Lycée Pierre-Gilles de Gennes **BCPST**
Mathématiques **2023-2024**

Notes de cours 02 Bis

Probabilités : l'espérance en version abstraite

Table des matières

⋆⋆⋆

On considère un espace probabilisé général $(\Omega, \mathscr{T}, \mathbb{P})$. Toutes les variables aléatoires apparaissant sont définies sur l'univers Ω.

Présentation heuristique

L'espérance d'une variable aléatoire numérique X doit être comprise comme la moyenne des valeurs prises par X lorsque que l'on moyenne sur toutes les configurations possibles du système.

Le mot espérance provient de la théorie des jeux : si dans un jeu (par exemple pile ou face), le succès A rapporte une somme donnée alors que l'échec non(A) ne rapporte rien, l'*espérance de gain* pour ce jeu est

$$\boxed{\text{Montant à gagner} \times \text{ probabilité de gagner}}$$

En formules, si A est un événement, on note $1\!\!1_A$ la variable aléatoire valant 1 si l'événement A survient, 0 sinon. Elle représente le gain du joueur si le montant à gagner est de 1. On a

$$\mathbb{E}(1\!\!1_A) = \mathbb{P}(A)$$

Si le montant à gagner en cas de succès est s, le gain du joueur est donné par la v.a.r $s.1\!\!1_A$, son espérance est

$$\mathbb{E}(s.1\!\!1_A) = s.\mathbb{P}(A) = s.\mathbb{E}(1\!\!1_A)$$

Si un jeu a plus de deux issues possibles, marquées par la famille complète d'événements incompatibles $A_1, A_2, A_3, \ldots$ avec une somme s_k à gagner en cas de survenue de l'événement A_k alors l'espérance de gain est

$$\boxed{\text{Montant } s_1 \times \text{ probabilité de } A_1 + \text{Montant } s_2 \times \text{ probabilité de } A_2 + ..}$$

1 Variables aléatoires réelles, espérance

1.1 Espérance abstraite

On admet qu'il existe une opération $\mathbb{E}$, agissant dans un premier temps sur les variables aléatoires réelles *positives*, dont les valeurs sont soit un nombre réel, soit le symbole ∞, et telle que

1. $\mathbb{E}(1\!\!1) = 1$, $\mathbb{E}(0) = 0$
2. Si $A \in \mathscr{T}$, $\mathbb{E}(1\!\!1_A) = \mathbb{P}(A)$,
3. Si X est une variable aléatoire réelle prenant des valeurs positives alors $\mathbb{E}(X)$ a toujours un sens. [1] Il s'agit soit d'un nombre réel positif, soit du symbole « infini » (∞).
4. Dans ce cas, on a l'équivalence $\mathbb{E}(X) = 0$ si et seulement si $\mathbb{P}(X = 0) = 1$.
5. (Croissance de l'espérance) Si X et Y sont deux variables aléatoires réelles prenant des valeurs positives et si $X \leq Y$, alors [2]
$$0 \leq \mathbb{E}(X) \leq \mathbb{E}(Y)$$

1. (Additivité de l'espérance) Si X et Y sont deux variables aléatoires réelles prenant des valeurs positives, alors, avec l'extension naturelle de cette écriture au cas où le symbole $+\infty$ apparaît,
$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$
2. (Homogénéité de l'espérance) Si X est une variable aléatoire réelle prenant des valeurs positives et $\lambda \in [0, +\infty[$, alors, avec conventions particulières pour gérer le cas $\mathbb{E}(X) = \infty$, on a
$$\mathbb{E}(\lambda.X) = \lambda.\mathbb{E}(X)$$

1. Ce qui est à comprendre c'est que $\mathbb{E}$ est un genre de moyenne. On moyenne sur tous les « ω »
2. $\mathbb{E}$ se comporte à beaucoup d'égards (linéarité, positivité) comme le symbole $\sum$ ou le symbole $\int$, ce qui va ressortir lors de diverses formules de transfert.

Variables aléatoires réelles admettant une espérance

Définition 1

Soit X une variable aléatoire réelle.

1. On dit que X est *intégrable* ou que X *admet une espérance* si $\mathbb{E}(|X|) \neq \infty$.
2. Si X est intégrable, $X^+ = \max(X, 0)$ et $X^- = -\min(X, 0)$ le sont aussi, sont à valeurs positives, on a
$$X = X^+ - X^-, \ |X| = X^+ + X^-,$$
et on pose
$$\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

---

1. La formule de CAVALIERI
$$\mathbb{E}(X) = \int_0^{+\infty} \mathbb{P}(X > t)\, dt$$
vue en exercice dans le chapitre précédent peut servir de définition pour le cas le plus général. En particulier, si X prend un nombre fini de valeurs réelles positives, $\mathbb{E}(X)$ désigne l'espérance de X au sens de la première année.
2. avec l'extension naturelle de cette écriture au cas où l'un des deux vaut ∞

RQ : ceci n'est pas sans rappeler des éléments de démonstration du théorème $ACV \Rightarrow CV$. C'est exprès, comme on le verra dans le chapitre sur les v.a. à densité. On a les règles de calcul suivantes.

