Variables aléatoires, concentration et loi des grands nombres

Approche intuitive de la « loi des grands nombres » : Si je ne connais pas une loi de probabilité, par exemple celle d'un dé truqué, je peux obtenir des informations en répétant l'expérience un grand nombre de fois.

Loi de la variable $X+Y$ connaissant le tableau des lois de X et Y.

Soit X une variable aléatoire prenant les valeurs a_i et Y une variable prenant les valeurs b_j.

La variable aléatoire $X+Y$ prend alors toutes les valeurs possibles a_i+b_j.

La loi de probabilité de $X+Y$: pour toute valeur w prise par $X+Y$, la probabilité $P(X+Y=w)$ est la somme des probabilités $P(\{X=a_i\}\cap\{Y=b_j\})$ telles que $a_i+b_j=w$

Exemple :

Les lois de X et de Y :

X \ Y	0	1	2	Loi de X
0	0 0,07	1 0,18	2 0,17	0,42
1	1 0,15	2 0,20	3 0,23	0,58
Loi de Y	0,22	0,38	0,40	1

La loi de $X+Y$

s	0	1	2	3
$P(X=s)$	0,07	0,33	0,37	0,23

Loi de la variable aX connaissant la loi de X.

La variable aléatoire aX prend exactement les valeurs $a\times a_i$.

La loi de probabilité de aX : pour toute valeur $a\times a_i$ prise par aX, on a $P(aX=a\times a_i)=P(X=a_i)$

Exemple

Loi de X

Valeur de X	0	1	2
Probabilité	0.22	0.38	0.40

Loi de $10X$

Valeur de $10X$	0	10	20
Probabilité	0.22	0.38	0.40

Propriété de l'espérance et de la variance de $X+Y$ et de aX

- $E(X+Y)=E(X)+E(Y)$
- $E(aX)=aE(X)$

et plus généralement $E(aX+bY)=aE(X)+bE(Y)$, on dit que la variance est linéaire.

- $V(X+Y)=V(X)+V(Y)$ lorsque X et Y sont indépendantes, c'est-à-dire que l'une n'influe pas sur l'autre.
- $V(aX)=a^2V(X)$

Remarque : Si X suit une loi binomiale alors X s'écrit $X_1+X_2+\cdots+X_n$ avec les X_i des variables aléatoires indépendantes suivant la même loi de Bernoulli de paramètre p. On retrouve alors $E(X)=np$ et $V(X)=np(1-p)$

Définition (échantillon d'une loi)

Un échantillon d'une loi de taille n est une liste $(X_1,X_2,\dots,X_n)$ de variables aléatoires indépendantes et identiques qui suivent toutes cette loi.

Exemple : Je prends le même train cinq jours par semaine. On admet que la variable aléatoire X qui compte le nombre de retard sur 5 jours suit une $\mathcal{B}(n=5;p=0.02)$. En répétant cette expérience pendant 10 semaines on construit un échantillon $(X_1,X_2,\dots,X_{10})$ de cette loi binomiale.

Propriété (espérance, variance, écart-type de la moyenne d'un échantillon d'une loi) :

Si $(X_1,X_2,\dots,X_n)$ est un échantillon d'une loi X dont l'espérance est m et la variance v alors :

- $E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\frac{nm}{n}=m$
- $V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\frac{nv}{n^2}=\frac{v}{n}$

Interprétation :

La moyenne de l'échantillon est identique à celle de la loi mais la dispersion diminue à mesure que n grandit.

Lorsqu'on réalise plusieurs mesures, chaque nouvelle mesure nous rapproche de la certitude que la moyenne empirique $\frac{X_1+X_2+\cdots+X_n}{n}$ est proche de l'espérance m.

Loi des grands nombres

On veut montrer que plus on prend de mesures, plus l'échantillon est grand, plus la moyenne de l'échantillon est proche de la moyenne de la loi échantillonnée. Donc quand on ne connait pas la moyenne d'une loi, il est judicieux de répéter des expériences. On va montrer la loi des grands nombres, c'est-à-dire

$$P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n}-m\right|\geqslant\varepsilon\right)\underset{n\to+\infty}{\longrightarrow}0 \text{ ceci pour tout } \varepsilon>0$$

Propriété : l'inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire quelconque d'espérance $E(X)=m$ et de variance $V(X)=v$. (il faut qu'elles existent)
Alors pour tout a réel :

