Probabilités conditionnelles, loi binomiale et loi de Poisson

Ex. 1 Le code génétique est basé sur les quatres lettres A, T, G, C et se lit en triplets. Sachant que les protéines sont composés de 20 acides aminés différents, expliquez pourquoi la lecture unitaire ou en couple n'est pas possible.

Ex. 2 Un scientifique souhaite effectuer un test sur n individus afin de déterminer s'ils sont porteurs d'une maladie assez rare. Ayant un budget réduit, il se demande quelle est la procédure la plus économique. Il compare donc deux stratégies.

a) S'il regroupe les n individus, il effectue un test groupé (tous les individus ensembles), puis s'il est positif, il effectue les tests individuellement. Montrez que le nombre moyen de tests dans ce cas est $((1-p)^n + (n+1)(1-(1-p)^n))$.

Supposons maintenant qu'ils sont malades avec une probabilité p suffisament petite pour avoir au plus un malade.

b) Il sépare les individus en r groupes. Si un groupe est positif, alors il test séparemment tous les individus du groupe. Montrez que le nombre moyen de test dans ce cas est $(r+\frac{n}{r}(1-(1-p)^n))$.

c) Déterminez la probabilité p pour laquelle il a intérêt à utiliser le test 1 d'abord dans le cas général, puis avec $n = 100, r = 10$.

Ex. 3 On considère un jugement pour lequel la condamnation doit être validée par au moins 8 des 12 membres du jury pour être effective. En admettant que les jurés se prononcent indépendamment les uns des autres et que la probabilité de décision correcte est θ, déterminez

a) la probabilité que la décision du jury soit correcte s'il est innocent.

b) la probabilité que la décision du jury soit correcte s'il est coupable.

c) la probabilité que la décision du jury entier soit correcte, sachant que l'accusé a une probabilité α d'être coupable.

Ex. 4 Dans un lac, on veut estimer le nombre de poissons. Après une première capture, 1492 poissons sont marqués. Lors d'une seconde capture de 2560 poissons, on dénombre 340 marqués. Estimez le nombre de poissons dans le lac et déterminez la loi des poissons marqués.

Ex. 5 On admet que le nombre d'erreurs par page dans un livre suit une loi de Poisson avec paramètre $\lambda = 0.5$. Calculez la probabilité qu'il y ait au moins une erreur sur cette page.

Ex. 6 Une installation électronique composée de plusieurs composants fonctionne si au moins deux tiers des composants est en état de marche. Chaque composant est en état de marche avec la même probabilité p. Le fonctionnement d'un composant est indépendant du fonctionnement des autres. Pour quelles valeurs de p, un système à 2 composants fonctionne-t-il plus souvent qu'un système à 4 composants ?

WWW.TAALIB24.COM

Ex. 7 Soit X une variable aléatoire distribuée selon une loi de Poisson de paramètre $\lambda > 0$.

a) Quelle est la valeur de λ qui maximise $P(X = 2)$? Quelle est la valeur de λ qui maximise $P(X = 2)P(X = 1)$?
Indication : Maximiser une fonction positive ou maximiser le logarithme de cette fonction sont des problèmes équivalents.

b) On considère une agglomération de $N = 100'000$ habitants. On modélise le nombre d'habitants qui sont infectées par le virus de la grippe chaque jour par une variable aléatoire de Poisson. L'expérience procède par sondage. Chaque jour, cinquante personnes non encore testées sont choisies au hasard dans la population et le nombre de personnes infectées dans ce groupe est comptabilisé. Après la première journée, vous enregistrez deux personnes malades. Sur la base de cette information, ajustez le modèle poissonnien; autrement dit, estimez la valeur du paramètre λ.

c) Utilisez les données ci-dessous pour estimer la probabilité que le onzième jour, sur les cinquante personnes testées, personne ne présentera de symptômes.

$$2 \quad 1 \quad 2 \quad 3 \quad 2 \quad 0 \quad 2 \quad 0 \quad 2 \quad 0$$

Indication : Trouvez la valeur de λ qui maximise $P(X = 2)P(X = 1) \cdots\cdots P(X = 0)$.

Ex. 8 a) Soit X une variable aléatoire binomiale de paramètres $n \geq 3$ et p. Si on suppose que la valeur de n est connue, quelle est la valeur de p qui maximise $P(X = 3)$?

b) Vous lancez une pièce de monnaie dix fois et obtenez deux fois PILE. Estimez la probabilité d'obtenir deux PILE ou plus lors de votre prochaine série de dix lancers.

Ex. 9 Pour évaluer le volume de métastases des poumons dans un groupe de souris, on découpe les organes en N fines tranches de $2mm$. Le volume est évalué comme suit

$$\hat{V} = \text{épaisseur} \times \text{aire moyenne associée à un point} \times (P_1 + \cdots + P_N),$$

où P_i représente le nombre de points (unités) présentant des tissus cancéreux dans la $i^{\text{ème}}$ tranche. L'aire moyenne vaut $0.046mm^2$. Si $(P_1 + \cdots + P_N)$ suit une loi de Poisson et que $P(P_1 + \cdots + P_N = 0) \approx 10^{-66}$, évaluez le volume moyen des métastases.