Proposition 2

Soit X, Y des variables aléatoires réelles *admettant une espérance* et λ, μ deux nombres réels,

1. (Linéarité de l'espérance) $Z = \lambda.X + \mu.Y$ est une variable aléatoire réelle *admettant une espérance* et
$$\mathbb{E}(Z) = \mathbb{E}(\lambda.X + \mu.Y) = \lambda.\mathbb{E}(X) + \mu.\mathbb{E}(Y)$$
2. (Croissance) Si $X \le Y$ alors $\mathbb{E}(X) \le \mathbb{E}(Y)$ avec égalité ssi $\mathbb{P}(X = Y) = 1$.
3. (Inégalité triangulaire)
$$|\mathbb{E}(\lambda.X + \mu.Y)| \le |\lambda|.\mathbb{E}(|X|) + |\mu|.\mathbb{E}(|Y|)$$

1.2 Des exemples.

Nombre fini de valeurs.

Dans le cas d'une v.a.r. X prenant un nombre fini de valeurs, on retrouve, par linéarité de $\mathbb{E}$, la formule de première année

Si X est une v.a.r., à valeurs dans l'ensemble fini $\{x_k, k \in \{1, \dots, K\}\} \subset \mathbb{R}$, alors

1. $X = \sum_{k=1}^{K} x_k 1\!\!1_{\{X=x_k\}}$
2. X admet une espérance,
$$\mathbb{E}(|X|) = \sum_{k=1}^{K} |x_k| \mathbb{P}(X = x_k) \text{ et } \mathbb{E}(X) = \sum_{k=1}^{K} x_k \mathbb{P}(X = x_k)$$

Exercice 1.— Soit n un entier naturel non nul, soit X une variable suivant la loi $\mathscr{U}_{\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, \frac{n}{n}\}}$. Soit $f : [0,1] \to \mathbb{R}$ une fonction continue.
1. Donner une formule pour $S_n(f) = \mathbb{E}(f(X))$. Appliquer cette formule à l'exemple $f : x \mapsto x^2$.
2. Quelle est, en général, la limite de $S_n(f)$ lorsque $n \to +\infty$.
3. Quelle est la valeur de cette limite pour l'exemple proposé ?

L'espérance d'une fonction d'une variable uniforme

On admet que l'on a la formule de transfert suivante :

Théorème 3: Formule de transfert générique, v.a. uniforme.

Supposons que $U \sim \mathscr{U}_{[0,1]}$. Si $h : \mathbb{R} \to \mathbb{R}$ est une fonction, on a
$$\mathbb{E}(h(U)) = \int_{-\infty}^{+\infty} h(u) 1\!\!1_{\{u \in [0,1]\}}\, du = \int_0^1 h(u)\, du$$
pourvu que cette dernière intégrale soit absolument convergente.
On convient que $\mathbb{E}(h(U)) = +\infty$ si h est positive et l'intégrale diverge vers $+\infty$.

Exemples :

1. Montrer que le formule de transfert en question est vraie au cas où h est constante par morceaux, *i.e.* h est de la forme $h = \sum_{k=1}^{K} x_k 1\!\!1_{I_k}$ où les I_k sont des intervalles disjoints de $\mathbb{R}$. (On pourra vérifier la formule dans le cas d'un seul intervalle et argumenter par linéarité des opérations).
2. Calculer $\mathbb{E}(U^2)$, $E(\frac{1}{1+U})$, $\mathbb{E}(\frac{1}{\sqrt{1-U^2}})$.
3. Calculer $\mathbb{E}(e^{-\lambda.U})$ où $\lambda \in \mathbb{R}$.
4. Les v.a.r $\ln U$, $\frac{1}{U}$, $\frac{1}{\sqrt{U}}$ sont-elles intégrables ? *i.e.* admettent-elles une espérance ?
5. Soit V une v.a.r uniforme sur $[0,1]$. Donner l'espérance de $U+V$, de $U-V$.
6. Soit $a < b$, $a, b \in \mathbb{R}$, V une v.a.r uniforme sur $[a,b]$. Donner la formule de transfert générique donnant $\mathbb{E}(h(V))$.

