

Probabilités, Statistiques, Combinatoire

Philippe Duchon

Université de Bordeaux – Licence Informatique

<http://www.labri.fr/perso/duchon/Enseignements/Probas/>

2019-2020

Previously...

- ▶ On a défini l'**espérance** (moyenne) d'une variable aléatoire
- ▶ On en a décrit certaines propriétés importantes (linéarité)
- ▶ On a également défini la **variance** (tendance à la dispersion)
- ▶ On n'a pas défini, mais on aurait dû : l'**écart-type** d'une variable aléatoire (la **racine carrée de la variance**) (notation fréquence : σ)
- ▶ **Note** : l'écart-type à la même "dimension" que les valeurs de la variable, et que l'espérance (si les valeurs sont exprimées en mètres, l'écart-type est exprimé également en mètres, alors que la variance est exprimée en mètres carrés)

Programme du jour

- ▶ **Inégalités sur des variables aléatoires** : inégalités de Markov et de Tchebycheff
- ▶ **Une application** : analyse d'un mécanisme de “rejet”
- ▶ **Une conséquence importante** : la “loi des grands nombres”

Inégalité de Markov

- ▶ **Inégalité de Markov** : si X est une variable aléatoire à **valeurs positives**, d'espérance m , alors pour tout $a > 0$,

$$\mathbb{P}(X \geq a.m) \leq \frac{1}{a}$$

- ▶ Autre formulation : pour tout $b > 0$, $\mathbb{P}(X \geq b) \leq \frac{m}{b}$
- ▶ **Attention** : il faut que la variable soit positive ; ne marche pas si on n'a pas de telle garantie !
- ▶ (une v.a. positive a probabilité au plus 1/10 d'être 10 fois plus grande que son espérance)
- ▶ **Remarque** : l'inégalité ne dit rien d'intéressant pour $a \leq 1$.

Preuve de l'inégalité de Markov

- ▶ X une v.a. à valeurs positives ou nulles, $x > 0$ fixé mais quelconque.
- ▶ On pose
 - ▶ $Y = x\mathbf{1}_{(X \geq x)} = \begin{cases} 0 & \text{si } X < x \\ x & \text{si } X \geq x \end{cases}$
 - ▶ $Z = X - Y = \begin{cases} X & \text{si } X < x \\ X - x & \text{si } X \geq x \end{cases}$
- ▶ On a $X = Y + Z$, donc $\mathbf{E}(X) = \mathbf{E}(Y) + \mathbf{E}(Z)$
- ▶ De plus $\mathbf{E}(Y) = x\mathbb{P}(X \geq x)$, et $Z \geq 0$ donc $\mathbf{E}(Z) \geq 0$
- ▶ **Donc** on a $\mathbf{E}(X) \geq x\mathbb{P}(X \geq x)$, soit

$$\mathbb{P}(X \geq x) \leq \frac{\mathbf{E}(X)}{x}.$$

Inégalité de Tchebycheff

- ▶ **Théorème** : si X est une variable aléatoire d'espérance m et de variance $v = \sigma^2$, alors on a, pour tout réel $x > 0$,

$$\mathbb{P}(|X - m| \geq x\sigma) \leq \frac{1}{x^2}.$$

- ▶ **Remarque** : la condition “ $|X - m| \geq x.\sigma$ ” correspond à “ $X \geq m + x.\sigma$ ou $X \leq m - x.\sigma$ ” : on regarde la probabilité d’être “loin de l’espérance”
- ▶ **Remarque** : pas de condition du type “ X positive”, mais en revanche il faut connaître la variance.
- ▶ Cette inégalité va nous permettre de montrer la “loi des grands nombres”

