Variables aléatoires - Loi des grands nombres

Probabilité - Chapitre 3

Table des matières

Dans ce chapitre, on considérera une expérience aléatoire dont l'univers, c'est-à-dire l'ensemble des issues, Ω est fini. On notera, ω_1, ω_2, ..., ω_m les m **issues** de l'expérience aléatoire, càd les éléments de Ω.

I Variables aléatoires et somme de variables aléatoires

I 1 Rappel sur la définition d'une variable aléatoire.

DEFINITION : Variable aléatoire

Une **variable aléatoire** X sur Ω est une **fonction** définie sur Ω (ensemble des éventualités) et à valeurs dans $\mathbb{R}$ (au lycée).

$$\boxed{X : \omega \mapsto X(\omega)}$$

On notera x_1, x_2, ..., x_n les n **valeurs** réelles prises par la variable aléatoire X et E l'ensemble de tous les x_i.
Ainsi, $\forall \omega_i \in \Omega$, $X(\omega_i) = x_j$ avec $1 \leqslant j \leqslant n$.
(*Faire diagramme de Venn, attention les cardinaux de ces 2 ensembles peuvent être différents, précisément :* $\mathrm{card}(E) \leqslant \mathrm{card}(\Omega)$).

Exemple 1 : On lance une pièce de monnaie. Si on obtient pile, on gagne 5 euros et si on obtient face, on gagne 2 euros.
On peut alors définir une variable aléatoire X correspondant au gain obtenu en euros.
X est définie sur l'univers $\Omega = \{\text{pile}; \text{face}\}$.
On a alors $X(\text{pile}) = 5$ et $X(\text{face}) = 2$. X peut donc prendre les valeurs 5 et 2.

Exemple 2 : On lance un dé cubique classique équilibré. Si la face 6 sort alors on gagne 5 euros, sinon on perd 1 euro.
Soit X la variable aléatoire associant aux issues de l'expérience leur gain algébrique.
On a donc $X(6) = 5$ et $X(1) = X(2) = X(3) = X(4) = X(5) = -1$
Ainsi $\Omega = \{1; 2; 3; 4; 5; 6\}$ et $E = \{5; -1\}$

I 2 Rappels sur l'application d'une transformation affine à une VA

DEFINITION : (*faire diagramme de Venn*)

n et m sont entiers naturels non nuls. Soit X une variable aléatoire définie sur $\Omega = \{\omega_1; \omega_2; ...\omega_m\}$ et x_1, x_2, ...x_n ses valeurs. Soit a et b deux réels.

Alors la variable aléatoire $Y = aX + b$ est telle que : $\boxed{Y : \omega_i \mapsto Y(\omega_i) = aX(\omega_i) + b}$, $\forall i \in \{1; 2; ...; m\}$.

Les **valeurs prises** par Y peuvent donc se noter : $\boxed{y_j = ax_j + b}$, $\forall j \in \{1; 2; ...; n\}$

PROPRIETE : démontrée en première et revue Chap P2

Avec les données de la définition précédente.

L'espérance est **linéaire** donc : $\boxed{E(Y) = E(aX + b) = aE(X) + b}$

La variance est telle que $\boxed{V(Y) = V(aX + b) = V(aX) = a^2V(X)}$

I 3 Somme de variables aléatoires

DEFINITION :

Si X et Y sont deux variables aléatoires définies sur un même univers Ω ,
alors la variable aléatoire Z, qui à tout élément ω de Ω associe $Z(\omega) = X(\omega) + Y(\omega)$, notée $Z = X + Y$, est **la somme des variables aléatoires de X et Y**.

Exercice 1

En période de réglage de ses machines, le contrôleur qualité d'une entreprise effectue de nombreux prélèvements dans la production et relève, pour chaque pièce, si elle a un défaut A, un ou deux défauts B. Compte-tenu du grand nombre de pièces prélevées, les fréquences indiquées dans ce tableau peuvent être assimilées à des probabilités.
X (resp. Y) est la variable aléatoire qui compte le nombre de défauts A (resp. B).