1.3 Inégalité de MARKOV

Théorème 4: Inégalité de MARKOV

Si X est une variable aléatoire réelle, *positive*, alors, pour tout seuil $\lambda > 0$,

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}$$

Démonstration. On se limite évidemment au cas où X admet une espérance. Comme X est positive, on a l'inégalité entre v.a.r :

$$1\!\!1_{\{X \geq \lambda\}} \leq \frac{X}{\lambda}.$$

Par croissance et linéarité de l'espérance, il vient

$$\mathbb{P}(X \geq \lambda) = \mathbb{E}(1\!\!1_{\{X \geq \lambda\}}) \leq \frac{\mathbb{E}(X)}{\lambda}.$$

□

2 Variables admettant une variance

2.1 Définition

Définition 5

Soit X une v.a.r. On dit que X est de carré intégrable [a] si X^2 est intégrable, *i.e.* admet une espérance.

a. on dira admet une variance à l'issue de ce chapitre

Exercice 2.— Soit $U \sim \mathscr{U}_{[0,1]}$.
1. U est de carré intégrable ?
2. $\frac{1}{\sqrt{U}}$ est intégrable sans être de carré intégrable ?

2.2 Inégalité de CAUCHY–SCHWARZ

Théorème 6: Inégalité de CAUCHY–SCHWARZ

Si X et Y sont deux v.a.r de carré intégrable alors $X.Y$ admet une espérance et on a

$$|\mathbb{E}(X.Y)| \leq \mathbb{E}(|X|.|Y|) \leq \mathbb{E}(|X|^2)^{\frac{1}{2}}\mathbb{E}(|Y|^2)^{\frac{1}{2}}$$

Démonstration. On commence par noter que si X et Y sont de carré intégrable alors toute combinaison linéaire de X et Y l'est aussi. En effet, concernant la somme,

$$0 \leq (X+Y)^2 \leq 4X^2 + 4Y^2$$

et comme par hypothèse, $\mathbb{E}(X^2)$ et $\mathbb{E}(Y^2)$ sont finis, 4 fois leur somme l'est aussi et, par croissance de l'espérance, $\mathbb{E}((X+Y)^2$ est finie.

Si λ, μ deux nombres réels, alors $\lambda.X$ et $\mu.Y$ sont aussi de carré intégrable et leur somme aussi.

Comme $X.Y = \frac{1}{4}((X+Y)^2 - (X-Y)^2)$, on en déduit que $X.Y$ admet une espérance.

Considérons maintenant la fonction $f : \lambda \in \mathbb{R} \mapsto \mathbb{E}((X+\lambda.Y)^2)$. Cette fonction est bien définie, à valeurs *réelles*.

On a, en développant le carré et en utilisant la linéarité de l'espérance d'une part et en utilisant la positivité de l'espérance, que, pour *tout* $\lambda \in \mathbb{R}$,

$$0 \leq \mathbb{E}((X+\lambda.Y)^2) = \mathbb{E}(X^2) + 2\lambda\mathbb{E}(XY) + \lambda^2\mathbb{E}(Y^2)$$

La fonction f est donc une fonction polynômiale, de degré au plus 2, positive.

1. Si $\mathbb{E}(Y^2) \neq 0$, f est de degré 2 et son discrimant (réduit) est *négatif* :

$$(\mathbb{E}(XY))^2 - \mathbb{E}(X^2).\mathbb{E}(Y^2) \leq 0,$$

   ce qui, une fois réarrangé donne l'inégalité attendue.

2. Si $\mathbb{E}(Y^2) = 0$, on a affaire à une fonction affine positive et donc sa pente est nulle : $\mathbb{E}(XY) = 0$ et l'inégalité est aussi vérifiée.

L'inégalité intermédiaire avec $\mathbb{E}(|X|.|Y|)$ provient de l'application de l'inégalité démontrée dans le cas positif aux variables $|X|$ et $|Y|$ qui sont elles aussi de carré intégrable. $\square$

Une conséquence en est la suivante :

Proposition 7. *Si X est de carré intégrable, elle admet une espérance et*

$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq \mathbb{E}(X^2)^{\frac{1}{2}}$$

Démonstration. On applique le théorème précédent à X et $Y = \mathbb{1}$. Y est de carré intégrable et $\mathbb{E}(Y^2) = \mathbb{E}(Y) = 1$. $\square$

2.3 Variance et covariance

Proposition–Définition 8

1. Si X est de carré intégrable[a], sa variance est définie par

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

2. Son *écart-type*, $\sigma(X) = \mathbb{V}(X)^{\frac{1}{2}}$.
3. Si X et Y sont de carré intégrable, leur covariance est définie par

$$\mathbb{C}ov(X,Y) = \mathbb{E}((X - \mathbb{E}(X)).(Y - \mathbb{E}(Y)))$$

4. On a $|\mathbb{C}ov(X,Y)| \leq \sigma(X).\sigma(Y)$.

a. La terminologie officielle est « X admet une variance ».

