Otherwise,

there is a nonzero natural number k such that $\alpha \notin \mathcal{L}^k_{\mu/\nu}$. Fix a ν-Martin-Löf test $\{\mathcal{U}_n\}_{n\in\mathbb{N}}$ where $\alpha \in \bigcap_n \mathcal{U}_n$ and a computable function $(n,i) \mapsto u_{n,i}$ such that for all n, the set $\mathcal{U}_n$ is the disjoint union of the basic open sets $[u_{n,1}], [u_{n,2}], \ldots$. For all n, let

$$\mathcal{V}_n = \bigcup \{[u_{n,i}] \colon i \in \mathbb{N} \text{ and } \mu(u_{n,i}) \leq k\ \nu(u_{n,i})\}\ .$$

Then $\{\mathcal{V}_n\}_{n\in\mathbb{N}}$ is a uniformly effectively sequence of open sets which, by definition, satisfies $\mu(\mathcal{V}_n) \leq k\ \nu(\mathcal{U}_n)$ for all n, hence $\{\mathcal{V}_n\}_{n\in\mathbb{N}}$ is a μ-Martin-Löf test by Lemma 1.3.2. But $\alpha \notin \mathcal{L}^k_{\mu/\nu}$, hence for all n, $\mu(\alpha{\restriction_n}) \leq k\ \nu(\alpha{\restriction_n})$ hence $\alpha \in \mathcal{V}_n$ for all n and thus $\alpha \in \bigcap_n \mathcal{V}_n$, and consequently $\alpha \notin \mu\mathbf{MLR}$.

(iii) Suppose $\mathcal{L}^\infty_{\mu/\nu} \cap \mu\mathbf{CR} = \mathcal{L}^\infty_{\nu/\mu} \cap \nu\mathbf{CR} = \emptyset$. Let $\alpha \notin \nu\mathbf{CR}$, i.e., there is a μ-martingale that succeeds on A and thus, by Proposition 3.1.9, there exists a computable measure ξ such that

$$\limsup_{n\to\infty} \frac{\xi(\alpha{\restriction_n})}{\nu(\alpha{\restriction_n})} = +\infty\ . \tag{3.5}$$

In case A is a member of $\mathcal{L}^\infty_{\mu/\nu}$, by assumption $\alpha \notin \mu\mathbf{CR}$ holds and we are done. Otherwise, the quotients $\mu(\alpha{\restriction_n})/\nu(\alpha{\restriction_n})$ are bounded from above and (3.5) remains valid with ν replaced by μ, which by Proposition 3.1.9 implies $\alpha \notin \mu\mathbf{CR}$.

■

Proposition 3.3.8 will be extremely useful in the construction of counter-examples. In fact we will use a slight variation of it. Indeed, we will give several constructions of measures μ where we only define the values of $\mu(u)$ for words u whose length is a power of 3. First notice that if a function $u \mapsto m(u)$ is computable when restricted to the words whose length is a power of 3, and if the condition

$$m(u) = \sum_{\{w\colon |w|=3|u|\}} m(w)$$

is satisfied for all such words, then m canonically extends to a computable measure μ (for the words u such that $3^s < |u| < 3^{s+1}$, set inductively (in decreasing order of length) $\mu(u) = m(u0) + m(u1)$). Similarly, if a function $u \mapsto d(u)$ is computable when restricted to the words whose length is a power of 3, and if the condition

$$(3^{|u|+1} - 3^{|u|})d(u) = \sum_{\{w\colon |w|=3|u|\}} d(w)$$

is satisfied for all such words, then d canonically extends to a computable λ-martingale.

That said, we need to make sure that things still work if we restrict our attention to words whose length is a power of 3. But this is quite naturally the case, as the cylinders $[u]$ generated by such words form a basis for the topology of 2^ω. For example, if $\mathcal{U}$ is an effectively open set, one can give an enumeration of $\mathcal{U}$ with such cylinders: instead of enumerating a cylinder $[w]$ where $3^s < |w| < 3^{s+1}$, just enumerate $\{[u] \colon w \sqsubset u \text{ and } |u| = 3^{s+1}\}$. Based on this observation, we introduce the following definition.

Definition 3.3.10. *Let μ and ν be computable measures. Then for every $k \in \mathbb{R}^+$ we put*

$$\widehat{\mathcal{L}}^k_{\mu/\nu} = \left\{\alpha \in 2^\omega : \sup_n \frac{\mu(\alpha\restriction_{3^n})}{\nu(\alpha\restriction_{3^n})} \geq k\right\} \qquad \text{and} \qquad \widehat{\mathcal{L}}^\infty_{\mu/\nu} = \bigcap_{k\in\mathbb{N}} \widehat{\mathcal{L}}^k_{\mu/\nu} \,.$$

Then we get the desired variant of Proposition 3.3.8.

Proposition 3.3.11. *For every pair μ and ν of computable measures the following equivalences hold.*

(i) $\mu \sim \nu$ *if and only if* $\mu(\widehat{\mathcal{L}}^\infty_{\mu/\nu}) = \nu(\widehat{\mathcal{L}}^\infty_{\nu/\mu}) = 0$,

(ii) $\mu\mathbf{MLR} = \nu\mathbf{MLR}$ *if and only if* $\widehat{\mathcal{L}}^\infty_{\mu/\nu} \cap \mu\mathbf{MLR} = \widehat{\mathcal{L}}^\infty_{\nu/\mu} \cap \nu\mathbf{MLR} = \emptyset$,

(iii) $\mu\mathbf{CR} = \nu\mathbf{CR}$ *if and only if* $\widehat{\mathcal{L}}^\infty_{\mu/\nu} \cap \mu\mathbf{CR} = \widehat{\mathcal{L}}^\infty_{\nu/\mu} \cap \nu\mathbf{CR} = \emptyset$.

Similarly, one obtains the following characterizations of Schnorr randomness and weak randomness, which can be verified by using savings martingales as discussed in Remark 1.4.8.

Proposition 3.3.12. *Let μ be a computable measure. A sequence α is μ-Schnorr random if and only if there exists no computable μ-martingale d and computable order g such that $d(\alpha\restriction_{3^n}) \geq g(n)$ for infinitely many n.*
A sequence α is μ-weakly random if and only if there exists no computable μ-martingale d and computable order g such that $d(\alpha\restriction_{3^n}) \geq g(n)$ for all n.

