

DS de l'UE 4TIN404U.
Probabilités, Statistiques et Combinatoire.
27 mars 2017. Durée 2h.
Barème approximatif : 4 , 7, 4, 9

Documents interdits, à l'exception d'une feuille A4 avec notes personnelles; téléphones et calculatrices interdits.

Exercice 1. [Tirage avec remise, sans remise] On dispose d'une urne qui contient 60 boules noires, 40 boules rouges.

- (i) Bob y tire 1 boule uniformément au hasard, et la remet; il fait cela 4 fois, et on suppose que les tirages sont indépendants. Quelle est la probabilité qu'il ait obtenu 2 boules noires et 2 boules rouges ?
 - (ii) Alice choisit au hasard 4 boules d'un coup (uniformément parmi tous les choix de 4 boules parmi 100). Quelle est la probabilité qu'elle obtienne 2 boules noires et 2 boules rouges ?
- (i) En fait, Bob a $6/10$ à chaque tirage d'avoir une boule noire, et $4/10$ d'en avoir une rouge. Mais alors, puisque les tirages sont indépendants, on voit que le nb de boules noires obtenues est une binomiale $B(4, 6/10)$. La quantité demandée est donc $\binom{4}{2}(6/10)^2(4/10)^2$ que l'on peut simplifier si on veut.
- (ii) Pour ne pas se tromper, il faut voir que l'on est en train de tirer un sous ensemble de 4 éléments parmi 100. Le cas que l'on recherche consiste à tirer 2 boules parmi les 60 premières, et 2 boules parmi les 40 autres: ce sont 2 sous ensembles de taille 2, que l'on cherche. La proba recherchée est donc $\binom{60}{2}\binom{40}{2}/\binom{100}{4}$, puisque les tirages de 4 boules parmi 100 sont supposés équiprobables. De nouveau, on peut simplifier un peu si on veut.

Exercice 2. On dispose d'un dé à 6 faces bien équilibré: la probabilité qu'il tombe sur i est $1/6$ pour i allant de 1 à 6. On le lance 100 fois, et on suppose que les résultats obtenus sont indépendants. On note $N_1, N_2, \dots, N_6$ le nombre de 1, 2, , $\dots$, 6 obtenus, de sorte que $N_1 + \dots + N_6 = 100$.

- (i) Calculer $\mathbb{P}(N_1 = n_1)$ pour tout $n_1 \in \mathbb{N}$ (on fera attention que la formule peut changer si n_1 est trop grand). Comment s'appelle la loi de N_1 ?
- (ii) Identifier la loi de $N_1 + N_2$ (on peut répondre à la question sans faire de calcul).
- (iii) Quelle est la variance de N_1 ? de N_2 ? de $N_1 + N_2$?
- (iv) En déduire la covariance $\text{cov}(N_1, N_2)$. Est-ce que N_1 et N_2 sont indépendants ?

(v) Montrer que pour $n_1 \geq 0, n_2 \geq 0, n_1 + n_2 \leq 100$,

$$\mathbb{P}(N_1 = n_1, N_2 = n_2) = \frac{100!}{n_1!n_2!(100 - n_1 - n_2)!} \frac{4^{100-n_1-n_2}}{6^{100}}.$$

(On pourra remarquer, si nécessaire que $\binom{100}{n_1} \binom{100-n_1}{n_2} = \frac{100!}{n_1!n_2!(100-n_1-n_2)!}$).

- (i) Bon alors, si n_1 n'est pas dans l'ensemble $\{0, \dots, 100\}$, alors $\mathbb{P}(N_1 = n_1) = 0$. Sinon, comme chaque fois qu'on lance le dé, on a une chance sur 6 d'avoir un 1, le nb de 1 est une binomiale $B(100, 1/6)$, d'où on tire $\mathbb{P}(N_1 = n_1) = \binom{100}{n_1} (1/6)^{n_1} (5/6)^{100-n_1}$ pour $0 \leq n_1 \leq 100$.
- (ii) On pourrait faire un calcul un peu difficile, mais on peut remarquer qu'à chaque lancer on a un 1 ou un 2 avec proba (cumulée) $1/3$. Alors la loi de $N_1 + N_2$ est simplement une binomiale $(100, 1/3)$.
- (iii) Dans le cours, on a vu que la variance d'une binomiale de paramètre (n, p) est $np(1-p)$. Du coup, comme N_2 suit aussi une loi binomiale $(100, 1/6)$, on a $\text{var}(N_1) = \text{var}(N_2) = 100 \times (1/6) \times (5/6) = 500/36$ et $\text{var}(N_1 + N_2) = 100 \times (1/3) \times (2/3) = 200/9$.
- (iv) On a vu en cours que $\text{var}(N_1 + N_2) = \text{var}(N_1) + \text{var}(N_2) + 2\text{cov}(N_1, N_2)$ (et d'ailleurs, c'est valable en toute généralité, pour toute paire de variables qui possèdent une variance). On résout: $\text{cov}(N_1, N_2) = (\text{var}(N_1 + N_2) - \text{var}(N_1) - \text{var}(N_2))/2 = (200/9 - 1000/36)/2 = -200/72$. On voit alors que N_1 et N_2 ne sont pas indépendants, car s'ils l'étaient, leur covariance serait nulle (On peut aussi montrer que ce n'est pas indépendant, en disant que $\mathbb{P}(N_1 = 100, N_2 = 100) = 0$ alors que $\mathbb{P}(N_1 = 100)\mathbb{P}(N_2 = 100) = (1/6^{100})^2 \neq 0$).
- (v) On commence par prendre un tirage qui correspond à $N_1 = n_1, N_2 = n_2$, par exemple,