Formule de KOENIG–HUYGENS

Proposition–Définition 9: Formule de KOENIG–HUYGENS

1. Si X admet une variance, sa variance vaut

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

2. Si X et Y admettent une variance, leur covariance vaut

$$\mathbb{C}ov(X,Y) = \mathbb{E}(X.Y) - \mathbb{E}(X).\mathbb{E}(Y)$$

Démonstration. On ne démontre que la première égalité, la seconde relevant de la même technique de calcul. On a, en développant le carré, par linéarité de l'espérance et le fait que $\mathbb{E}(C) = C$ pour une constante C,

$$\mathbb{V}(X) = \mathbb{E}(X^2 - 2X.\mathbb{E}(X) + \mathbb{E}(X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X).\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

□

Centrage, normalisation

Une variable aléatoire *positive* est d'espérance nulle si et seulement si $\mathbb{P}(X = 0) = 1$. (On dit que X est nulle *presque sûrement*.)

Une variable aléatoire est de variance nulle, $\mathbb{V}(X) = 0$, si et seulement si $\mathbb{P}(X = \mathbb{E}(X)) = 1$. Autrement dit ssi X est une constante presque sûrement.

Supposons que X est une variable aléatoire admettant une variance.

— si C une constante (*i.e.* une variable ne prenant presque sûrement qu'une valeur), on a

$$\mathbb{V}(X + C) = \mathbb{V}(X)$$

— si λ est une constante, $\mathbb{V}(\lambda.X) = \lambda^2\mathbb{V}(X)$.

On peut se servir de ces deux propriétés pour centrer, puis normaliser une v.a. non constante X, admettant une variance.

$$X^* = \frac{X - \mathbb{E}(X)}{\sigma(X)}$$

est une variable d'espérance nulle (centrée), de variance 1. On dit qu'elle est *centrée-réduite*.

Remarquer que si X est mesurée dans une certaine unité physique, alors $\mathbb{E}(X)$, $\sigma(X)$ sont dans la même unité et X^* est sans unité.

Exercice 3.—
1. Donner espérance et variance d'une v.a $U \sim \mathscr{U}_{[0,1]}$.
2. Donner espérance et variance d'une v.a $V \sim \mathscr{U}_{[a,b]}$.
3. Quelle est la centrée réduite de U ? de V ?

Exercice 4.— Si $X \sim \mathscr{B}\left(n, \frac{1}{2}\right)$. Donner l'histogramme de X^*. Dans quel ensemble fini X^* prend-elle ses valeurs ?

Exercice 5.— Même question si $X \sim \mathscr{U}_{\{0,\ldots,2n\}}$ où $n \in \mathbb{N}^*$.

La règle de la somme

Proposition 10

Si X et Y sont de carré intégrable alors

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathbb{C}ov(X,Y)$$

Exercice 6.— Donner une formule pour $\mathbb{V}(\sum_{n=1}^{N} X_n)$ où les X_n sont de carré intégrable.

2.4 Inégalité de Bienaymé–Tchebycheff

On a vu que l'espérance d'une v.a.r est une sorte de moyenne des valeurs que peut prendre X. Les valeurs de X se répartissent donc de part et d'autre $\mathbb{E}(X)$.

La valeur de la variance mesure quant à elle, assez grossièrement, la *façon* dont les valeurs prises par X se répartissent autour de $\mathbb{E}(X)$.

L'inégalité de Bienaymé–Tchebycheff est une traduction de l'idée de répartition de X autour de son espérance.

Théorème 11: Inégalité de Bienaymé–Tchebycheff

Si X est de carré intégrable, on a, pour tout seuil $\lambda > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda) \leq \frac{\mathbb{V}(X)}{\lambda^2}$$

Démonstration. Il suffit d'appliquer l'inégalité de Markov à la variable positive $Z = (X - \mathbb{E}(X))^2$ et de remarquer que les événements $(|X - \mathbb{E}(X)| > \lambda)$ et $(Z > \lambda^2)$ sont égaux. On a $\mathbb{E}(Z) = \mathbb{V}(X)$. □

Théorème 12 (Inégalité de BIENAYMÉ–TCHEBYCHEFF)**.** *Si X admet une variance, X non constante, on a, pour tout seuil $\lambda > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| > \lambda.\sigma(X)) \leq \frac{1}{\lambda^2}$$

Notamment, si $\lambda = 2$, on obtient qu'avec probabilité supérieure à 75%, X est distant de la constante $\mathbb{E}(X)$ de moins de 2 écart-types. **Ceci, quelle que soit la distribution de X.**

Cette inégalité n'est pertinente que pour $\lambda > 1$.

