Chapitre 20. Espérance et variance.

Dans tout le chapitre, sauf mention contraire, (Ω, P) désigne un espace probabilisé fini.

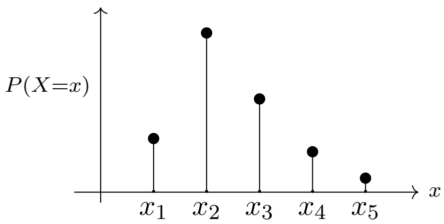
1 Espérance

1.a Définition

Définition :

Soit X une variable aléatoire réelle sur Ω. On appelle espérance de X le nombre :

Il s'agit donc de la moyenne des valeurs prises par X, pondérées par leur probabilité d'apparition.



Intuitivement, cela représente la valeur moyenne que l'on espère obtenir si on répète un très grand nombre de fois l'expérience.

Exemple : on lance un dé pipé : le dé tombe sur la face 6 avec probabilité $\frac{1}{3}$, sur les faces 1, 2 et 3 avec probabilité $\frac{1}{6}$, et sur les faces 4 et 5 avec probabilité $\frac{1}{12}$. Quelle valeur obtient-on en moyenne ?

Proposition :

Soit X une variable aléatoire réelle sur Ω. On a : $E(X) = \displaystyle\sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$.

Illustration :

1.b Les exemples à connaître par cœur

Proposition :

- Si X est une variable constante égale à λ, alors
- Si X est une variable de Bernoulli de paramètre p, alors
- Si X est une variable binomiale de paramètres n et p, alors

Démonstration 1

En particulier, si $X = \mathbb{1}_A$ avec A un événement, alors

1.c Théorème de transfert

Théorème :

Soit X une variable aléatoire réelle sur Ω, et g une fonction définie au moins sur $X(\Omega)$, à valeurs réelles.

$$E\left(g(X)\right) = \sum_{x \in X(\Omega)} g(x) P(X = x)$$

Ce théorème est très pratique ! Il permet de calculer l'espérance de $g(X)$ sans connaître la loi de $g(X)$, seulement celle de X.
(Juste avec la définition d'espérance, il faudrait calculer $\displaystyle\sum_{y \in g(X)(\Omega)} yP(g(X) = y)$ et donc calculer d'abord toutes les probabilités $P(g(X) = y)$...)

Exemple : Déterminer l'espérance de X^2 si la loi de X est donnée par :

x	-1	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Il y a une version similaire de ce théorème pour les couples :
Théorème :

(X, Y) un couple de variables aléatoires sur Ω, et g une fonction à valeurs réelles, telle que $g(X, Y)$ soit bien définie.

$$E\big(g(X, Y)\big) =$$

Cela va nous servir à calculer, par exemple, $E(XY)$, $E(e^{X+Y})$...
Ce théorème s'adapte au cas d'un n-uplet de variables aléatoires $(X_1, \ldots, X_n)$:

Reprenons l'exemple de couple de variables aléatoires dans le chapitre précédent (partie 3) :

$X\backslash Y$	0	1	2
0	$\frac{4}{9}$	0	0
1	0	$\frac{2}{9}$	$\frac{2}{9}$
2	0	$\frac{1}{9}$	0

Calculons $E(XY)$:

Le théorème de transfert permet de démontrer le résultat très important suivant :

Théorème :

> Si X et Y sont des variables aléatoires réelles indépendantes, alors $E(XY) = E(X)E(Y)$.

Démonstration 2

Cela s'étend à un n-uplet $(X_1, \ldots, X_n)$ de variables aléatoires réelles indépendantes : on a alors $E(X_1 \ldots X_n) = E(X_1) \ldots E(X_n)$.

⚠ La réciproque est fausse en général ! Il y a des variables aléatoires qui vérifient $E(XY) = E(X)E(Y)$ sans qu'elles soient indépendantes. Nous verrons un contre-exemple en TD.

Remarque : Par contre, pour calculer $E(X + Y)$ ou $E(X_1 + \cdots + X_n)$, on ne passera pas par le théorème de transfert, car on va voir que l'espérance est linéaire !

1.d Propriétés de l'espérance

Proposition :

> Soient X et Y des variables aléatoires réelles sur Ω.
>
> - (Linéarité) Pour tout $(\lambda, \mu) \in \mathbb{R}^2$, $E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$.
> - (Positivité) Si $X \geq 0$ (c'est-à-dire X à valeurs positives), alors $E(X) \geq 0$.
> - (Croissance) Si $X \geq Y$, alors $E(X) \geq E(Y)$.
> - (Inégalité trinagulaire) $|E(X)| \leq E(|X|)$.

Démonstration 3

Remarque : En particulier, grâce à la linéarité, si a et b sont des réels : $E(aX + b) =$

Définition :

> Une variable aléatoire réelle X sur Ω est dite <u>centrée</u> si $E(X) = 0$.

