CONCENTRATION ET LOI DES GRANDS NOMBRES

Rappel : un échantillon de taille n d'une loi de probabilité est une liste de n variables indépendantes suivant cette loi.

Dans ce chapitre, pour une variable aléatoire X, on notera $\mathrm{E}(\mathrm{X})$ son espérance et $\mathrm{V}(\mathrm{X})$ sa variance (on considère une variable aléatoire X dont l'espérance et la variance sont finies).

I. L'inégalité de Bienaymé-Tchebychev

DÉFINITION

Une variable aléatoire réelle X définie sur un univers Ω est dite ***positive*** si elle prend ses valeurs dans $\mathbb{R}^+$.

PROPRIÉTÉ INÉGALITÉ DE MARKOV

Si X est une variable aléatoire réelle positive : $\forall\ \delta>0,\ \mathrm{p}(\mathrm{X}\geqslant\delta)\leqslant\dfrac{\mathrm{E}(\mathrm{X})}{\delta}$.

delta

Démonstration :

Soit X une variable aléatoire réelle positive, dont on note $x_1,\ x_2,\dots,\ x_n$ les n valeurs.

Soit $\delta>0$. On note I l'ensemble des valeurs x prises par X telles que $x<\delta$, et S l'ensemble des valeurs x prises par X telles que $x\geqslant\delta$.

$$\mathrm{E}(\mathrm{X})=\sum_{i=1}^{n} x_i\,\mathrm{p}(\mathrm{X}=x_i)=\underline{\sum_{x\in\mathrm{I}} x\,\mathrm{p}(\mathrm{X}=x)+\sum_{x\in\mathrm{S}} x\,\mathrm{p}(\mathrm{X}=x)}$$

Or : • $\underline{\sum_{x\in\mathrm{I}} x\,\mathrm{p}(\mathrm{X}=x)\geqslant 0}$ puisque pour tout entier $i\in[\![1\,;n]\!]$, $x_i\geqslant 0$ et $\mathrm{p}(\mathrm{X}=x_i)\geqslant 0$

• $\sum_{x\in\mathrm{S}} x\,\mathrm{p}(\mathrm{X}=x)\geqslant\sum_{x\in\mathrm{S}}\delta\,\mathrm{p}(\mathrm{X}=x)$

ie $\sum_{x\in\mathrm{S}} x\,\mathrm{p}(\mathrm{X}=x)\geqslant\delta\sum_{x\in\mathrm{S}}\mathrm{p}(\mathrm{X}=x)$

ie $\underline{\sum_{x\in\mathrm{S}} x\,\mathrm{p}(\mathrm{X}=x)\geqslant\delta\,\mathrm{p}(\mathrm{X}\geqslant\delta)}$.

D'où : $\mathrm{E}(\mathrm{X})\geqslant\delta\,\mathrm{p}(\mathrm{X}\geqslant\delta)$ et, δ étant strictement positif, $\mathrm{p}(\mathrm{X}\geqslant\delta)\leqslant\dfrac{\mathrm{E}(\mathrm{X})}{\delta}$.

Attention : le caractère universel de cette inégalité a pour contrepartie le fait qu'elle est loin d'être optimale.

Andreï Andreïevitch Markov[1] (1856-1922) était un mathématicien russe →

1 Anecdote : il est cité par le rappeur Freeze Corleone dans son morceau *Tarkov* sur l'album LMF : « J'suis là pour les sous, à propos des chiffres, s/o Andreï Markov » (« s/o » veut dire « shout out » soit "dédicace" en anglais).

PROPRIÉTÉ INÉGALITÉ DE BIENAYMÉ-TCHEBYCHEV

Si X est une variable aléatoire réelle positive : $\forall\ \delta>0\ ,\ p(|X-E(X)|\geqslant\delta)\leqslant\dfrac{V(X)}{\delta^2}$.

Démonstration :

1. À l'aide de l'inégalité de Markov, majorer la probabilité $p((X-E(X))^2\geqslant\delta^2)$.

2. Justifier que $p((X-E(X))^2\geqslant\delta^2)=p(|X-E(X)|\geqslant\delta)$.

3. Conclure

REMARQUE : sous les mêmes hypothèses, on obtient aussi « $\forall\ \delta>0\ ,\ p(|X-E(X)|<\delta)\geqslant 1-\dfrac{V(X)}{\delta^2}$ ».

Ce théorème doit son nom aux mathématiciens Irénée-Jules Bienaymé (1796 - 1878), qui fut le premier à le formuler, et Pafnouti Tchebychev (1821 - 1894) qui le démontra.

EXEMPLE C1

Une variable aléatoire X suit la loi binomiale de paramètres $n=850$ et $p=0{,}16$.

On note μ l'espérance de X et σ son écart-type.

1. À l'aide de l'inégalité de Bienaymé-Tchebychev, minorer la probabilité $p(\mu-2\sigma<X<\mu+2\sigma)$.