Preuve de Tchebycheff

- ▶ X une v.a., $m = \mathbf{E}(X)$, $\sigma^2 = \mathbf{Var}(X)$, $x > 0$ fixé mais quelconque.
- ▶ On pose $Z = (X - m)^2$, et on applique l'inégalité de Markov à Z :

$$\mathbb{P}(Z \geq x^2) \leq \frac{\mathbf{E}(Z)}{x^2}$$

- ▶ Or, $\mathbf{E}(Z) = \sigma^2$; et
 $(Z > x^2) = ((X - m)^2 \geq x^2) = (|X - m| \geq x)$.
- ▶ On a donc

$$\mathbb{P}(|X - m| \geq x) \leq \frac{\sigma^2}{x^2}$$

- ▶ Forme alternative : en posant $x' = x/\sigma$,

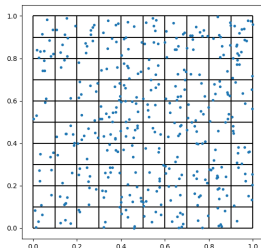
$$\mathbb{P}(|X - m| \geq x' \cdot \sigma) \leq \frac{1}{x'^2}$$

Conséquences de l'inégalité de Tchebycheff

- ▶ L'événement $\{|X - m| \geq a\}$, c'est $\{X \geq m + a\} \cup \{X \leq m - a\}$
- ▶ L'inégalité permet donc de majorer, au choix, $\mathbb{P}(X \geq m + a)$ (pour montrer qu'il est improbable que X soit "très grande"), ou $\mathbb{P}(X \leq m - a)$ (pour montrer qu'il est improbable que X soit "très petite").
- ▶ Particulièrement utile si on ne connaît pas la loi exacte de X , ou si les calculs sont compliqués (tout se ramène à identifier deux paramètres, espérance et variance)
- ▶ Il existe d'autres inégalités, "plus performantes", par exemple si on sait que X peut s'exprimer comme une somme de variables indépendantes ; mais on ne les abordera pas dans ce cours.

Une application : points aléatoires (cf TD)

On tire n points uniformément dans $[0, 1] \times [0, 1]$ (carré unité)



($n = 400$) En divisant le carré en 100 petits carrés d'aire $1/100$ chacun, on observe que 2 carrés ne contiennent aucun point, 5 carrés en contiennent 1, 15 carrés contiennent 2 points... et un carré en contient 9.

Est-ce “normal” ? “louche” ?

Points aléatoires (suite)

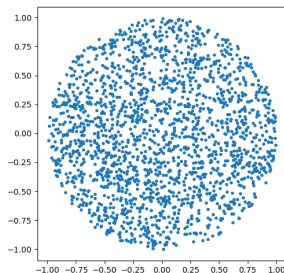
- ▶ Chaque point a probabilité $1/100$ de “tomber” dans un petit carré donné.
- ▶ **Donc**, pour un petit carré donné, le nombre de points qui tombent dedans suit la loi binomiale de paramètre $(n, 1/100)$.
- ▶ La probabilité qu’une telle binomiale prenne la valeur 0 est $0.99^{400} \simeq 0.01795$
- ▶ **Donc**, l’espérance du nombre de petits carrés sans points, est 100×0.01795 , soit environ 1.8
- ▶ Même calcul pour les carrés à 1 point : la proba de valoir 1 pour la binomiale est $400 \times (0.99)^{399} \times 0.01 \simeq 0.0725$.
- ▶ **Donc**, l’espérance du nombre de petits carrés contenant 1 point, est environ $100 \times 0.0725 \simeq 7.25$.
- ▶ Les valeurs observées (2 et 5) ne sont pas si éloignée que cela des espérances (1.8 et 7.25) ; pour en dire plus, il faudrait calculer des variances.

Points aléatoires : variances

- ▶ On a 100 variables aléatoires $I_1, I_2, \dots, I_{100}$, indicatrices de “le carré numéro k est vide” : toutes sont des Bernoulli de paramètre 0.01795.
- ▶ Donc leur somme N (nombre de carrés vides) est d'espérance 1.795.
- ▶ **Attention** : elles ne sont pas indépendantes (la probabilité qu'elles soient toutes égales à 1 est 0), on ne peut pas calculer la variance de N en multipliant la variance de I_1 par 100.
- ▶ **Admis** : la variance de $N = I_1 + \dots + I_{100}$ est **inférieure** à $100\text{Var}(I_1)$.
- ▶ On connaît la variance d'une binomiale : $\text{Var}(I_1) = p(1 - p) \simeq 0.01762$.
- ▶ On peut donc majorer la variance de N : $\text{Var}(N) \leq 1.762$. (En réalité, elle est proche de 1.762)
- ▶ En posant $\sigma = \sqrt{\text{Var}(N)}$ (“écart-type”), on obtient $\sigma \leq 1.33$.
- ▶ Pour N , la valeur observée (2) est dans l'intervalle $[m - \sigma, m + \sigma]$: ce n'est pas atypique.
- ▶ Même calcul pour les carrés contenant 1 point.