As a further step towards the construction of counter-examples, we show the following proposition.

Proposition 3.3.13. *Let $\alpha \in \lambda\mathbf{SR}$, and suppose that α is $\mathbf{0}'$-computable. There exists a computable measure μ such that $\alpha \notin \mu\mathbf{SR}$ and*

$$\widehat{\mathcal{L}}^\infty_{\mu/\lambda} = \emptyset \qquad \text{and} \qquad \widehat{\mathcal{L}}^\infty_{\lambda/\mu} = \{\alpha\} \,.$$

Proof. We will in fact construct a computable λ-martingale d such that: $d(\alpha\restriction_{3^n})$ tends to 0 as n tends to infinity, in such a way that $d(\alpha\restriction_{3^n}) \leq 1/n$ for infinitely many n and if $\beta \neq \alpha$, $d(\beta\restriction_{3^n})$ will be eventually constant. Then, setting $\mu(u) = \lambda(u)d(u)$ for all words u, μ will be as desired. By the above discussion, we will only define $d(u)$ for those words u whose length is a power of 3, which we do by stages: at stage s, $d(u)$ will be defined for all words u of length 3^s.

Since α is $\mathbf{0}'$ computable, it is the pointwise limit of a sequence of words $\{w_s\}_{s\in\mathbb{N}}$. We can moreover assume that $\lim_{s\to+\infty} |w_s| = +\infty$, that $|w_s| \leq 3^s$ for all s, and that w_s is a prefix of α for infinitely many s.

Let $E = \{u1^{2|u|} : u \in 2^{<\omega}\}$. Notice that every λ-Schnorr random sequence has only finitely many prefixes in E. Hence, up to changing a finite number of its bits,

we can assume that α has no prefix in E. Let us now proceed to the construction of d.

Stage $s = 0$. Set $d(\epsilon) = d(0) = d(1) = 1$.

Stage $s+1$. Suppose $d(u)$ is defined for a word u of length 3^s. We define $d(u')$ for every extension u' of u of length 3^{s+1} as follows:

- if u is not an extension of w_s, set $d(u') = d(u)$
- if u is an extension of w_s, and u' is not in E (i.e. $u' \neq u1^{2|u|}$) set $d(u') = \frac{d(u)}{s+1}$ and then set $d(u1^{2|u|})$ in such a way that the average value of

$$\{d(u') : u \sqsubset u' \text{ and } |u'| = 3^{s+1}\}$$

is equal to $d(u\restriction_{3^s})$

We turn to the verification.

Claim 1: The function $n \mapsto d(\beta\restriction_{3^n})$ is eventually constant for all $\beta \neq \alpha$. Proof: if $\beta \neq \alpha$, there exists s_0 such that if $s > s_0$, w_s is not a prefix of β (because the sequence w_s pointwise converges to $\alpha \neq \beta$), and thus, by construction of d, for all $s > s_0$, $d(\beta\restriction_{3^s}) = d(\beta\restriction_{3^{s_0}})$.

Claim 2: The function $n \mapsto d(\alpha\restriction_{3^n})$ tends to 0 and $d(\alpha\restriction_{3^n}) \leq 1/n$ for infinitely many n. Proof: this is a direct consequence of the definition of d. Since α has no prefix in E, by construction of d, one has for all s either $d(\alpha\restriction_{3^{s+1}}) = d(\alpha\restriction_{3^s})$ or $d(\alpha\restriction_{3^{s+1}}) = d(\alpha\restriction_{3^s})/(s+1)$. Hence; $s \mapsto d(\alpha\restriction_{3^s})$ is non-increasing and is smaller than $1/s$ for all s such that w_s is a prefix of α, which happens infinitely often.

Let us now consider $\mu = \lambda\, d$. By the above discussion, $\widehat{\mathcal{L}}^{\infty}_{\mu/\lambda} = \emptyset$, $\widehat{\mathcal{L}}^{\infty}_{\lambda/\mu} = \{\alpha\}$. Moreover, if we consider the μ-martingale $d' = \frac{\lambda}{\mu}$ we see that for infinitely many s, $d'(\alpha\restriction_{3^s}) \geq s$. Hence, by Proposition 3.3.12, $\alpha \notin \mu\mathbf{SR}$.

■

Proposition 3.3.14. *(i) There exists a computable measure μ such that $\lambda \sim \mu$ and nonetheless $\lambda\mathbf{MLR} \neq \mu\mathbf{MLR}$, $\lambda\mathbf{CR} \neq \mu\mathbf{CR}$, $\lambda\mathbf{SR} \neq \mu\mathbf{SR}$.*
(ii) There exists a computable measure μ such that $\lambda\mathbf{MLR} = \mu\mathbf{MLR}$ and $\lambda\mathbf{CR} \neq \mu\mathbf{CR}$.
(iii) There exists a computable measure μ such that $\lambda\mathbf{CR} = \mu\mathbf{CR}$ and $\lambda\mathbf{SR} \neq \mu\mathbf{SR}$.

Proof. (i) Let α be an $\mathbf{0}'$-computable member of $\lambda\mathbf{MLR}$ (such as Chaitin's constant Ω). Let μ be, by Proposition 3.3.13, a computable measure such that $\widehat{\mathcal{L}}^{\infty}_{\mu/\lambda} = \emptyset$, $\widehat{\mathcal{L}}^{\infty}_{\lambda/\mu} = \{\alpha\}$ and $\alpha \notin \mu\mathbf{SR}$. Since $\lambda(\{\alpha\}) = 0$, by Proposition 3.3.11, we have $\lambda \sim \mu$. Moreover, since $\alpha \in \lambda\mathbf{MLR} \subset \lambda\mathbf{CR} \subset \lambda\mathbf{SR}$, and $\alpha \notin \mu\mathbf{SR}$, it follows that

$\lambda\mathbf{MLR} \neq \mu\mathbf{MLR}$, $\lambda\mathbf{CR} \neq \mu\mathbf{CR}$, $\lambda\mathbf{SR} \neq \mu\mathbf{SR}$.