À partir d'une variable aléatoire réelle X quelconque, on obtient une variable centrée en prenant :

On se servira souvent de la linéarité pour aller vite dans nos calculs d'espérance.

Exemples :

a) On lance deux fois de suite un dé à 6 faces et on note S la somme des résultats obtenus. Déterminer $E(S)$.

b) Retrouver plus rapidement l'espérance d'une variable binomiale grâce à la linéarité.

c) Une urne contient N_r boules rouges numérotées, et N_b boules blanches numérotées. On tire simultanément n boules, avec $1 \leq n \leq N_r + N_b$.
Déterminer l'espérance du nombre X de boules rouges tirées.

Démonstration 4

2 Variance, écart-type et covariance

2.a Variance et écart-type

Définition :

Soit X une variable aléatoire réelle sur Ω.
On appelle variance de X le nombre :

$$V(X) =$$

C'est un réel positif.
On appelle écart-type de X le nombre $\sigma(X) =$
C'est un réel positif.

L'espérance était un *indicateur de position* : cela donne une "valeur moyenne" de X en prenant en compte les différentes probabilités d'apparition des différentes valeurs de X.
La variance est un *indicateur de dispersion* : cela mesure à quel point la variable X est susceptible de s'éloigner de sa valeur moyenne $E(X)$... Intuitivement :

- Si $V(X)$ est petite, X est « le plus souvent » proche de $E(X)$.
- Si $V(X)$ est grande, X est prend « souvent » des valeurs éloignées de $E(X)$.

On s'intéresse parfois à l'écart-type pour des questions d'homogénéité : l'intérêt de l'écart-type par rapport à la variance est qu'il s'exprime dans la même unité que X (s'il y en a une).

Exemple : Soient X et Y de lois :

x	5	15
$P(X=x)$	$\frac{1}{2}$	$\frac{1}{2}$

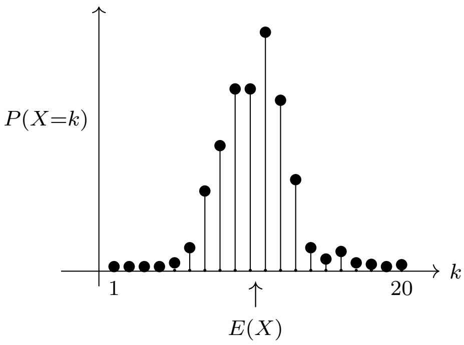
et

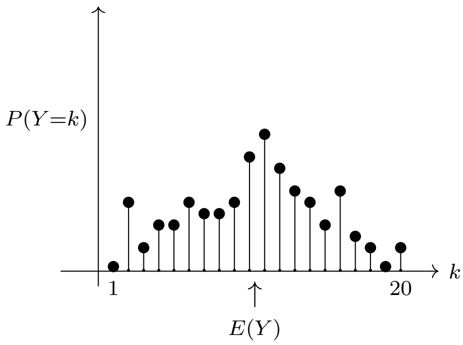
y	9	11
$P(Y=y)$	$\frac{1}{2}$	$\frac{1}{2}$

.

On constate facilement que X et Y ont même espérance, égale à 10.
Calculons leurs variances respectives :

Illustration : Ici les deux variables aléatoires X et Y sont à valeurs dans $\{1, \dots, 20\}$, elles ont la même espérance (environ 10), mais X est d'écart-type 2, Y est d'écart-type 5 :





2.b Propriétés de la variance

Proposition :

(Formule de Huygens) Soit X une variable aléatoire réelle sur Ω.

Démonstration 5

Cette "forme développée" de la variance est très souvent utile : la plupart du temps, on aura déjà calculé $E(X)$, donc pour calculer $V(X)$, il suffira de calculer $E(X^2)$.
De façon générale, comme il s'agit d'une formule à 3 termes, dès qu'on a 2 des 3 termes, on en déduit le dernier.

Proposition :

Pour a et b des réels et X une variable aléatoire : $V(aX + b) =$

Démonstration 6

Définition :

Une variable aléatoire réelle X sur Ω est dite centrée réduite si $E(X) = 0$ et si $V(X) = 1$.

À partir d'une variable aléatoire réelle X quelconque de variance non nulle, on obtient une variable centrée réduite en prenant

2.c Les exemples à connaître par cœur

Proposition :

- Si X est une variable de Bernoulli de paramètre p, alors
- Si X est une variable binomiale de paramètres n et p, alors

Démonstration 7

2.d Covariance

Définition :

Soient X et Y des variables aléatoires réelles sur Ω.
On appelle covariance de X et Y le nombre :

$$\text{cov}(X, Y) =$$

On a les propriétés immédiates suivantes :
- la covariance est symétrique : $\text{cov}(X, Y) = \text{cov}(X, Y)$
- La variance de X est la covariance de X avec elle-même : $V(X) = \text{cov}(X, X)$.