$$\underbrace{1, \dots, 1}_{n_1}, \underbrace{2, \dots, 2}_{n_2}, y_{n_1+n_2+1}, \dots, y_{100}$$

avec les y_i tout, sauf des 1 ou des 2. On voit qu'on doit avoir $N_3 + N_4 + N_5 + N_6 = 100 - n_1 - n_2$. Tout se passe comme si on avait un dé à 3 faces, une face "1", une face "2" et une face "3-4-5-6"; on veut compter quelle est la proba qu'il tombe sur la face 1, n_1 fois, n_2 fois sur la 2, $100 - n_1 - n_2$ sur la face "3-4-5-6", avec proba $1/6$ de tomber sur la 1ere, $1/6$ sur la 2eme $4/6$ sur la 3eme. Chaque tirage de ce type a une proba

$$\frac{1}{6^{n_1}} \frac{1}{6^{n_2}} \left(\frac{4}{6}\right)^{100-n_1-n_2} = \frac{4^{100-n_1-n_2}}{6^{100}}.$$

Reste à voir combien de tirages font l'affaire: On voit que les temps $t_1, \dots, t_{n_1}$ où apparaissent les 1 forment un sous ensemble de $\{1, \dots, 100\}$, les temps $t'_1, \dots, t'_{n_2}$ où apparaissent les 2, un sous ensemble de $\{1, \dots, 100\} \setminus \{t_1, \dots, t_{n_1}\}$, ceci fixant les temps où on a ni 1 ni 2. Ainsi, il y $\binom{n}{n_1}$ tels choix pour les t_i , puis $\binom{100-n_1}{n_2}$ pour les t'_i . Ca fait $\binom{n}{n_1} \binom{100-n_1}{n_2}$ possibilités.

Exercice 3. Soit X et Y deux variables de Bernoulli de paramètre $1/2$ indépendantes.

- (i) Calculer la loi de $X - Y$.
- (ii) Est-il possible de trouver la loi d'un couple de variables aléatoire (X', Y') telle que les lois de X' et de Y' sont aussi des lois de Bernoulli $1/2$ mais telle que la loi de $X' - Y'$ soit différente de celle de $X - Y$?

(i) Clairement $X - Y$ ne peut prendre que 3 valeurs: soit 0, soit 1 soit -1 puisque X et Y valent 0 ou 1. On a, en utilisant l'indépendance de X et Y ,

$$\mathbb{P}(X - Y = 0) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 0, Y = 0) = 1/4 + 1/4 = 1/2$$

et

$$\mathbb{P}(X - Y = 1) = \mathbb{P}(X = 1, Y = 0) = 1/4$$

et même chose pour $\mathbb{P}(X - Y = -1)$.

- (ii) Comme on souhaite que X' et Y' soient des lois de Bernoulli $1/2$, alors, le couple (X', Y') ne peut prendre que les valeurs $\{(0, 0), (1, 1), (0, 1), (1, 0)\}$. On représente ça par le tableau suivant:

$X' \setminus Y'$	0	1
0	a	b
1	c	d

avec dans la case, la proba d'avoir la valeur spécifiée. On veut que X' et Y' soient des Bernoulli $1/2$: il faut donc que $a + b + c + d = 1$, $a + b = 1/2$, $c + d = 1/2$, $a + c = 1/2$, $b + d = 1/2$. Et si on veut que $X' - Y'$ n'ait pas la même loi que précédemment, par exemple, il suffit que $\mathbb{P}(X' - Y' = 1) = \mathbb{P}(X' = 0, Y' = 1) = b$ ne valent pas $1/4$. Alors, on prend $b = 1/6$, et on déduit tout le reste: $a = 1/3$, $c = 1/6$ et $d = 1/3$. Et ça fait l'affaire, et la réponse est donc oui ! (mais il y avait aussi plein d'autres façons de faire).

Exercice 4. On lance un dé équilibré à 6 faces, et on note $D_1, D_2, \dots$ les résultats obtenus. On suppose que les D_i sont des variables indépendantes. On calcule successivement les sommes des premiers résultats obtenus, en posant $S_1 = D_1$, $S_2 = D_1 + D_2$, $S_3 = D_1 + D_2 + D_3$; autrement dit, pour tout n on pose $S_n = D_1 + \dots + D_n$.

- (i) Calculer la loi de S_2
- (ii) Expliquer pourquoi S_n a une probabilité positive de prendre n'importe quelle valeur entre n et $6n$, et une probabilité nulle pour les autres.