(ii) Let β be an $\mathbf{0}'$-computable sequence such that $\beta \in \lambda\mathbf{CR} \setminus \lambda\mathbf{MLR}$. Such a sequence exists by Theorem 2.2.18, where we constructed a $\mathbf{0}'$-computable sequence that was λ-computably random, but nonetheless had prefixes of very low Kolmogorov complexity (hence by Schnorr theorem was not λ-Martin-Löf random). By Proposition 3.3.13, there exists a computable measure μ such that $\widehat{\mathcal{L}}^{\infty}_{\mu/\lambda} = \emptyset$, $\widehat{\mathcal{L}}^{\infty}_{\lambda/\mu} = \{\beta\}$ and $\beta \notin \mu\mathbf{SR}$ (a fortiori $\beta \notin \mu\mathbf{CR}$). By Proposition 3.3.11, we have $\lambda\mathbf{MLR} = \mu\mathbf{MLR}$ (since $\beta \notin \lambda\mathbf{MLR}$) and $\lambda\mathbf{CR} \neq \mu\mathbf{CR}$ (since $\beta \in \lambda\mathbf{CR} \setminus \mu\mathbf{CR}$).

(iii) By Corollary 2.2.26, there exists a sequence γ which is left-c.e. (hence $\mathbf{0}'$-computable) and not Church stochastic (hence non-computably random) and Schnorr random. Let γ be such a sequence. By Proposition 3.3.13, there exists a computable measure μ such that $\widehat{\mathcal{L}}^{\infty}_{\mu/\lambda} = \emptyset$, $\widehat{\mathcal{L}}^{\infty}_{\lambda/\mu} = \{\gamma\}$ and $\gamma \notin \mu\mathbf{SR}$. By Proposition 3.3.11, we have $\lambda\mathbf{CR} = \mu\mathbf{CR}$ (since $\gamma \notin \lambda\mathbf{CR}$) and $\lambda\mathbf{SR} \neq \mu\mathbf{SR}$ (since $\gamma \in \lambda\mathbf{SR} \setminus \mu\mathbf{SR}$).

■

Proposition 3.3.15. *There exists a computable measure μ such that $\lambda\mathbf{SR} = \mu\mathbf{SR}$, $\lambda\mathbf{CR} \neq \mu\mathbf{CR}$ and $\lambda\mathbf{MLR} \neq \mu\mathbf{MLR}$*

Proof. First, we prove the following lemma.

Lemma 3.3.16. *Let μ and ν be two computable measures and $\alpha \in 2^{\omega}$. If $\alpha \in \nu\mathbf{SR} \setminus \mu\mathbf{SR}$, then there exists a computable order g such that $\frac{\nu(\alpha\restriction_{3^n})}{\mu(\alpha\restriction_{3^n})} \geq g(n)$ infinitely often.*

Subproof. Let $\alpha \in \nu\mathbf{SR} \setminus \mu\mathbf{SR}$. By Proposition 3.1.9, and Proposition 3.3.12, there exists a computable measure ξ and a computable order g such that $\frac{\xi(\alpha\restriction_{3^n})}{\mu(\alpha\restriction_{3^n})} \geq g(n)$ for infinitely many n. Since $\alpha \in \nu\mathbf{SR}$ and since $\sqrt{g}$ is a computable order, for almost all n, $\frac{\xi(\alpha\restriction_{3^n})}{\nu(\alpha\restriction_{3^n})} \leq \sqrt{g(n)}$. Hence, for infinitely many n: $\frac{\nu(\alpha\restriction_{3^n})}{\mu(\alpha\restriction_{3^n})} \geq \frac{\xi(\alpha\restriction_{3^n})}{\mu(\alpha\restriction_{3^n})} \frac{\nu(\alpha\restriction_n)}{\xi(\alpha\restriction_{3^n})} \geq \frac{g(n)}{\sqrt{g(n)}} = \sqrt{g(n)}$. □

Let Ω be Chaitin's constant, which is in $\lambda\mathbf{MLR}$. We will construct, in a very similar way as for Proposition 3.3.13 a computable measure μ such that $\widehat{\mathcal{L}}^{\infty}_{\mu/\lambda} = \emptyset$, $\widehat{\mathcal{L}}^{\infty}_{\lambda/\mu} = \{\Omega\}$ but this time, we want Ω to be μ-Schnorr random. By the above lemma, it will be sufficient to ensure that $\frac{\lambda(\Omega\restriction_{3^n})}{\mu(\Omega\restriction_{3^n})}$ tends to infinity more slowly than any computable order. Hence, we will again construct a λ-martingale d such that $\lim_n d(\Omega\restriction_{3^n}) = 0$ and if $\beta \neq \Omega$, $d(\beta\restriction_{3^n})$ will be eventually constant, ensuring that $d(\Omega\restriction_{3^n})$ decreases very slowly.

Since Ω is a left-c.e. sequence, let $(w_s)_{s\in\mathbb{N}}$ be a sequence of words, increasing for the lexicographic order and such that Ω is the pointwise limit of this sequence. We transform $(w_s)_{n\in\mathbb{N}}$ into a reduced form $(w'_s)_{n\in\mathbb{N}}$ as follows. Set $w'_0 = \epsilon$. Then, by induction, if $w'_0, \ldots, w'_s$ are already defined, w'_{s+1} is defined to be the shortest prefix of w_{s+1} that does not belong to $w'_0, \ldots, w'_s$. This way, the sequence $(w'_s)_{n\in\mathbb{N}}$ is computable (since $(w'_s)_{n\in\mathbb{N}}$ is) and has the following properties:

Lemma 3.3.17.
(a) For all s, all the strict prefixes of w'_s belong to $w'_0, \dots, w'_{s-1}$.
(b) $(w'_s)_{n\in\mathbb{N}}$ is increasing for the lexicographic order.
(c) All prefixes of Ω appear in the sequence $(w'_s)_{n\in\mathbb{N}}$.
(d) $(w'_s)_{n\in\mathbb{N}}$ pointwise converges to Ω.
(e) $|w_s| \leq s$ for all s.

Subproof. (a) For all s, by definition of w'_s, the prefix u of w'_s of length $|w'_s| - 1$ belongs to $w'_0, \dots, w'_{s-1}$. We get the desired result by induction.