X $\downarrow$ et Y $\rightarrow$	0	1	2
0	0,07	0,18	0,17
1	0,15	0,20	0,23

1. A partir de ces données, déterminer la loi de probabilité de X.
2. A partir de ces données, déterminer la loi de probabilité de Y.
3. A partir de ces données, déterminer la loi de probabilité de S=X+Y.

Exercice 2

Une urne contient six boules : trois vertes, numérotées 1, 2 et 3 et trois rouges, numérotées 5, 6 et 7. On effectue un tirage au hasard d'une boule dans cette urne. Soit X la variable aléatoire qui prend la valeur 1 si la boule est verte et 10 si elle est rouge. Soit Y la variable aléatoire qui prend la valeur du numéro inscrit sur la boule tirée.

1. Donner les valeurs prises par la variable aléatoire X.
2. Donner les valeurs prises par la variable aléatoire Y.
3. Donner les valeurs prises par la variable aléatoire $Z = X + Y$.

I 4 Espérance d'une somme de variables aléatoires

PROPRIETE : donnée dans le Chap P2

Si X et Y sont deux variables aléatoires définies sur un même univers Ω alors :

L'espérance de la variable aléatoire $Z = X + Y$ est $E(Z) = \boxed{E(X+Y) = E(X) + E(Y)}$

REMARQUE(S) :

$E(aX + b) = aE(X) + b$ et $E(X + Y) = E(X) + E(Y)$ traduisent **la linéarité de l'espérance**.

DEMONSTRATION :

Cf. manuel p 408, démonstration sur un cas particulier. Cas général complexe à mettre en forme avec rigueur.

Exercice 3

On lance 5 dés équilibrés et on compte la somme des nombres obtenus.
Soit X la variable aléatoire correspondant à cette somme.
On peut écrire X sous la forme $X = X_1 + X_2 + X_3 + X_4 + X_5$ où X_k (pour $k \in \{1; 2; 3; 4; 5\}$) correspond au résultat du dé numéroté k.

1. Quel est l'ensemble des valeurs prises par X ?
2. Calculer l'espérance de X.
3. Si maintenant on pose $Y = 5X_1$. Quel est l'ensemble des valeurs prises par Y ?
4. Calculer l'espérance de Y.

Exercice 4

On joue à un jeu se déroulant en deux étapes.
Dans la phase 1, on lance un dé équilibré à six faces.
Si le résultat obtenu est 1 ou 6 alors on gagne 9 points ; sinon on perd 6 points.

Dans la phase 2, on lance une pièce équilibrée. Si on obtient face, on gagne 6 points ; sinon on perd 2 points.

Soit X la variable aléatoire correspondant au nombre total de points obtenus.
Calculer de deux manières différentes $E(X)$.

I 5 Variance d'une somme de variables aléatoires INDEPENDANTES

DEFINITION : généralisation

Soient X_1, X_2, ..., X_n, n variables aléatoires à valeurs respectives dans E_1, ..., E_n.

On dit que X_1, X_2, ..., X_n sont **indépendantes** (ou **mutuellement indépendantes**) lorsque, pour tout $x_1 \in E_1$, $x_2 \in E_2$, ..., $x_n \in E_n$:

$P(X_1 = x_1 \cap X_2 = x_2 \cap ... \cap X_n = x_n) = P(X_1 = x_1) \times P(X_2 = x_2) \times ... \times P(X_n = x_n)$.

REMARQUE(S) :

(HP) Ne pas confondre **mutuellement indépendants** et **deux à deux indépendants**.

Par exemple, dans le cas du lancer de deux dés, si on pose :
X_1 (resp. X_2) la VA prenant la valeur 1 si le premier (resp. second) dé est pair et 0 sinon,
et Y la VA prenant la valeur 1 si la somme des deux résultats est paire et 0 sinon.

On vérifiera que X_1, X_2 et Y sont **deux à deux indépendants** mais **pas mutuellement indépendants**. En effet :

- Connaître la parité de la somme n'aide pas à connaître celle d'un dé.
- Connaître la parité d'un dé n'aide pas à déterminer celle de la somme, ni de l'autre dé.
- En revanche, connaître la parité des deux dés permet de connaître celle de la somme.