Points dans le disque

- ▶ On veut tirer des points **uniformément** dans le **disque** centré en $(0,0)$, et de rayon 1.
- ▶ **Idée qui semble marcher** : on tire des points dans un carré qui contient le disque, par exemple $[-1, 1] \times [-1, 1]$; et on rejette ceux qui ne sont pas dans le disque (on tire des points dans le carré, jusqu'à en obtenir n qui sont dans le disque).
- ▶ **Deux questions** : est-ce que c'est "correct" (est-ce que les points ont la bonne loi) ? est-ce que ça ne risque pas d'être inefficace ?



Points dans le disque : correction

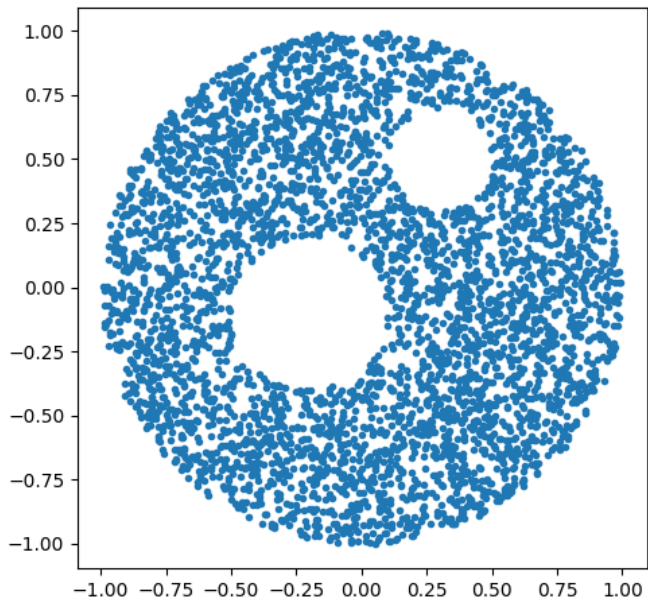
- ▶ C'est un schéma classique de "rejet" : on a une séquence de variables aléatoires $X_1, X_2, \dots$, indépendantes, de même loi uniforme sur un ensemble B ; et un ensemble $A \subset B$ de valeurs "intéressantes", à garder.
- ▶ On pose Y , la valeur de la première des X_i qui tombe dans A : $K = \min\{k > 0 : X_k \in A\}$ et $Y = X_K$.
- ▶ Alors, Y est uniforme sur l'ensemble A .
- ▶ Cœur du calcul : pour tout $k > 0$ et tous $a, a' \in A$, $\mathbb{P}(K = k, X_k = a) = \mathbb{P}(K = k, X_k = a')$, donc, en sommant sur k , $\mathbb{P}(Y = a) = \mathbb{P}(Y = a')$.
- ▶ Le calcul correspondant marche pour les lois uniformes sur un ensemble B fini, mais aussi pour la "loi uniforme" dans une surface (comme pour les points dans le disque), ou un volume.
- ▶ Donc, le rejet pour nos ensemble est correct : on passe de "uniforme dans le carré" à "uniforme dans le disque"
- ▶ La même méthode marcherait aussi pour tirer des points dans une autre forme (bornée) que le disque (il faut savoir dire si un point est dans la bonne zone).