(b) We prove by induction that $w'_0 <_{lex} \dots <_{lex} w'_t$. Suppose this is true at stage t. For sake of contradiction, suppose this is not true at stage $t+1$ i.e. $w'_{t+1} \leq_{lex} w'_s$ for some $s \leq t$. Hence

$$w'_{t+1} \leq_{lex} w'_s \leq_{lex} w_s \leq_{lex} w_{t+1}$$

(the second inequality comes from the definition of w'_s, the third one from the fact that the w are increasing). Since w'_{t+1} is a prefix of w_{t+1}, the above inequalities tell us that w'_{t+1} is also a prefix of w'_s. But by (a), this means that $w'_{t+1} = w'_r$ for some $r < s$. This contradicts the definition of w'_{t+1}.

(c) Let u be a prefix of Ω. Since the sequence $(w_s)_{s\in\mathbb{N}}$ pointwise converges to Ω, there exists some t such that for all $s \geq t$, u is a prefix of w_s. On the other hand, by (b) the sequence $(w'_s)_{n\in\mathbb{N}}$ is increasing for $\leq_{lex}$, it has no repetition, hence $\lim_t |w_t| = +\infty$. Let $r \geq t$ such that $|w_r| \geq |u|$. Then, w'_r is a prefix of w_r (by definition), and is longer than u, which is also a prefix of w_r (since $r \geq t$). Hence, u is a prefix of w'_r. Applying (a), this tells us that u appears in the sequence $(w'_s)_{n\in\mathbb{N}}$.

(d) This is a direct consequence of (b) and (c).

(e) This follows from (a) and a straightforward induction. □

We now return to the proof of Proposition 3.3.15. Like in the proof of Proposition 3.3.13, we set $E = \{u1^{2|u|} : u \in 2^{<\omega}\}$. Clearly, Ω has only finitely many prefixes in E, so up to modifying it on a finite numer of bits, we can assume that it has no prefix in E.

We construct a computable λ-martingale d such that $d(\Omega \upharpoonright_{3^s})$ tends to 0 slowly (this will be made precise below), and if $\beta \neq \Omega$, $d(\beta\upharpoonright_n)$ is eventually constant:

Stage $s = 0$. Set $d(\epsilon) = d(0) = d(1) = 1$.

Stage $s+1$. Suppose $d(u)$ is defined for a word u of length 3^s. We define $d(u')$ for every extension u' of u of length 3^{s+1} as follows:

- if u is not an extension of w'_s, set $d(u') = d(u)$
- if u is an extension of w'_s, and u' is not in E (i.e. $u' \neq u1^{2|u|}$) set $d(u') = \frac{1}{|w'_s|+1}$ and then set $d(u1^{2|u|})$ in such a way that the average value of

$$\{d(u') : u \sqsubseteq u' \text{ and } |u'| = 3^{s+1}\}$$

is equal to $d(u)$

It is not obvious that the second case is always well defined: we need to make sure that $d(u) \geq \frac{1}{|w'_s|+1}$ when u is an extension of w'_s of length 3^s (i.e. d needs to have enough capital at u to distribute among the extensions of u). We prove this using Lemma 3.3.17, by induction. This is clearly true for $s = 0$. For the induction step, suppose this is true for some all $t < s$. Let u be an extension of w'_s of length 3^s. By Lemma 3.3.17, there exists $t < s$ such that w'_t is a the prefix of w'_s of length $|w'_s| - 1$ (since the sequence w' is increasing for the lexicographic order, this also implies that none of the w'_r for $t < r < s$ are prefixes of u). By the induction hypothesis, we have $d(u\restriction_{3^t}) \geq \frac{1}{|w'_t|+1}$. By construction of d, we then have $d(u\restriction_{3^{t+1}}) \geq \frac{1}{|w'_t|+1}$. For all $t < r < s$, since w'_r is not a prefix of u, still by construction of d, $d(u\restriction_{3^{r+1}}) = d(u\restriction_{3^r})$. This means that $d(u\restriction_{3^s}) = d(u\restriction_{3^t})$, hence $d(u\restriction_{3^s}) \geq \frac{1}{|w'_t|+1} \geq \frac{1}{|w'_s|+1}$, which completes the induction.

Let us now check that d is as desired. Let $\alpha \in 2^\omega$. For all s, we have $d(\alpha\restriction_{3^{s+1}}) = d(\alpha\restriction_{3^s})$, unless w'_s is a prefix of α. Since the w'_s pointwise converge to Ω, if $\alpha \neq \Omega$, α has only finitely many prefixes in $(w'_s)_{n\in\mathbb{N}}$, hence $d(\alpha\restriction_{3^s})$ is eventually constant (and positive by construction). Let us now study the behaviour of $d(\Omega\restriction_{3^s})$. By Lemma 3.3.17, there exist $t_0 < t_1 < t_2 < \dots$ such that $\Omega\restriction_i = w'_{t_i}$ for all i (and if s is not a t_i, w'_s is not a prefix of Ω. By construction of d, it is easy to check that for all i: $d(\Omega\restriction_{3^{t_i+1}}) = \frac{1}{|w_{t_i}+1|} = \frac{1}{i+1}$ and for all $s \in (t_i, t_{i+1}]$, $d(\Omega\restriction_{3^s}) = d(\Omega\restriction_{3^{t_i+1}}) = \frac{1}{i+1}$. Hence, $d(\Omega\restriction_{3^s})$ is nonincreasing and tends to 0 as s tends to $+\infty$. However, it tends to 0 very slowly:

Lemma 3.3.18. *For every computable order g, $d(\Omega\restriction_{3^s}) \geq 1/g(s)$ for almost all s.*

Subproof. Suppose that there exists a computable order g such that $d(\Omega\restriction_{3^s}) \leq 1/g(s)$ for infinitely many s. Pick such an s. Let i be such that $s \in (t_i, t_{i+1}]$. We have by the above discussion:

$$\frac{1}{i+1} = d(\Omega\restriction_{3^{t_i+1}}) = d(\Omega\restriction_{3^s}) \leq \frac{1}{g(s)} \leq \frac{1}{g(t_i+1)}$$