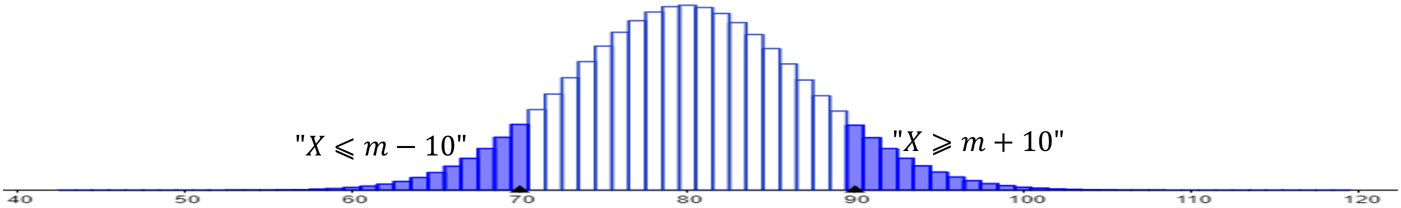
$$P(|X-m|\geqslant a)\leqslant\frac{v}{a^2}$$

Interprétation *:*

Cette inégalité majore le risque qu'une variable aléatoire s'écarte de sa moyenne, c'est-à-dire que $|X-m|\geqslant a$ se produise pour un a donné. Ce risque est géré par la variance. Plus a est grand, plus ce risque est faible.
En appliquant cette propriété à $\frac{X_1+X_2+\cdots+X_n}{n}$ on obtiendra la loi des grands nombres.

Exemple :

Soit X suivant une $\mathcal{B}(n=200,p=0.4)$. L'événement "$|X-m|\geqslant 10$" est représenté par les 2 zones suivantes :



L'inégalité de Bienaymé donne $P(|X-m|\geqslant 10)\leqslant\frac{200\times0.4\times0.6}{10^2}=0.48$

Conséquence (*inégalité de concentration et loi des grands nombres*):

En appliquant cette propriété à $X=\frac{X_1+X_2+\cdots+X_n}{n}$, avec $(X_1,X_2,\ldots,X_n)$ un échantillon d'une loi d'espérance m et de variance v, en sachant qu'on a $E(X)=m$ et $V(X)=\frac{v}{n}$

on obtient l'inégalité de concentration : Pour tout $a>0$ on a :

$$P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n}-m\right|\geqslant a\right)\leqslant\frac{v}{na^2}$$

et lorsque $n\to+\infty$ on obtient la loi (faibles) des grands nombres : Pour tout $\varepsilon>0$ on a :

$$P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n}-m\right|\geqslant\varepsilon\right)\underset{n\to+\infty}{\longrightarrow}0$$

L'inégalité de Irénée-Jules Bienaymé et Pafnouti Tchebychev

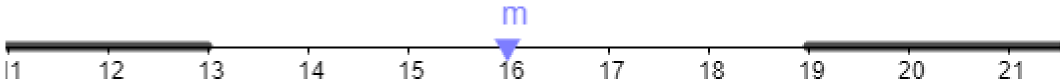
Soit X une variable aléatoire quelconque d'espérance $E(X) = m$ et de variance $V(X) = v$. (il faut qu'elles existent)
Alors pour tout a réel :

$$P(|X - m| \geqslant a) \leqslant \frac{v}{a^2}$$

Démonstration :
Il s'agit de partitionner l'ensemble des valeurs $X(\Omega)$ en 2 ensembles, définis par la valeur de a :

- $Z_{ext} = \{x_i \; tels \; que \; |x_i - m| \geqslant a\}$
- $Z_{int} = \{x_i \; tels \; que \; |x_i - m| < a\}$

Sur ce schéma la zone autour de m est Z_{int} et la zone en noire est Z_{ext}



On cherche à montre que

$$P(X \in Z_{ext}) \leqslant \frac{v}{a^2}$$

Pour cela exprimons la variance :

$$V(X) = \sum_{x_i \in X(\Omega)} p_i (x_i - m)^2$$

$$V(X) = \sum_{x_i \in Z_{int}} p_i (x_i - m)^2 + \sum_{x_i \in Z_{ext}} p_i (x_i - m)^2$$

Car si $x_i \in Z_{ext}$ on a $|x_i - m| \geq a$
donc $(x_i - m)^2 \geq a^2$

$$V(X) \geq \sum_{x_i \in Z_{int}} p_i (x_i - m)^2 + a^2 \sum_{x_i \in Z_{ext}} p_i$$

$$V(X) \geq 0 + a^2 P(X \in Z_{ext})$$

On obtient comme convenu $P(X \in Z_{ext}) \leqslant \frac{v}{a^2}$