Points dans le disque : complexité

- ▶ Que peut-on dire sur le nombre (variable aléatoire) de points du carré qu'on va tirer pour en obtenir n dans le disque ?
- ▶ Chaque point a probabilité $\pi/4$ de “tomber” dans le disque (aire du disque, divisée par l'aire du carré).
- ▶ **Donc**, le nombre de points à tirer pour en obtenir un dans le disque, suit la loi **géométrique** de paramètre $p = \pi/4 \simeq 0.78$. Espérance $4/\pi \simeq 1.27$, variance $(1 - p)/p^2 \simeq 0.35$.
- ▶ Le nombre total de tirages pour obtenir 100 points dans le disque est donc la somme de n géométriques indépendantes, toutes de paramètre $\pi/4$. (C'est une variable suivant une loi “binomiale négative”)
- ▶ On pourrait calculer la loi, mais on peut donner facilement l'espérance et la variance : $m = n/p$, variance $\sigma^2 = n(1 - p)/p^2$.
- ▶ **Numériquement, pour $n = 1000$:** $m \simeq 1273$, $\sigma^2 \simeq 347.9$ ($\sigma \simeq 18.65$).

Complexité (suite)

- ▶ D'après l'inégalité de Tchebycheff ($x = 10$) : la probabilité que le tirage de 1000 points du disque, demande de tirer plus de $m + 10\sigma \simeq 1459$ points du carré, est inférieure à $1/100$.
- ▶ Au final c'est assez efficace : avec bonne probabilité, on "paie" moins de 50% de plus ; en moyenne, seulement 27% de plus, par rapport au tirage dans le carré.
- ▶ L'efficacité dépend du rapport entre les aires.



Loi des grands nombres

- ▶ **Intuitivement**, si une expérience a probabilité p de “réussir”, en la répétant (de manière indépendante) un “grand” nombre de fois, la **proportion** des essais qui “réussissent” devrait être proche de p .
- ▶ **Plus généralement**, si on a une séquence de v.a. indépendantes de même loi $X_1, X_2, \dots$, toutes d'espérance m , on s'attend à ce que $S_n = (X_1 + \dots + X_n)/n$ soit également proche de m .
- ▶ (C'est effectivement plus général : quand les X_i sont des Bernoulli de paramètre p , la proportion d'essais réussis est bien S_n)
- ▶ Un tel phénomène porte le nom de **loi (faible) des grands nombres**, et c'est un théorème (au moins si les X_i ont une variance).

Loi des grands nombres

Théorème (loi faible des grands nombres)

Soit $X_1, X_2, \dots$ une suite de variables aléatoires indépendantes, toutes de même loi ; on suppose que les X_i ont une espérance m , et une variance $\sigma^2 < \infty$.

Alors, en posant $S_n = \frac{1}{n} \sum_{k=1}^n X_k$, on a, pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(m - \epsilon \leq S_n \leq m + \epsilon) = 1$$

Loi des grands nombres : preuve

- ▶ Chaque X_i est d'espérance m , variance σ^2 .
- ▶ Donc $X_1 + \dots + X_n$ est d'espérance $n.m$, de variance $n.\sigma^2$.
- ▶ Donc S_n est d'espérance m , de variance σ^2/n .
- ▶ On peut appliquer l'inégalité de Tchebycheff :

$$\mathbb{P}(|S_n - m| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon}$$

- ▶ Par passage au complémentaire :

$$\mathbb{P}(m - \epsilon < S_n < m + \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon}$$

- ▶ Le terme de droite (minoration de la proba) tend vers 1, donc la proba à gauche tend forcément vers 1.

Commentaires sur la loi des grands nombres

- ▶ Il existe une version “plus forte”, qu'on ne prouvera pas :

$$\mathbb{P}(\lim_{n \rightarrow \infty} S_n = m) = 1$$

- ▶ D'une certaine manière, ça justifie toutes les expériences qu'on a pu faire en simulation : chaque fois qu'on peut simuler l'expérience, et la répéter (indépendamment) suffisamment de fois, la *fréquence* des essais “réussis” devrait *converger* vers la *probabilité* de réussite.
- ▶ Pour faire plus précis (ex : statistiques), il faudrait savoir évaluer la *vitesse de convergence* : pour une valeur fixée de n , quelles valeurs de S_n sont “suffisamment éloignées de m ” pour qu'on soit amené à remettre en cause le modèle (loi de X) ?