Hence for infinitely many i, $g(t_i + 1) \leq i + 1$, which implies $t_i < g^{-1}(i + 1)$. Set $r_i = g^{-1}(i + 1)$ for all i. The sequence of r_i is computable. Moreover, for all i such that $r_i > t_i$ (which happen infinitely often), $\Omega\restriction_i$ is a prefix of w'_{r_i} because

$$\Omega\restriction_i = w'_{t_i} \leq_{lex} w'_{r_i} \leq_{lex} \Omega$$

Therefore, for all such i, $\Omega\restriction_i$ can be retrieved from r_i, which implies $K(\Omega\restriction_i) \leq K(r_i) + O(1) \leq K(i) + O(1) \leq O(\log i)$. This can only happen for finitely many i, a contradiction. □

We finally put everything together. We define the measure μ by $\mu(w) = d(w)\lambda(w)$ for all $w \in 2^{<\omega}$. It is computable since d is (and it is a measure by Proposition 3.1.9). It remains to show that μ satisifes the conclusion of the Theorem. First, Second, we have seen that $d(\Omega\restriction_{3^s})$ (which is equal to $\frac{\mu(\Omega\restriction_{3^s})}{\lambda(\Omega\restriction_{3^s})}$) tends to 0, hence $\frac{\lambda(\Omega\restriction_{3^s})}{\mu(\Omega\restriction_{3^s})}$ tends to $+\infty$, hence $\Omega \notin \mu\mathbf{CR}$. Since $\Omega \in \lambda\mathbf{MLR}$ (and a fortiori

$\Omega \in \lambda\mathbf{CR}$), this proves that $\lambda\mathbf{MLR} \neq \mu\mathbf{MLR}$ and $\lambda\mathbf{CR} \neq \mu\mathbf{CR}$. To see that $\lambda\mathbf{SR} = \mu\mathbf{SR}$, we use Lemma 3.3.16. We have seen that $s \mapsto d(\beta\restriction_{3^s})$ is eventually constant for $\beta \neq \Omega$. For all $\beta \neq \Omega$, by Lemma 3.3.16: $\beta \in \lambda\mathbf{SR} \Leftrightarrow \mu\mathbf{SR}$. For Ω, the ratio $\frac{\lambda(\Omega\restriction_{3^s})}{\mu(\Omega\restriction_{3^s})}$ (equal to $\frac{1}{d(\Omega\restriction_{3^s})}$) tends to $+\infty$ but, by Lemma 3.3.18, slower than any computable order. Hence, $\Omega \in \mu\mathbf{SR}$. This proves that $\lambda\mathbf{SR} = \mu\mathbf{SR}$ and completes the proof. ■

Proposition 3.3.19. *There exist a computable measure μ such that μ and λ are consistent and $\lambda\mathbf{WR} \neq \mu\mathbf{WR}$*

Proof. Let δ be the measure such that $\delta(\{0^\omega\}) = 1$ (which is clearly computable). Set $\mu = \delta/2 + \lambda/2$. λ and μ are consistent: let $\mathcal{X} \subseteq 2^\omega$. If $0^\omega \in \mathcal{X}$, then $\mu(\mathcal{X}) = 1/2 + \lambda(\mathcal{X})/2$ and if $0^\omega \notin \mathcal{X}$, $\mu(\mathcal{X}) = 1/2 + \lambda(\mathcal{X})/2$. In both cases, it is impossible to have $\mu(\mathcal{X}) = 0$ and $\lambda(\mathcal{X}) = 1$. On the other hand, $0^\omega \in \mu\mathbf{WR}$ and $0^\omega \notin \lambda\mathbf{WR}$. ■

We now come to our last counter-example.

Proposition 3.3.20. *There exists a computable probability measure μ such that $\mu\mathbf{WR} = \nu\mathbf{WR}$ and $\mu \not\sim \nu$.*

Despite the fact that weak randomness is not a very good notion of randomness, this result is particularly interesting. Indeed, by regularity, two Borel measures on 2^ω are equivalent if and only if they have the same closed nullsets. The above proposition shows that this cannot be effectivized: two computable Borel measures on 2^ω can have the same effectively closed nullsets and yet not be equivalent (indeed, having the same weakly random sequence exactly means having the same effectively closed sets).

Proof. We will construct a computable measure μ such that λ and μ have the same weakly random sequences and yet are not equivalent. As in the proof of Proposition 3.3.15, the construction will be done by constructing a λ-martingale d and setting $\mu = d\,\lambda$. And here again, we will only define d on words the length of which is a power of 3, the values on the other words being implicitely defined.

Our proof involves $\mathbf{0}'$-Martin-Löf randomness, and its characterization by computable upper bounds for C proven in Theorem 2.3.22: there exists a c.u.b $C^\star$ of C such that every α is $\mathbf{0}'$-Martin-Löf randomness if and only if $C^\star(\alpha\restriction_n) \geq n - O(1)$ for infinitely many n.

In fact, in the rest of the proof, we will require $C^\star$ to have another property: we would like to have $C^\star(u1^{2|u|}v) \leq 2|u| + |v| + O(1)$ for all words u, v. We can assume that it is the case. Indeed, for all $u, v \in 2^{<\omega}$, we have $C(u1^{2|u|}v) \leq |u| + |v| + O(\log|u|)$. Hence, if we define a function $\widehat{C}$ on $2^{<\omega}$ by $\widehat{C}(u1^{2|u|}v) = 2|u| + |v|$, and $\widehat{C}(w) = +\infty$ if w is not of this type, we get a c.u.b. for C. Taking $C^{\star\star} = \min(\widehat{C}, C^\star)$, we get a computable upper bound for C which satisfies the above property and can replace $C^\star$ in Theorem 2.3.22 (according to Remark 2.3.23)

The set of $\mathbf{0}'$-Martin-Löf random sequences has measure 1, and by definition of $C^\star$, is equal to the nested countable union

$$\bigcup_{c\in\mathbb{N}}\{\alpha:\ \exists^\infty n\ C^\star(\alpha\restriction_n)\geq n-c\}$$

Thus, there exists some $c_0\in\mathbb{N}$ such that

$$\mathcal{R}=\{\alpha:\ \exists^\infty n\ C^\star(\alpha\restriction_n)\geq n-c_0\}$$

has positive λ-measure.