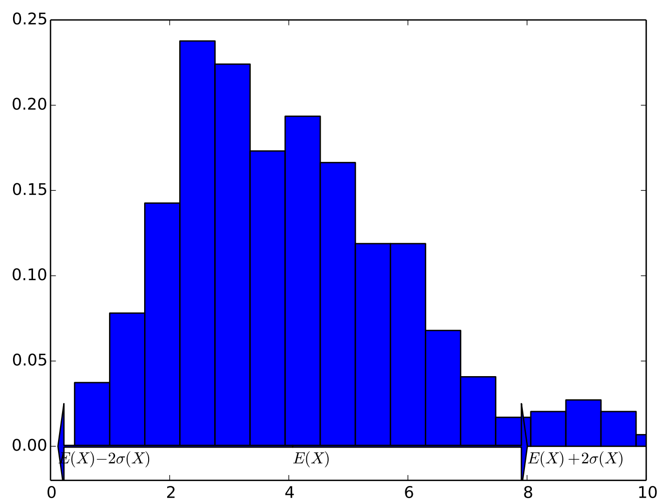

FIGURE 1 – BIENAYMÉ–TCHEBYCHEFF : Au moins 75% de la population est à moins de 2 écarts-types de la moyenne.

Exercice 7.— Sur $n = 100$ palourdes péchêes en Bretagne sud, on effectue la mesure de leur diamètre. On note d_k le diamètre de la palourde portant le numéro $k \in \{1, \dots, n\}$. On regroupe ces mesures en un tableau et on calcule (via Excel par exemple) leur moyenne $\overline{d} = \frac{1}{n}\sum_{k=1}^{n} d_k$ et l'écart-type $\sigma = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(d_k - \overline{d})^2}$.

1. On prend une palourde au hasard et on note D son diamètre. Calculer $\mathbb{E}(D)$ et $\mathbb{V}(D)$ en fonction de $\overline{d}$ et σ.

2.a. Soit $\alpha > 0$. Enoncer l'inégalité de BIENAYMÉ–TCHEBYCHEFF pour D au seuil $\alpha\sigma$.

2.b. Comment choisir α pour être sûr que moins d'un quart des palourdes ont un diamètre qui diffère du diamètre moyen de plus de $\alpha.\sigma$ en valeur absolue.

3. En fait, on met nos n palourdes dans un sac que l'on secoue, ce qui fait que les plus grandes palourdes vont plus facilement être tirées. On admet que la palourde numéro k est tirée avec probabilité

$$p_k = \frac{d_k}{n.\overline{d}}$$

3.a. Vérifier que $\sum_{k=1}^{n} p_k = 1$.

3.b. On note D' le diamètre de la palourde tirée au sort suivant ce mode. Pourquoi, heuristiquement, $\mathbb{E}(D') \geq \mathbb{E}(D)$?

3.c. Donner une formule pour $\mathbb{E}(D')$ en fonction de $\mathbb{E}(D^2)$.

3.d. Démontrer mathématiquement l'inégalité heuristique.

2.5 Couples de variables, coeff. de corrélation et régression linéaire

Le problème de la régression linéaire est le suivant. On dispose de deux variables X (non constante) et Y, admettant ue variance. Faire la régression linéaire de Y sur X consiste à trouver la variable $\hat{Y}$, fonction affine de X « la plus proche » de Y *au sens des moindres carrés*, *i.e.* trouver deux réels a et b tels que

$$\mathbb{E}((Y-(aX+b))^2)$$

soit minimal. On pose alors $\hat{Y} = aX+b$.

Théorème 13: Droite de régression

La régression linéaire de Y sur X est la v.a.

$$\hat{Y} = \frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)}(X-\mathbb{E}(X))+\mathbb{E}(Y)$$

ou alors, en variables centrées et réduites, si X et Y sont non constantes,

$$\hat{Y^*} = \rho(X,Y)X^*$$

où $\rho(X,Y) = \frac{\mathbb{C}ov(X,Y)}{\sigma(X)\sigma(Y)}$ est le *coefficient de corrélation linéaire*.