PROPRIETE : (admise)

Si X et Y sont deux variables aléatoires **INDÉPENDANTES** définies sur un même univers Ω alors :

La variance de la variable aléatoire $Z = X + Y$ est $V(Z) = \boxed{V(X+Y) = V(X) + V(Y)}$

DEMONSTRATION :

ex 76 (complexe et HP) et 77 p 405 LLS (poly sur le blog)

Exercice 5

Reprendre les exemples précédents et calculer la variance de chaque VA.

II Application à un échantillon de n VA indépendantes de même loi de probabilité

DEFINITION :

Soit n un entier naturel non nul. Soit X une variable aléatoire définie sur l'ensemble Ω des issues d'une expérience aléatoire.

Un **échantillon de taille n de la loi de probabilité** de X est une liste $(X_1,\ X_2,\ ...,\ X_n)$, de variables aléatoires **indépendantes** et **identiques**, suivant toutes cette **même** loi de probabilité.
On peut alors définir :

• la **variable aléatoire somme** S_n d'un échantillon de taille n de la loi de X par $\boxed{S_n = \sum_{i=1}^{n} X_i = X_1 + X_2 + ... + X_n}$

• la **variable aléatoire moyenne** M_n d'un échantillon de taille n de la loi de X par $\boxed{M_n = \frac{1}{n} S_n = \frac{1}{n}\sum_{i=1}^{n} X_i}$

PROPRIETE : Somme

Soit n un entier naturel non nul. Soit X une variable aléatoire définie sur un univers Ω.
Soit S_n la variable aléatoire somme d'un échantillon de taille n de la variable aléatoire X. On a alors :

$$\boxed{E(S_n) = nE(X)} \qquad \boxed{V(S_n) = nV(X)} \qquad \boxed{\sigma(S_n) = \sqrt{n}\ \ \sigma(X)}$$

DEMONSTRATION :

- La linéarité de l'espérance donne $E(S_n) = E(X_1) + E(X_2) + ... + E(X_n)$.
  Or ces variables aléatoires suivent la même loi, elles ont donc la même espérance : $E(X)$. Donc $E(S_n) = E(X) + E(X) + ... + E(X) = nE(X)$.

- De même, les variables aléatoires X_i (pour $i \in \{1; 2; ...; n\}$ sont **indépendantes** et suivent la même loi donc on la même variance $V(X)$, on a donc $V(S_n) = V(X_1) + V(X_2) + ... + V(X_n) = V(X) + V(X) + ... + V(X) = nV(X)$.

- Par passage à la racine, $\sigma(S_n) = \sqrt{V(S_n)} = \sqrt{nV(X)} = \sqrt{n} \times \sqrt{V(X)} = \sqrt{n} \ \ \sigma(X)$.

PROPRIETE : Moyenne

Soit n un entier naturel non nul.
Soit M_n la variable aléatoire moyenne d'un échantillon de taille n de la variable aléatoire X. On a alors :

$$\boxed{E(M_n) = E(X)} \qquad \boxed{V(M_n) = \frac{1}{n}V(X)} \qquad \boxed{\sigma(M_n) = \frac{\sigma(X)}{\sqrt{n}}}$$

DEMONSTRATION :

- $E(M_n) = E\left(\frac{1}{n}S_n\right) = \frac{1}{n}E(S_n) = \frac{1}{n} \times nE(X) = E(X)$.
- $V(M_n) = V\left(\frac{1}{n}S_n\right) = \left(\frac{1}{n}\right)^2 V(S_n) = \left(\frac{1}{n}\right)^2 \times nV(X) = \frac{1}{n}V(X)$. Rappel : $V(aX) = a^2V(X)$
- $\sigma(M_n) = \sqrt{V(M_n)} = \sqrt{\frac{1}{n}V(X)} = \frac{\sqrt{V(X)}}{\sqrt{n}} = \frac{\sigma(X)}{\sqrt{n}}$.

REMARQUE(S) :

$V(M_n) = \frac{1}{n}V(X)$ montre que la variance diminue quand la taille de l'échantillon augmente.

Elle quantifie la fluctuation d'échantillonnage, càd l'**écart moyen** entre les valeurs prises par la VA et son espérance (précisément, le carré de cet écart).