For all $\alpha\in 2^\omega$, define the function h_α by

$$h_\alpha(s)=\#\big\{0\leq t<s:\exists n\in(3^t,3^{t+1}]\ C^\star(\alpha\restriction_n)\geq n-c_0\big\}$$

If $\alpha\in\mathcal{R}$, h_α is an order. If $\alpha\notin\mathcal{R}$, h_α is eventually constant. As proven by Nies, Stephan and Terwijn [51], if $\alpha\in\mathcal{R}$, then there is no computable order g such that $g\leq h_\alpha$. Suppose otherwise. Then

$$\alpha\in\Big\{\beta\in 2^\omega:\ \forall s\ \#\{k\leq 3^s:\ C^\star(\beta\restriction_k)\geq k-c_0\}\geq g(s)\Big\}$$

Notice that the right-hand-side of the above relation is an effectively closed set, which by definition of $C^\star$ and Theorem 2.3.22 contains only $\mathbf{0}'$-λ-Martin-Löf random sequences. This is a contradiction since by the Low Basis theorem, every non-empty effectively closed class contains a $\mathbf{0}'$-computable sequence.

We now construct the martingale d. Set $d(0)=d(1)=1$. Suppose $d(u)$ is defined for all u of length 3^s. Define inductively d on words of length 3^{s+1} as follows. Let u be a word of length 3^s. For all extension w of u such that $|w|=3^{s+1}$ and which is different from $u1^{(3^{s+1}-3^s)}$:

- if there exists some $n\in(3^s,3^{s+1}]$ such that $C^\star(w\restriction_n)\geq n-c_0$, set $d(w)=d(u)/2$
- otherwise, set $d(w)=d(u)$

Clearly, d is computable. There are three cases for the behavior of d. If $\alpha\notin\mathcal{R}$ and α has finitely many prefixes in E, there are finitely many n such that $C^\star(\alpha\restriction_n)\geq n-c_0$, hence $d(\alpha\restriction_{3^{s+1}})=d(\alpha\restriction_{3^s})$ for almost all s, by construction of d. If $\alpha\notin\mathcal{R}$ and α has infinitely many prefixes in E, then α has prefixes of type $u1^{2|u|}$ for arbitrarily long u. But for all such u, the complexity of an extension uv satisfies $C^\star(u1^{2|u|}v)\leq 2|u|+|v|+O(1)$ the right-hand side being smaller than $|u1^{2|u|}v|-c_0$ for u long enough. And thus, for such a long u, by construction of d one will have $d(u1^{2|u|}v)=d(u1^{2|u|})$ for all v. Finally, if $\alpha\in\mathcal{R}$, it is in particular $\mathbf{0}'$-λ-Martin-Löf random hence there are only finitely many prefixes of α of type $u1^{2|u|}$. Thus, up to a fixed positive multiplicative constant: $d(\alpha\restriction_{3^s})=2^{-h_\alpha(s)}$ for all s (by construction of d). Notice that in all the above cases $n\mapsto d(\alpha\restriction_{3^n})$ is bounded from above.

Set $\mu = d\,\lambda$. Let us prove that μ is as desired.

Claim 1: μ is computable. Proof: this is obvious since d is.

Claim 2: μ and λ are not equivalent. Proof: by definition, $\lambda(\mathcal{R}) > 0$. On the other hand, $d' = 1/d = \lambda/\mu$ is a μ-martingale (Proposition 3.1.9) which on every $\alpha \in \mathcal{R}$ satisfies $d'(\alpha\restriction_{3^s}) = 2^{h_\alpha(s)}$ for all s up to a positive multiplicative constant. Since h_α is an order when $\alpha \in \mathcal{R}$, this proves that d' succeeds on all $\alpha \in \mathcal{R}$. Hence, $\mathcal{R} \cap \mu\mathbf{CR} = \emptyset$ and thus $\mu(\mathcal{R}) = 0$.

Claim 3: $\mu\mathbf{WR} \subseteq \lambda\mathbf{WR}$. Proof: let $\alpha \notin \lambda\mathbf{WR}$. There exists a computable λ-martingale d_0 and a computable order g such that $d_0(\alpha\restriction_{3^n}) \geq g(n)$ for all n. Since on every α, $d(\alpha\restriction_{3^n}) = \frac{\mu(\alpha\restriction_{3^n})}{\lambda(\alpha\restriction_{3^n})}$ is bounded from above, say by a constant $r > 0$, the μ-martingale $d_1 = d_0\,\frac{\lambda}{\mu}$ satisfies $d_1(\alpha\restriction_n) \geq \frac{1}{r} d_0(\alpha\restriction_n) \geq \frac{g(n)}{r}$. This proves that $\alpha \notin \mu\mathbf{WR}$.

Claim 4: $\lambda\mathbf{WR} \subseteq \mu\mathbf{WR}$. Proof: suppose $\alpha \notin \mu\mathbf{WR}$, and let us show that $\alpha \notin \lambda\mathbf{WR}$. Since $\alpha \notin \mu\mathbf{WR}$, by there exists a computable μ-martingale d_2 and a computable order f such that $d_2(\alpha\restriction_n) \geq f(n)$ for all n.

We distinguish two cases. If $\alpha \notin \mathcal{R}$, we have seen that $d(\alpha\restriction_{3^n}) = \frac{\mu(\alpha\restriction_{3^n})}{\lambda(\alpha\restriction_{3^n})}$ is eventually constant, hence bounded from above, say by a constant $r' > 0$ and thus, with the same argument as above, the λ-martingale $d_3 = d_2\,\frac{\mu}{\lambda}$ satisfies $d_3(\alpha\restriction_{3^n}) \geq \frac{g'(n)}{r}$ for all n. Since $\frac{g'(n)}{r}$ is an order, this implies $\alpha \notin \lambda\mathbf{WR}$.

In the second case, i.e. $\alpha \in \mathcal{R}$, observe that $\frac{d_2}{d'}$ is a (computable) λ-martingale. Since $\alpha \in \mathcal{R}$, α is in particular $\mathbf{0}'$-Martin-Löf random, hence $\frac{d_2(\alpha\restriction_{3^n})}{d'(\alpha\restriction_{3^n})}$ is bounded from above, say by a constant $r'' > 0$. Hence:

$$d'(\alpha\restriction_{3^n}) = d_2(\alpha\restriction_{3^n})\,\frac{d'(\alpha\restriction_{3^n})}{d_2(\alpha\restriction_{3^n})} \geq \frac{f(n)}{r''}$$

Recall that $d'(\alpha\restriction_{3^s}) = 2^{h_\alpha(s)}$. It follows that

$$2^{h_\alpha(s)} \geq \frac{f(3^s)}{r''}$$

and hence

$$h_\alpha(s) \geq \log\left(\frac{f(3^s)}{r''}\right)$$

But $\log\left(\frac{f(3^s)}{r''}\right)$ is an order, which contradicts the fact that h_α majorizes no computable order. Hence, the second case cannot happen (and the first does yield $\alpha \notin \lambda\mathbf{WR}$).