Démonstration. Posons, pour $a,b \in \mathbb{R}$, $f(a,b) = \mathbb{E}((Y-(aX+b))^2)$ et développons cette expression. On a

$$f(a,b) = \mathbb{E}(Y^2)+a^2\mathbb{E}(X^2)+b^2-2a\mathbb{E}(X.Y)-2b\mathbb{E}(Y)+2ab\mathbb{E}(X)$$

Ce serait plus agréable si les variables X et Y étaient centrées, il y aurait deux termes de moins. Reprenons l'expression de départ et forçons le centrage.

$$\begin{aligned} f(a,b) &= \mathbb{E}((\underbrace{Y-\mathbb{E}(Y)}_{=\tilde{Y}}-(\underbrace{a}_{=\alpha}\underbrace{(X-\mathbb{E}(X)}_{=\tilde{X}}+\underbrace{b-\mathbb{E}(Y)+a\mathbb{E}(X)}_{=\beta}))^2) = \mathbb{E}((\tilde{Y}-(\alpha\tilde{X}+\beta))^2) \\ &= g(\alpha,\beta) \end{aligned}$$

où on a posé $\alpha = a$, $\beta = b-\mathbb{E}(Y)+a\mathbb{E}(X)$. Trouver un couple (a,b) qui rende $f(a,b)$ minimal revient à trouver un couple (α,β) qui rende $g(\alpha,\beta)$ minimal.

Développons l'expression de $g(\alpha,\beta)$, comme nous avons développé $f(a,b)$, le calcul est le même :

$$\begin{aligned} g(\alpha,\beta) &= \underbrace{\mathbb{E}(\tilde{Y}^2)}_{=\mathbb{V}(Y)}+a^2\underbrace{\mathbb{E}(\tilde{X}^2)}_{=\mathbb{V}(X)}+b^2-2a\underbrace{\mathbb{E}(\tilde{X}.\tilde{Y})}_{=\mathbb{C}ov(X,Y)}-2b\underbrace{\mathbb{E}(\tilde{Y})}_{=0}+2ab\underbrace{\mathbb{E}(\tilde{X})}_{=0} \\ &= \mathbb{V}(Y)+\beta^2+\alpha^2\mathbb{V}(X)-2\alpha\,\mathbb{C}ov(X,Y) \\ \text{(forme canonique)} &= \underbrace{\beta^2}_{\geq 0}+\underbrace{\mathbb{V}(X)(\alpha-\frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)})^2}_{\geq 0}+\frac{\mathbb{V}(Y)\mathbb{V}(X)-\mathbb{C}ov(X,Y)^2}{\mathbb{V}(X)} \\ &\geq g(\frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)},0) \end{aligned}$$

Ceci démontre que le minimum de $g(\alpha,\beta)$ est (uniquement) atteint en $(\frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)},0)$ et donc que le minimum de f est atteint en

$$a=\frac{\mathbb{C}ov(X,Y)}{\mathbb{V}(X)}, b=\mathbb{E}(Y)-a.\mathbb{E}(X)$$

Ce qui est ce que l'on cherche. On peut noter que le minimum vaut

$$\frac{\mathbb{V}(Y)\mathbb{V}(X)-\mathbb{C}ov(X,Y)^2}{\mathbb{V}(X)}$$

Cette quantité est positive, par l'inégalité de CAUCHY–SCHWARZ et ne vaut 0 qu'en cas d'égalité dans cette inégalité, *i.e.* lorsque Y est fonction affine de X. □

Pour retrouver la formule de $\hat{Y}$, on cherche a et b tels que $\mathbb{C}ov(X,\hat{Y})=\mathbb{C}ov(X,Y)$ et $\mathbb{E}(\hat{Y})=\mathbb{E}(Y)$.

1. Si $\mathbb{C}ov(X,Y)=0$ (*i.e.* $\rho(X,Y)=0$), on dit que X et Y sont *non corrélées*.
2. Si $\rho(X,Y)>0$, on dit que X et Y sont *positivement corrélées*.
3. Si $\rho(X,Y)<0$, on dit que X et Y sont *négativement corrélées*.

Exercice 8.— Soit $X\sim\mathscr{U}_{[0,1]}, Y=X^2$.
1. Donner les valeurs de $\mathbb{V}(X)$, $\mathbb{V}(Y)$, $\mathbb{C}ov(X,Y)$.
2. Donner une formule pour $\hat{Y}$ la régression linéaire de Y sur X.
3. En Python, tirer 100 nombres x_i suivant la loi de X de façon indépendante. Tracer les nuages de points (x_i,y_i) le nuage $(x_i,\hat{y}_i)$ et tracer la droite des moindres carrés calculée statistiquement.

Proposition 14. *Si X et Y sont non corrélées, on a*

$$\mathbb{V}(X+Y)=\mathbb{V}(X)+\mathbb{V}(Y)$$

Si $X_1,\dots,X_n$ sont deux à deux non corrélées, i.e. $\forall i,j\in\{1,\dots,n\}, i\neq j\Rightarrow\mathbb{C}ov(X_i,X_j)=0$ alors

$$\mathbb{V}(X_1+\cdots+X_n)=\mathbb{V}(X_1)+\cdots+\mathbb{V}(X_n)$$

3 Indépendance et espérance.

Indépendance d'événements

Définition 15. *Soit $(\Omega,\mathscr{T},\mathbb{P})$ un espace probabilisé*

1. *Soit $A,B\in\mathscr{T}$ deux événements. On dit que A et B sont indépendants si $\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B)$.*
2. *Soit $A_1,\dots,A_n\in\mathscr{T}$ une famille d'événements. On dit qu'ils sont (mutuellement) indépendants si pour toute sous-famille $A_{i_1},\dots,A_{i_k}$, $i_1<i_2<\cdots<i_k$,*

$$\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k})=\mathbb{P}(A_{i_1})\cap\dots\mathbb{P}(A_{i_k})$$