III Inégalités de Markov et de Bienaymé-Tchebychev

III 1 Inégalité de Markov

PROPRIETE : Inégalité de Markov

Soit X est une variable aléatoire réelle **positive ou nulle** (toutes ses valeurs sont des réels positifs ou nuls) et $E(X)$ son espérance.

On a alors $\forall a \in]0\,;+\infty[$, $\boxed{P(X \geqslant a) \leqslant \dfrac{E(X)}{a}}$.

DEMONSTRATION :

Soit X une variable aléatoire à valeurs positives ou nulles dont on note x_i les n valeurs pour i, entier entre 1 et n.

Par définition, on a $E(X) = \displaystyle\sum_{i=1}^{n} x_i P(X = x_i)$.

Séparons cette somme en deux blocs selon les valeurs supérieures ou égales à a et celles strictement inférieures à a.

On obtient : $E(X) = \displaystyle\sum_{x_i \geqslant a} x_i P(X = x_i) + \sum_{x_i < a} x_i P(X = x_i)$.

Pour tout entier i entre 1 et n, on sait que $x_i \geqslant 0$ (on a posé que X prenait des valeurs positives)

et par définition d'une probabilité $P(X = x_i) \geqslant 0$ donc $\displaystyle\sum_{x_i < a} x_i P(X = x_i) \geqslant 0$.

Par suite, $E(X) \geqslant \sum_{x_i \geqslant a} x_i P(X = x_i)$. Par ailleurs, dans cette partie de somme, on a $\forall i \in \{1;2;...;n\}$, $\boxed{x_i \geqslant a}$ donc $E(X) \geqslant \sum_{x_i \geqslant a} aP(X = x_i)$.

Ainsi, $E(X) \geqslant a \sum_{x_i \geqslant a} P(X = x_i)$, soit $E(X) \geqslant a\ P(X \geqslant a)$.

Et puisque $a > 0$, on a donc $\boxed{P(X \geqslant a) \leqslant \dfrac{E(X)}{a}}$.

REMARQUE(S) :

$$\boxed{P(X \geqslant a) \leqslant \frac{E(X)}{a}}$$

- Cela signifie que la probabilité que les valeurs prises par X soient supérieures à a est d'autant **plus petite que a est grand**.
- Cette inégalité permet de trouver un majorant mais **pas nécessairement le plus petit possible**.
- Si $a \leqslant E(X)$ alors l'inégalité de Markov n'a **aucun intérêt** car $\dfrac{E(X)}{a} \geqslant 1$ et on obtient donc $P(X \geqslant a) \leqslant 1$ ce qui est évident.

Exercice 6 (Résolu)

En 2015, le salaire brut mensuel moyen en France était de 2 442 euros.
On choisit un salarié au hasard et on note X la VA donnant son salaire.

. . . .

Les salaires étant positifs ou nuls, on sait que X est une VA positive ou nulle et on peut donc appliquer l'inégalité de Markov sur un exemple :

$$P(X \geqslant 3 \times 2442) \leqslant \frac{2442}{3 \times 2442} \text{ soit } P(X \geqslant 3 \times 2442) \leqslant \frac{1}{3}.$$

La probabilité que le salaire brut mensuel d'un français soit supérieure ou égale à 3 fois le salaire moyen est donc inférieure ou égale à $\dfrac{1}{3}$.

Exercice 7

Une usine produit en moyenne 35 pièces par semaine.
On note X la VA donnant le nombre de pièces produites par semaine.
Que peut-on dire de la probabilité que l'usine produise plus de 70 pièces par semaine ?

III 2 Inégalité de Bienaymé-Tchebychev

PROPRIETE : Inégalité de Bienaymé-Tchebychev

Soit X est une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$.

On a alors, $\forall a \in]0\,;+\infty[$, $\boxed{P(|X-E(X)| \geqslant a) \leqslant \dfrac{V(X)}{a^2}}$.

REMARQUE(S) :

- Cela signifie que la probabilité que les valeurs prises par X s'écartent d'au moins a de l'espérance $E(X)$ est d'autant plus petite que a est grand.
- L'intervalle $[E(X)-a\,;E(X)+a]$ est un intervalle de fluctuation de la VA X.
- Sous les mêmes conditions, l'inégalité peut s'écrire également $\boxed{P(|X-E(X)| < a) \geqslant 1-\dfrac{V(X)}{a^2}}$.