Putting Claim 3 and Claim 4 together, this finishes the proof. ■

Bibliography

[1] N. Alon and J. Spencer. *The probabilistic method*. Wiley-Interscience, second edition, 2000.

[2] K. Ambos-Spies, E. Mayordomo, Y. Wang, and X. Zheng. Resource-bounded balanced genericity, stochasticity and weak randomness. In *Symposium on Theoretical Aspects of Computer Science (STACS 1996)*, volume 1046 of *Lecture Notes in Computer Science*, pages 63–74. Springer, 1996.

[3] E. Asarin. Some properties of Kolmogorov Δ-random sequences. *Theory of Probability and its Applications*, 32:507–508, 1987.

[4] R. Ash. *Probability and measure theory*. Academic Press, second edition, 1999.

[5] V. Becher and S. Figueira. An example of a computable absolutely normal number. *Theoretical Computer Science*, 270(1-2):947–958, 2002.

[6] V. Becher, S. Figueira, and R. Picchi. Turing's unpublished algorithm for normal numbers. *Theoretical Computer Science*, 377(1-3):126–138, 2007.

[7] L. Bienvenu. Constructive equivalence relations for computable probability measures. In *Computer Science - Theory and Applications, First International Computer Science Symposium in Russia (CSR 2006)*, volume 3967 of *Lecture Notes in Computer Science*, pages 92–103. Springer, 2006.

[8] L. Bienvenu. Kolmogorov-Loveland stochasticity and Kolmogorov complexity. In *Symposium on Theoretical Aspects of Computer Science (STACS 2007)*, volume 4393 of *Lecture Notes in Computer Science*, pages 260–271. Springer, 2007.

[9] L. Bienvenu and W. Merkle. Effective randomness for computable probability measures. In *International Conference on Computability and Complexity in Analysis (CCA 2006)*, volume 167 of *Electronic Notes in Computer Science*, pages 117–130, 2007.

[10] L. Bienvenu and W. Merkle. Reconciling data compression and Kolmogorov complexity. In *International Colloquium on Automata, Languages and Pro-*

gramming (ICALP 2007), volume 4596 of *Lecture Notes in Computer Science*, pages 643–654. Springer, 2007.

[11] L. Bienvenu, W. Merkle, and A. Shen. A simple proof of Miller-Yu theorem. Accepted for publication in Fundamenta Informaticae.

[12] G. Chaitin. Information-theoretical characterizations of recursive infinte strings. *Theoretical Computer Science*, 2:45–48, 1976.

[13] G. Chaitin. Incompleteness theorems for random reals. *Advances in Applied Mathematics*, 8:119–146, 1987.

[14] R. Cilibrasi and P. Vitanyi. Clustering by compression. *IEEE Transactions on Information Theory*, 51(4), 2005.

[15] D. Cohn. *Measure theory*. Birkhäuser Boston, 1994.

[16] B. Durand and N. Vereshchagin. Kolmogorov-Loveland stochasticity for finite strings. *Information Processing Letters*, 91(6):263–269, 2004.

[17] K. Falconer. *The geometry of fractal sets*. Cambridge Unversity Press, 1985.

[18] P. Gács. On the symmetry of algorithmic information. *Soviet Mathematics Doklady*, 15:1477–1480, 1974.

[19] P. Gács. Exact expressions for some randomness tests. *Z. Math. Log. Grdl. M.*, 26:385–394, 1980.

[20] P. Gács. Every set is reducible to a random one. *Information and Control*, 70:186–192, 1986.

[21] P. Gács. Uniform test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341(1-3):91–137, 2005.

[22] F. Hausdorff. Dimension und äusseres Mass. *Mathematishce Annalen*, 79:157–179, 1919.

[23] J. Hitchcock. *Effective fractal dimension: foundations and applications*. PhD dissertation, Iowa State University, Ames, 2003.

[24] M. Hoyrup and C. Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. http://arxiv.org/abs/0709.0907.

[25] J. Jacod and P. Protter. *Probability Essentials*. Springer, 2003.

[26] C. Jockusch. Degrees of generic sets. In F. Drake and S. S. Wainer, editors, *Recursion theory: its generalizations and applications*. Cambridge Unversity Press, 1980.

[27] C. Jockusch and R. Soare. Π^0_1 classes and degrees of theories. *Transaction of the American Mathematical Society*, 173:33–56, 1972.

[28] S. Kakutani. On equivalence of infinite product measures. *Annals of Mathematics*, 49(214-224), 1948.

[29] A. Kolmogorov. On tables of random numbers. *Sankhya Series A*, 25:369–376, 1963.

[30] L. Kraft. A device for quantizing, grouping, and coding amplitude modulated pulses. Master's thesis, Massachusetts Institute of Technology, 1949.

[31] A. Kučera. Measure, Π^0_1 classes, and complete extensions of PA. *Lecture Notes in Mathematics*, 1141:245–259, 1985.

[32] S. Kurtz. *Randomness and genericity in the degrees of unsolvability.* PhD dissertation, University of Illinois at Urbana, 1981.

[33] J. Lathrop and J. Lutz. Recursive computational depth. *Information and Computation*, 153(1):139–172, 1999.

[34] L. Levin. *Some theorems on the algorithmic approach to probability theory and information theory.* Dissertation in mathematics, Moscow, 1971.

[35] L. Levin. The concept of random sequence. *Doklady Akademii Nauk SSSR*, 212:548–550, 1973.

[36] L. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61:15–37, 1984.

[37] M. Li and P. Vitanyi. *An introduction to Kolmogorov complexity and its applications.* Graduate Texts in Computer Science. Springer, second edition, 1997.

[38] D. Loveland. A new interpretation of the von mises concept of random sequence. *Z. Math. Log. Grdl. M.*, 12:279–294, 1966.