Comme pour la variance, on a une forme "développée" de la covariance :
Proposition :

Soient X et Y des variables aléatoires réelles sur Ω.

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

Démonstration 8

Lorsque $\text{cov}(X, Y) = 0$, on dit que les variables sont décorrélées ; cela équivaut à $E(XY) = E(X)E(Y)$.
D'après la partie 1 :
Proposition :

Si X et Y sont des variables aléatoires réelles indépendantes, alors $\text{cov}(X, Y) = 0$.

Ainsi X et Y indépendantes $\Longrightarrow$ X et Y décorrélées, mais la réciproque est fausse !

2.e Variance d'une somme

La covariance permet de calculer la variance d'une somme :
Proposition :

Soient X et Y des variables aléatoires réelles sur Ω.

$$V(X + Y) = V(X) + 2\text{cov}(X, Y) + V(Y)$$

Démonstration 9

On notera l'analogie avec l'identité remarquable $(a + b)^2 = a^2 + 2ab + b^2$.

Cette formule se généralise à la somme de n variables aléatoires réelles $X_1, \dots, X_n$:

$$V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{1 \leq i < j \leq n} \operatorname{cov}(X_i, X_j)$$

En particulier, si les variables aléatoires en jeu sont indépendantes, toutes les covariances qui appa-raissent sont nulles donc :

Proposition :

- Si X et Y sont des variables aléatoires réelles $\boxed{\text{indépendantes}}$ sur Ω, alors

$$V(X+Y) = V(X) + V(Y)$$

- Si $X_1, \dots, X_n$ sont des variables aléatoires réelles $\boxed{\text{indépendantes}}$ sur Ω, alors

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

(Re)trouvons la variance d'une variable binomiale à l'aide de ce résultat :

Démonstration 10

3 Inégalités probabilistes

3.a Inégalité de Markov

Théorème :

Soit X une variable aléatoire réelle <u>positive</u> sur Ω.

$$\forall\, a > 0, \quad P\big(X \geq a\big) \leq \frac{E(X)}{a}$$

Démonstration 11

Si X est une variable aléatoire pas forcément positive, on peut appliquer le résultat à $|X|$:
$\forall\, a > 0, \quad P\big(|X| \geq a\big) \leq \dfrac{E(|X|)}{a}$.

3.b Inégalité de Bienaymé-Tchebychev

Théorème :

Soit X une variable aléatoire réelle sur Ω.

$$\forall\, a > 0, \quad P\big(|X - E(X)| \geq a\big) \leq \frac{V(X)}{a^2}$$

Démonstration 12

Interprétation :

Cette inégalité confirme l'intérêt de la variance comme mesure de dispersion par rapport à la moyenne.

Exemple : Au second tour d'une élection, on a deux candidats A et B. Notons p la proportion de la population qui souhaite voter pour A.
Pour un sondage, on veut interroger un grand nombre n de personnes, de sorte qu'en notant X_n le nombre de personnes interrogées souhaitant voter pour A, la quantité $\dfrac{X_n}{n}$ soit proche de p.
On souhaite que $Y = \dfrac{X_n}{n}$ s'écarte de p de moins de 1%.
Déterminer n pour cela arrive avec une probabilité supérieure à 95%.

✎ **Démonstration** 13

Plus généralement, les inégalités de Markov et de Bienaymé-Tchebychev sont des *inégalités de concentration* ; elles donnent une borne sur la probabilité que X "dévie" d'une certaine valeur, l'espérance pour l'inégalité de Bienaymé-Tchebychev.

Cela permet de justifier que les objets qu'on a construits dans ce cours de probabilité ont une interprétation intuitive en termes de fréquence !
Par exemple, si on s'intéresse à une grandeur que l'on obtient au cours d'une expérience aléatoire donnée, on peut effectuer n fois cette expérience aléatoire, et noter X_i la valeur obtenue au cours du $i$ème essai. Les variables X_i sont donc indépendantes et de même loi ; notons μ leur espérance et σ leur écart-type, de sorte que la variance soit σ^2.
La valeur moyenne obtenue au cours des n essais (moyenne "empirique") est donc $S_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
Son espérance vaut :

$$E(S_n) =$$

Sa variance vaut :

$$V(S_n) =$$

Donc, d'après l'inégalité de Bienaymé-Tchebychev, pour tout $a > 0$,

$$P\big(|S_n - \mu| \geq a\big) \leq \frac{1}{n}\frac{\sigma^2}{a^2}$$

Grâce au facteur $\dfrac{1}{n}$, cela tend donc vers 0 lorsque n tend vers $+\infty$!
Ainsi, plus on fait d'essais, moins cette moyenne S_n a de chances de s'éloigner de l'espérance μ.
Plus précisément, la probabilité que la valeur de S_n soit éloignée (en valeur absolue) d'un petit $a > 0$ de l'espérance tend vers 0, même en prenant a "petit" !

Plan du cours