3. *Soit $A_1,\dots,A_n,\dots\in\mathscr{T}$ une famille (infinie) d'événements. On dit qu'ils sont (mutuellement) indépendants si toute sous-famille finie est indépendante.*

Exercice 9.— Montrer que si A et B sont indépendants, A et $\overline{B}$ le sont aussi.

Exercice 10.— Montrer que si A est indépendant de lui-même alors $\mathbb{P}(A)=1$ ou $\mathbb{P}(A)=0$.

V.a. indépendantes : deux définitions équivalentes

Définition 16 (Première définition)**.** *1. Deux v.a.r X et Y sont* indépendantes *si pour tous intervalles I, J de $\mathbb{R}$,*

$$\mathbb{P}(X \in I \text{ et } Y \in J) = \mathbb{P}(X \in I).\mathbb{P}(Y \in J)$$

2. De façon analogue, les v.a.r $X_1, \ldots, X_n$ sont dites (mutuellement) indépendantes *si pour tous intervalles $I_1, \ldots, I_n$,*

$$\mathbb{P}(X_1 \in I_1 \text{ et } \ldots X_n \in I_n) = \mathbb{P}(X_1 \in I_1) \ldots \mathbb{P}(X_n \in I_n)$$

Traduisons ceci en termes d'espérance en utilisant la relation fondamentale $\mathbb{P}(A) = \mathbb{E}(1\!\!1_A)$.
X et Y sont indépendantes si pour tous intervalles I et J,

$$\mathbb{E}(1\!\!1_{\{X\in I\}}.1\!\!1_{\{Y\in J\}}) = \mathbb{E}(1\!\!1_{\{X\in I\}}).\mathbb{E}(1\!\!1_{\{X\in J\}})$$

Prenons maintenant une fonction f de la forme $f = \sum_{k=1}^{K} f_k 1\!\!1_{I_k}$ et $g = \sum_{\ell=1}^{L} g_\ell 1\!\!1_{J_\ell}$ où $(I_k)_{1\leq k\leq K}$ et $(J_\ell)_{1\leq \ell\leq L}$ sont deux familles finies d'intervalles de $\mathbb{R}$. Un peu d'algèbre (on pourra tester avec des familles de deux intervalles) utilisant uniquement la linéarité de l'espérance montre que

$$\mathbb{E}(f(X).g(Y)) = \mathbb{E}(f(X)).\mathbb{E}(g(Y))$$

On a bien évidemment un énoncé similaire pour une famille $X_1, \ldots, X_n$ de variables indépendantes en prenant des fonctions $f_1, \ldots, f_n$ en escalier sur $\mathbb{R}$:

$$\mathbb{E}(\prod_{\nu=1}^{n} f_\nu(X_\nu)) = \prod_{\nu=1}^{n} \mathbb{E}(f_\nu(X_\nu))$$

Définition 17: Deuxième définition équivalente

1. Deux v.a.r X et Y sont *indépendantes* si pour toutes fonctions $f, g : \mathbb{R} \to \mathbb{R}$

$$\mathbb{E}(f(X).g(Y)) = \mathbb{E}(f(X)).\mathbb{E}(g(Y))$$

pourvu que ces espérances soient bien définies.

2. De façon analogue, les v.a.r $X_1, \ldots, X_n$ sont dites *(mutuellement) indépendantes* si pour toutes fonctions $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$, pourvu que les termes de cette formule soient définis,

$$\mathbb{E}(\prod_{\nu=1}^{n} f_\nu(X_\nu)) = \prod_{\nu=1}^{n} \mathbb{E}(f_\nu(X_\nu))$$

Lemmes d'indépendance

Lemme 18 (des coalitions)**.** *Si $X_1, \ldots, X_n, Y_1, \ldots, Y_p$ sont indépendantes et si $u : \mathbb{R}^n \to \mathbb{R}$, $v : \mathbb{R}^p \to R$ alors, en posant*

$$U := u(X_1, \ldots, X_n) \text{ et } V := v(Y_1, \ldots, Y_p),$$

les variables U et V sont indépendantes.