[39] J. Lutz. Almost everywhere high nonuniform complexity. *Journal of Computer and System Sciences*, 44:220–258, 1992.

[40] J. Lutz. Dimension in complexity classes. *SIAM Journal on Computing*, 32:1236–1259, 2003.

[41] J. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187:49–79, 2003.

[42] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[43] E. Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Information Processing Letters*, 84:1–3, 2002.

[44] W. Merkle. The complexity of stochastic sequences. In *IEEE Conference on Computational Complexity (Complexity 2003)*, pages 230–235. IEEE Computer Society, 2003.

[45] W. Merkle. The Kolmogorov-Loveland stochastic sequences are not closed under selecting subsequences. *Journal of Symbolic Logic*, 68:1362–1376, 2003.

[46] W. Merkle and N. Mihailovic. On the construction of effective random sets. In *Mathematical Foundations of Computer Science (MFCS 2002)*, volume 2420, pages 568–580, 2002.

[47] W. Merkle, J. S. Miller, A. Nies, J. Reimann, and F. Stephan. Kolmogorov-Loveland randomness and stochasticity. *Annals of Pure and Applied Logic*, 138(1-3):183–210, 2006.

[48] J. S. Miller. Every 2-random real is Kolmogorov random. *Journal of Symbolic Logic*, 69(3):907–913, 2004.

[49] J. S. Miller and L. Yu. On initial segment complexity and degrees of randomness. *Transaction of the American Mathematical Society*, to appear.

[50] A. A. Muchnik, A. Semenov, and V. Uspensky. Mathematical metaphysics of randomness. *Theoretical Computer Science*, 207(2):263–317, 1998.

[51] A. Nies, F. Stephan, and S. Terwijn. Randomness, relativization and Turing degrees. *Journal of Symbolic Logic*, 70:515–535, 2005.

[52] C. Schnorr. *Zufälligkeit und Wahrscheinlichkeit*, volume 218 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin-Heidelberg-New York, 1971.

[53] C. Schnorr. Process complexity and effective random tests,. *Journal of Computer and System Sciences*, 7:376–388, 1973.

[54] A. Shen. On relations between different algorithmic definitions of randomness. *Soviet Mathematics Doklady*, 38:316–319, 1989.

[55] R. Solovay. Draft of a paper (or series of papers) on Chaitin's work. Unpublished notes, 215 pages, 1975.

[56] A. Turing. A note on normal numbers. In J. Britton, editor, *Collected works of Alan Turing*, pages 117–119. North Holland, Amsterdam, 1992.

[57] M. van Lambalgen. *Random sequences*. PhD dissertation, University of Amsterdam, Amsterdam, 1987.

[58] J. Ville. *Etude critique de la notion de collectif*. Gauthiers-Villars, Paris, 1939.

[59] R. von Mises. Grundlagen der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919.

[60] V. Vovk. On a criterion for randomness. *Soviet Mathematics Doklady*, 294(6):1298–1302, 1987.

[61] Y. Wang. *Randomness and Complexity*. PhD dissertation, University of Heidelberg, 1996.

[62] Y. Wang. A separation of two randomness concepts. *Information Processing Letters*, 69(3):115–118, 1999.

[63] A. Zvonkin and L. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83–124, 1970.

Index

Résumé

Cette thèse est une contribution à l'étude des différentes notions effectives d'aléatoire pour les objets individuels (essentiellement les suites binaires finies ou finies). Dans le premier chapitre nous considérons les approches de l'aléatoire par la théorie des jeux (martingales et stratégies) que nous comparons à l'approche historique par les fréquences qui remonte au début du 20ème siècle avec les travaux de von Mises. Le résultat principal de ce chapitre est une relation explicite entre la vitesse de gain d'une martingale (ou stratégie) sur une suite binaire et le biais des sous-suites extraites. Le second chapitre porte sur les liens existant entre les différentes définitions d'aléatoire pour les suites binaires infinies et la notion de complexité de Kolmogorov, définie comme étant la taille du plus court programme qui génère un objet donné. De nombreux résultats sont déjà connus dans ce domaine. Nous présentons une approche nouvelle, en utilisant non pas la complexité de Kolmogorov elle-même, mais ses bornes supérieures calculables. Cette approche est unificatrice, en ce sens qu'elle permet de caractériser précisément une grande variété de notions d'aléatoire, dont certaines pour qui la complexité de Kolmogorov échoue. Le troisième et dernier chapitre étudie l'extension des notions effectives d'aléatoire à des mesures de probabilité calculables quelconques, et plus particulièrement les relations d'équivalence qu'elles induisent sur ces mesures (où deux mesures sont équivalentes si elles ont les mêmes éléments aléatoires). Une preuve constructive (par les martingales) du théorème de Kakutani (qui caractérise l'équivalence entre les mesures de Bernoulli généralisées) y est notamment obtenue. Enfin, nous discutons en toute généralité (c'est-à-dire pour des mesures quelconques) les relations d'équivalence induites, dont nous donnons une classification complète.

Summary

This thesis is a contribution to the study of the different notions of effective randomness for individual objects (mainly binary sequences, finite or infinite). In the first chapter, we consider various game-theoretic approaches to randomness (via martingales and strategies), and we compare them to the historical approach by frequency stability, which goes back to the work of von Mises in the beginning of the 20th century. The principal result of the first chapter is an explicit relation between the "speed of success" of a martingale (or strategy) on a sequence and the bias of the selected subsequences. The second chapter focuses on the links between the various randomness notions for infinite sequences and the notion of Kolmogorov complexity (or program-size complexity), defined to be the size of the shortest program which outputs a given finite object. Many results are already known in this direction. We present a new approach, using computable upper bounds of Kolmogorov complexity instead of Kolmogorov complexity itself. This turns out to be a very unifying approach, in the sense that it allows us to characterize a wide variety of randomness notions, even some for which Kolmogorov complexity fails. The third and last chapter studies the extension of all randomness notions to wider classes of probability measures, and more specifically the equivalence relations induced by the randomness notions (where we say that two measures are equivalent if they have the same random sequences). A constructive proof of Kakutani's theorem (a criterion of equivalence for generalized Bernoulli measures) is presented. Finally, in great generality (i.e. for arbitrary computable measures), we give a complete hierarchical classification of the induced equivalence relations.