  En effet, $P(|X-E(X)| \geqslant a) = 1-P(|X-E(X)| < a)$, donc d'après BT, $1-P(|X-E(X)| < a) \leqslant \dfrac{V(X)}{a^2}$

  càd : $1-\dfrac{V(X)}{a^2} \leqslant P(|X-E(X)| < a)$.

DEMONSTRATION :

Comme $a>0$, les inégalités $|X-E(X)| \geqslant a$ et $|X-E(X)|^2 \geqslant a^2$ sont équivalentes.

De plus, la VA $|X-E(X)|^2$ est positive ou nulle. On applique donc l'inégalité de Markov à la VA $|X-E(X)|^2$ et au réel a^2.

Ainsi, $P(|X-E(X)|^2 \geqslant a^2) \leqslant \dfrac{E(|X-E(X)|^2)}{a^2}$.

Or, par définition, $E(|X-E(X)|^2) = V(X)$ donc $P(|X-E(X)|^2 \geqslant a^2) \leqslant \dfrac{V(X)}{a^2}$

on a donc bien $P(|X-E(X)| \geqslant a) \leqslant \dfrac{V(X)}{a^2}$.

Exercice 8 (Résolu)

Le taux moyen de glycémie dans une population est de 1g.L^{-1} avec une variance de 0,1.
Une personne présente un taux X critique si son taux ne se situe pas dans l'intervalle $]0,5\,;1,5[$.

....

Cet événement se traduit par l'inégalité $|X-E(X)| \geqslant 0,5$.

Sa probabilité vérifie donc $P(|X-E(X)| \geqslant 0,5) \leqslant \dfrac{0,1}{0,5^2}$ soit $P(|X-E(X)| \geqslant 0,5) \leqslant 0,4$.

Ainsi la probabilité qu'une personne présente un taux critique est inférieure ou égale à 0,4.

Exercice 9

Lors d'une saison de football, le nombre moyen de buts par match est de 2,5 avec une variance de 1,1.
Majorer la probabilité que le match suivant ne se termine pas par 2 ou 3 buts.

III 3 Inégalité de Bienaymé-Tchebychev appliquée au cas particulier de la loi binomiale (récap)

1) On suppose une **épreuve de Bernoulli** à deux issues : succès et échec. La probabilité du succès étant un nombre réel p de l'intervalle $[0\ ;\ 1]$. On se souvient que la variable aléatoire X associée, prenant la valeur 1 en cas de succès et 0 sinon, suit la **loi de Bernoulli**.

x_i	1	0
$P(X = x_i)$	p	$q = 1 - p$

Nous avons montré au **chapitre P2** que :
l'**espérance** de X est $\boxed{E(X) = p}$, sa **variance** est $\boxed{V(X) = p(1-p)}$ et son **écart-type** est $\boxed{\sigma = \sqrt{p(1-p)}}$

2) On répète maintenant n fois cette épreuve de Bernoulli, de manière identique et indépendante, constituant ainsi un **schéma de Bernoulli** d'ordre n.

A chaque i-ème épreuve, (avec i entier naturel compris entre 0 et n inclus), on considère la variable aléatoire X_i de Bernoulli, qui suit la loi de X, càd la loi de Bernoulli, celle de la variable aléatoire X. La liste des X_i constitue un **échantillon de taille** n de la loi de probabilité de X.

On définit également la variable aléatoire $S_n = X_1 + X_2 + X_3 + ... + X_n$, **somme** des variables aléatoires précédentes. Cette variable aléatoire S_n compte le nombre de succès obtenus parmi les n épreuves réalisées et d'après le cours, elle suit la loi Binomiale de paramètres n et p.