2. À l'aide de la calculatrice, déterminer une valeur approchée à 10^{-3} de $p(\mu-2\sigma<X<\mu+2\sigma)$.

EXEMPLE A1

p. 469 SF2

Une variable aléatoire Y suit la loi binomiale de paramètres 700 et 0,72.

a. Majorer la probabilité $p(|Y-504|\geqslant\sqrt{700})$ à l'aide de l'inégalité de Bienaymé-Tchebychev.

b. Estimer $p(|Y-504|\geqslant\sqrt{700})$ par simulation.

II. Loi des grands nombres

DÉFINITION

Soit X une variable aléatoire. La ***variable aléatoire moyenne*** M_n d'un échantillon $(X_1 ; X_2 ; \ldots ; X_n)$ de taille n de la loi de X est : $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

PROPRIÉTÉS

Soit X une variable aléatoire. Pour tout échantillon $(X_1 ; X_2 ; \ldots ; X_n)$ de taille n de la loi de X dont la variable aléatoire moyenne est M_n : $E(M_n) = E(X)$ et $V(M_n) = \frac{1}{n} V(X)$.

Démonstrations :

- $E(M_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right)$

$= \frac{1}{n}\sum_{i=1}^{n} E(X_i)$ d'après la linéarité de l'espérance

Or, pour tout entier i de $[1 ; n]$: $E(X_i) = E(X)$

donc : $E(M_n) = \frac{1}{n} \times n\, E(X) = E(X)$.

- $V(M_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \left(\frac{1}{n}\right)^2 V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right)$

Or, les X_i sont indépendantes et $V(X_i) = V(X)$

donc $V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} V(X) = n\, V(X)$.

D'où : $V(M_n) = \frac{1}{n^2} \times n\, V(X) = \frac{1}{n} V(X)$.

PROPRIÉTÉ INÉGALITÉ DE CONCENTRATION

Soit X une variable aléatoire. Pour tout échantillon $(X_1 ; X_2 ; \ldots ; X_n)$ de taille n de la loi de X dont la variable aléatoire moyenne est M_n :

$$\forall\ \delta > 0,\ p(|M_n - E(X)| \geqslant \delta) \leqslant \frac{V(X)}{n\delta^2}.$$

Démonstration :

EXEMPLE A2


p. 469 SF3

Une variable aléatoire X suit la loi de Bernoulli de paramètre 0,13.
Pour un échantillon de taille n de la loi de X, on note M_n la variable aléatoire moyenne associée.

a. Justifier que $p(|M_n-0{,}13|\geqslant 0{,}12)\leqslant \dfrac{377}{48n}$.

b. En déduire la valeur de n telle que l'inégalité de concentration assure que $p(|M_n-0{,}13|\geqslant 0{,}12)\leqslant 0{,}05$.
Interpréter.

> **THÉORÈME** LOI (FAIBLE) DES GRANDS NOMBRES
>
> Soit X une variable aléatoire. Pour tout échantillon $(X_1;X_2;\dots;X_n)$ de taille n de la loi de X dont la variable aléatoire moyenne est M_n :
>
> $$\lim_{n\to+\infty} p(|M_n-E(X)|\geqslant\delta)=0.$$

Autrement dit, plus la taille n d'un échantillon d'une variable aléatoire X est grande, plus l'écart entre la moyenne de cet échantillon et l'espérance de la variable aléatoire X est proche de 0.

Démonstration :

EXEMPLE C2

On effectue n lancers successifs indépendants d'une pièce équilibrée. On associe à chaque tirage n°i la variable aléatoire X_i prenant la valeur 0 si on obtient « face », la valeur 1 sinon.

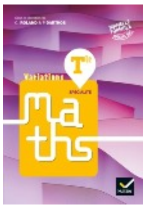
On pose $S_n=X_1+X_2+\dots+X_n$ (nombre de « pile » obtenu) et $M_n=\dfrac{S_n}{n}$.

1. On effectue 10 000 lancers. Démontrer que $p\left(\left|M_n-\dfrac{1}{2}\right|\geqslant 0{,}01\right)\leqslant\dfrac{1}{4}$, autrement dit que pour 10 000 lancers, la probabilité que la proportion de « pile » obtenue soit strictement inférieure à 0,49 ou strictement supérieure à 0,51 est inférieure à 0,25.

2. On souhaite que l'écart entre la proportion de « pile » obtenue et $\dfrac{1}{2}$ soit strictement inférieur à 0,01.

Avec l'inégalité de concentration, à partir de quelle valeur de n le risque d'erreur est inférieur ou égal à 5 % ?

→ BILAN DU CHAPITRE & TRAVAIL EN AUTONOMIE ←



- Fiche bilan → p.470
- QCM 11 questions corrigées → p.470
- Exercices corrigés → 21 à 25 p.472

• Méthodes et exercices corrigés en vidéo : → maths-et-tiques : tsm-clgn-ym

(cc) BY-NC-SA