Lemme 19. *Si $X_1, \ldots, X_n$ sont indépendantes et si $u_1, \ldots, u_n : \mathbb{R} \to \mathbb{R}$ alors, en posant*

$$U_1 := u_1(X_1), \ldots U_n := u_n(X_n),$$

les variables $U_1, \ldots, U_n$ sont indépendantes.

Exercice 11.— Soit E une. v.a.r à valeurs dans $]0,+\infty[$ admettant une variance, α une v.a.r à valeurs dans $]0,\frac{\pi}{2}[$, $m>0$, $g>0$ deux constantes. On suppose que E et α sont indépendantes.

On pose $D=\frac{2.E.\sin(2\alpha)}{m.g}$ la distance parcourue par une sagaie lancée avec énergie (cinétique) initiale E et angle de tir α. On pose $S=\sin(2\alpha)$.

1. Montrer que S est à valeurs dans $]0,1]$ et admet une variance.
2. Exprimer l'espérance de D en fonction de m, g, $\mathbb{E}(E)$ et $\mathbb{E}(S)$.
3. Exprimer la variance de D en fonction de m, g, $\mathbb{E}(E)$, $\mathbb{E}(S)$, $\mathbb{V}(E)$ et $\mathbb{V}(S)$.
4. Montrer la formule (propagation de l'incertitude relative en physique)

$$\frac{\mathbb{V}(D)}{\mathbb{E}(D)^2}=\frac{\mathbb{V}(E)}{\mathbb{E}(E)^2}.\frac{\mathbb{V}(S)}{\mathbb{E}(S)^2}+\frac{\mathbb{V}(E)}{\mathbb{E}(E)^2}+\frac{\mathbb{V}(S)}{\mathbb{E}(S)^2}$$

Ces quantités dépendent-elles de m et g ?
5. (Optionnel). Calculer $\mathbb{E}(S)$, $\mathbb{V}(S)$ sous l'hypothèse $\alpha\sim\mathscr{U}_{]0,\frac{\pi}{2}[}$.

Exercice 12.— Un chasseur dispose d'une seule sagaie (de masse $m=1kg$), la cible est un troupeau de bisons à une distance entre 20 et 30m.

Il développe au lancer une énergie E comprise entre $E_{min}=50J$ et $E_{max}=350J$, qu'il ne contrôle pas.

Son angle de tir α, lui même incontrolé, est compris entre $\frac{\pi}{8}$ et $\frac{3\pi}{8}$.

Comme notre individu ne contrôle rien (il ne va pas survivre longtemps), on suppose qu'énergie et angle de tir sont indépendants.

1. Que valent espérance et écart-type de la distance (on arrondira aux entiers les plus proches) à laquelle la sagaie est lancée ? Quelle information obtient-on en appliquant l'inégalité de BIENAYMÉ–TCHEBYCHEFF au seuil de deux écarts-types ?

Indication:On rappelle qu'avec une énergie de lancer E, un angle de tir α, le point d'impact est à distance $d=\frac{2.E.\sin(2\alpha)}{m.g}$. On pourra interpréter « incontrolé » comme une uniformité dans le tirage au sort.

2. Tracer l'histogramme de la distance pour $NS=100000$ simulations de lancer. (Faire 50 « bins »).
3. Evaluer numériquement la probabilité de toucher un animal. Vérifier la cohérence de moyenne et écart-type de vos simulations avec espérance et écart-type théoriques.

Corrélation et indépendance

Proposition 20

1. Si X et Y sont deux variables admettant une variance, *indépendantes*, elles sont non corrélées : $\mathbb{C}ov(X,Y)=0$.
2. **La réciproque est très fausse, contrairement à ce que croit le monde entier.**

Démonstration. 1. Si X et Y sont indépendantes, en prenant $f(x)=x-\mathbb{E}(X)$ et $g(y)=y-\mathbb{E}(Y)$, $f(X).G(Y)$, $f(X)$, $g(Y)$ sont intégrables et on a donc par indépendance $\mathbb{C}ov(X,Y)=\mathbb{E}(f(X).g(Y))=\mathbb{E}(f(X))\mathbb{E}(g(Y))=0$

2. Prenons un exemple assez élémentaire. Soit X une v.a.r uniformément distribuée sur l'ensemble fini $\{-1,0,+1\}$ et $Y=X^2$. On a alors $\mathbb{E}(X)=0$, $\mathbb{E}(Y)=\frac{2}{3}$, $\mathbb{E}(XY)=\mathbb{E}(X^3)=0$ et $\mathbb{C}ov(X,Y)=0$.
Ces deux variables ne sont pas indépendantes car

$$\mathbb{P}(X=1 \text{ et } Y^2=1)=\mathbb{P}(X=1)=\frac{1}{3},\mathbb{P}(X=1)\mathbb{P}(Y^2=1)=\frac{2}{9}$$

$\square$