- (iii) Prenons un n fixé. En supposant que vous connaissez $\mathbb{P}(S_{n-1} = k)$ pour tout k , comment pouvez vous calculer $\mathbb{P}(S_n = m)$ pour tout m ?
- (iv) Donnez un algorithme A pour calculer $\mathbb{P}(S_n = m)$ pour tout n et pour tout m (bien sûr, on ne suppose pas ici, que l'on connaît $\mathbb{P}(S_{n-1} = k)$ pour tout k).
- (v) On note T le nombre de fois qu'il faut lancer le dé pour atteindre une somme de 100 au moins (pour la première fois): c'est-à-dire $T = \min\{t : S_t \geq 100\}$.

(a) Expliquez pourquoi

$$\mathbb{P}(T > t) = \mathbb{P}(S_t < 100).$$

(b) Fournir alors un algorithme B pour calculer $\mathbb{P}(T > t)$ (on pourra utiliser l'algorithme A au besoin).

(c) Comment peut-on déduire $\mathbb{P}(T = k)$ si on connaît $\mathbb{P}(T > t)$ pour tout t ?

(i) On utilise la formule des probabilités totales... On somme sur toutes les valeurs possibles pour le premier dé. On a

$$\begin{aligned}\mathbb{P}(S_2 = k) &= \sum_{j=1}^6 \mathbb{P}(D_1 = j, D_2 = k - j) \\ &= \sum_{j=1}^6 \mathbb{P}(D_1 = j) \mathbb{P}(D_2 = k - j) = 1/6 \sum_{j=1}^6 \mathbb{P}(D_2 = k - j)\end{aligned}$$

et bien sûr, $\mathbb{P}(D_2 = k - j)$ fait 0 si $k - j > 6$ ou $k - j < 1$, et $1/6$ sinon. On voit que $\mathbb{P}(S_2 = 2) = 1/36, \mathbb{P}(S_2 = 3) = 2/36, \dots, \mathbb{P}(S_2 = 7) = 6/36, \mathbb{P}(S_2 = 8) = 5/36, \dots, \mathbb{P}(S_2 = 12) = 1/6$.

(ii) Au minimum, tous les dés tombent sur 1, et au max, tous les dés tombent sur 6. Donc, $n \leq S_n \leq 6n$. Toutes les valeurs peuvent être atteintes entre n et $6n$, car partant d'une configuration non maximale de dés, on peut toujours ajouter un à un des dés (de valeur minimale), pour incrémenter de 1 le total.

(iii) On utilise simplement la formule des probabilités totales,

$$\mathbb{P}(S_n = m) = \sum_{j=1}^6 \mathbb{P}(S_{n-1} = m - j, D_n = j) = \sum_{j=1}^6 \mathbb{P}(S_{n-1} = m - j) \mathbb{P}(D_n = j)$$

car le résultat du dé numéro n ne dépend pas des précédents, il est indépendant de S_{n-1} . Le fait qu'on connaît $\mathbb{P}(S_{n-1} = m - j)$ suffit pour voir que la calcul peut être fait, vu que $\mathbb{P}(D_n = j) = 1/6$.

(iv) Il suffit de faire un algorithme récursif: on connaît effectivement $\mathbb{P}(S_1 = j) = \mathbb{P}(D_1 = j)$ pour tout j (c'est $1/6$ si $1 \leq j \leq 6$, 0 sinon). Donc un algo récursif consiste juste à dire :

Calcul de $\mathbb{P}(S_n = k)$:

- si $n = 1$, alors retourner $1/6$ si $1 \leq k \leq 6$, 0 sinon.

- si $n > 1$, alors retourner $\sum_{j=1}^6 \mathbb{P}(S_{n-1} = m - j) \mathbb{P}(D_n = j) = \sum_{j=1}^6 \mathbb{P}(S_{n-1} = m - j)/6$.

On peut aussi procéder à l'aide de boucles.

- (v) (a) L'identité $\mathbb{P}(T > t) = \mathbb{P}(S_t < 100)$ provient de l'identité des événements $\{T > t\}$ et $\{S_t < 100\}$: si $S_t < 100$, eh bien on n'a pas encore atteint 100 au temps t , donc il faut attendre davantage pour atteindre 100: donc $T > t$. Réciproquement si $T > t$, ça veut dire qu'au temps t on n'a pas atteint 100... donc la somme $S_t < 100$. Puisque les 2 événements $\{T > t\}$ et $\{S_t < 100\}$ correspondent exactement à la même situation (au temps t), ils ont même proba.
- (b) Pour calculer $\mathbb{P}(T > t)$ il suffit de calculer $\mathbb{P}(S_t < 100) = \sum_{k=1}^{99} \mathbb{P}(S_t = k)$. Or l'algo A , permet de calculer précisément $\mathbb{P}(S_t = k)$. L'algo B consiste donc juste à faire la somme des 99 termes que l'on calcule avec l'algo A .
- (c) Comme on connaît $\mathbb{P}(T > k)$ pour tout k , on voit que $\mathbb{P}(T = k) = \mathbb{P}(T > k-1) - \mathbb{P}(T > k)$. That's all !