Nous avons montré au **chapitre P2** que : l'**espérance** de S_n est $\boxed{E(S_n) = np}$, sa **variance** est $\boxed{V(S_n) = np(1-p)}$ et son **écart-type** est $\boxed{\sigma(S_n) = \sqrt{np(1-p)}}$

3) Enfin, posons M_n la variable aléatoire **moyenne** de l'échantillon précédemment défini : $M_n = \dfrac{1}{n} S_n$. On montre alors que :

l'**espérance** de M_n est $\boxed{E(M_n) = E\left(\dfrac{1}{n} S_n\right) = \dfrac{1}{n} E(S_n) = \dfrac{1}{n} \times np = p}$ (linéarité de l'espérance), sa **variance** est $\boxed{V(M_n) = V\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}V(S_n) = \frac{1}{n^2} \times np(1-p) = \frac{p(1-p)}{n}}$

et son **écart-type** est $\boxed{\sigma(M_n) = \sqrt{V(M_n)} = \sqrt{\frac{p(1-p)}{n}}}$

4) Maintenant appliquons l'**inégalité de Bienaymé-Tchebychev** aux deux variables aléatoires S_n et M_n :

$$\boxed{P(|S_n - np| \geqslant a) \leqslant \frac{np(1-p)}{a^2}} \text{ et } \boxed{P(|M_n - p| \geqslant a) \leqslant \frac{p(1-p)}{n \times a^2}}$$

Exercice 10

On lance 100 fois un dé équilibré tétraédrique, numéroté de 1 à 4. On observe la face cachée du dé et le succès est obtenu lorsque cette face est celle numérotée 4. Étudier $P(21 \leqslant S_n \leqslant 29)$ en utilisant l'inégalité de BT.

IV Loi des grands nombres

IV 1 Inégalité de concentration

PROPRIETE : Inégalité de concentration

Soit X est une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$.

On pose M_n **variable aléatoire moyenne** d'un échantillon de taille n de X ; autrement dit, $M_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$,

où les variables X_i sont **indépendantes**, **identiques** et suivent toutes la **même** loi de probabilité (celle de X). On a alors

$$\forall a \in]0\,;+\infty[,\ \boxed{P(|M_n - E(X)| \geqslant a) \leqslant \frac{V(X)}{na^2}} \quad \text{(inégalité de concentration)}$$

DEMONSTRATION :

On applique l'inégalité de Bienaymé-Tchebychev à M_n. On se souvient que $E(M_n) = E(X)$ et $V(M_n) = \dfrac{1}{n}V(X)$ donc

$$P(|M_n - E(M_n)| \geqslant a) \leqslant \frac{V(M_n)}{a^2} \Longleftrightarrow P(|M_n - E(X)| \geqslant a) \leqslant \frac{V(X)}{na^2}$$

Exercice 11

On effectue n lancers successifs supposés indépendants d'une pièce équilibrée.
On associe à chaque lancer i la variable aléatoire X_i prenant la valeur 0 si on obtient face et 1 si on obtient pile.
On pose $S_n = X_1 + X_2 + ... + X_n$ la variable aléatoire donnant le nombre de piles obtenu lors de ces n lancers.
On pose $M_n = \dfrac{S_n}{n}$.

1. Donner la loi de probabilité de X_i (pour i entier entre 1 et n).
2. Déterminer l'espérance $E(X_i)$ et la variance $V(X_i)$.
3. Appliquer l'inégalité de concentration avec $n = 10\ 000$ et $a = 0,01$ et interpréter ce résultat.
4. On souhaite maintenant que l'écart entre la proportion de pile obtenu et $\dfrac{1}{2}$ soit inférieur ou égal à 0,01.
   Quelle est la valeur minimale de n pour que le risque d'erreur soit inférieur ou égal à 5% ?

IV 2 Loi faible des grands nombres

PROPRIETE : Loi faible des grands nombres

Soit (X_n) un échantillon d'une variable aléatoire. On pose $M_n = \frac{1}{n}\sum_{k=1}^{n} X_k$. Alors pour tout réel a strictement positif,

$$\boxed{\lim_{n\to+\infty} P(|M_n - E(X)| \geqslant a) = 0}$$

DEMONSTRATION :

On applique l'inégalité de concentration à la VA M_n. Ainsi, $0 \leqslant P(|M_n - E(X)| \geqslant a) \leqslant \frac{V(X)}{na^2}$

Or $\lim_{n\to+\infty} \frac{1}{n} = 0$ donc $\lim_{n\to+\infty} \frac{V(X)}{na^2} = 0$

Donc d'après le théorème des gendarmes, $\lim_{n\to+\infty} P(|M_n - E(X)| \geqslant a) = 0$.