Ex. 10 Soit X_1 et X_2 deux variables aléatoires indépendantes distribuées selon la loi de Poisson de paramètres λ_1 et λ_2 respectivement. Calculez $P(X_1 + X_2 = x)$. Que pouvez-vous en déduire ?

Ex. 11 Lors d'une expérience, on a évalué la sensibilité de deux types de cellules par rapport à un produit mutagène en estimant le taux de mutation λ. L'expérience procède comme suit. On dépose dans le fond de $n = 100$ puits des échantillons de solution contenant chacun $N = 10^5$ cellules. Ensuite, on recouvre ces échantillon d'une solution liquide contenant une certaine concentration du produit mutagène. On laisse agir et au final, on comptabilise n_0, le nombre de puits dans lesquels aucune mutation n'est apparue.

a) Si l'on souhaite que la probabilité qu'un puit présente une colonie de cellules mutées s'élève à $1/2$, quelle est la relation qui lie N à λ ?

On a effectué cette expérience pour différentes concentration du produit et dans chaque cas, on a répété la mesure cinq fois. Les données récoltées sont fournies ci-après. Pour chacune des combinaison *type de cellule - concentration du produit mutagène* :

WWW.TALIB24.COM

b) Evaluez le taux de mutation moyen $\bar{\lambda}$.

c) Déterminez pour ce taux de mutation moyen $\bar{\lambda}$ le nombre de puits qui n'ont présenté aucune mutation.

d) Discutez vos résultats.

Type	Concentration	Fraction mutée
TK6	0.00	0.0000027
TK6	0.00	0.0000078
TK6	0.00	0.0000042
TK6	0.00	0.0000018
TK6	0.00	0.0000036
TK6	1.25	0.0000020
TK6	1.25	0.0000097
TK6	1.25	0.0000199
TK6	1.25	0.0000020
TK6	1.25	0.0000092
TK6	2.50	0.0000010
TK6	2.50	0.0000102
TK6	2.50	0.0000111
TK6	2.50	0.0000117
TK6	2.50	0.0000139
MT1	0.00	0.0000394
MT1	0.00	0.0001168
MT1	0.00	0.0000608
MT1	0.00	0.0000614
MT1	0.00	0.0001510
MT1	1.25	0.0000400
MT1	1.25	0.0001300
MT1	1.25	0.0000692
MT1	1.25	0.0000817
MT1	1.25	0.0001450
MT1	2.50	0.0000686
MT1	2.50	0.0001621
MT1	2.50	0.0001778
MT1	2.50	0.0000932
MT1	2.50	0.0002130

Fraction de cellules mutées lors d'une expérience avec une substance mutagène (PhIP). L'expérience a été effectuée pour deux types de cellules.

1ère colonne : type de cellule

2ème colonne : concentration de la substance mutagène

3ème colonne : fraction de cellules mutées

WWW.TALIB24.COM

Corrigé Probabilité conditionnelle, loi binomiale et loi de Poisson

Ex. 1 Si un nucléotide codait une protéine, il ne pourrait exister que quatre protéines différentes. Dans le cas d'une lecture en couple, il n'existerait que $4 \cdot 4 = 16$ protéines. Les nucléotides ne peuvent donc se lire qu'en triplet.

Ex. 2 a) On calcule le nombre de tests à effectuer dans chacun des cas. Si le test est négatif, on fait un test avec probabilité $(1-p)^n$ car aucune personne n'est malade. Si le test est positif, on effectue alors n tests en plus de l'initial. Cet évènement se réalise avec une probabilité $\mathrm{P}(\text{Au moins un malade}) = 1 - \mathrm{P}(\text{Aucun malade}) = 1 - (1-p)^n$.

b) D'après le même raisonnement, il réalise r tests avec probabilité 1. A ceci s'ajoute éventuellement les tests des membres du groupe positif (unique car un seul malade potentiel), donc n/r avec probabilité $\mathrm{P}(\text{Au moins un malade}) = 1 - \mathrm{P}(\text{Aucun malade}) = 1 - (1-p)^n$.

c) On cherche la valeur p pour laquelle

$$(1-p)^n + (n+1)(1-(1-p)^n) < r + \frac{n}{r}(1-(1-p)^n).$$

Après regroupement des termes adéquats, on doit déterminer p tel que

$$(1-\frac{r}{n}) < (1-p)^n,$$

c'est-à-dire

$$p < 1 - \sqrt[n]{1-\frac{r}{n}},$$

ce qui donne pour l'application numérique approximativement $p \lesssim 0.001$.

Ex. 3 a) Pour que le jury rende une sentence correcte s'il est innocent, il faut qu'entre cinq et douze juges prennent une décision correcte. Cet évènement a pour probabilité $\sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$.

b) Pour que le jury rende une sentence correcte s'il est coupable, il faut qu'entre huit et douze juges prennent une décision correcte. La probabilité de cet évènement a pour valeur $\sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$.

c) Il suffit d'utiliser la formule des probabilités totales pour ce calcul en conditionnant sur le fait d'être coupable. On trouve $\alpha \sum_{i=8}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i} + (1-\alpha) \sum_{i=5}^{12} \binom{12}{i} \theta^i (1-\theta)^{12-i}$. Notez que l'on a supposé une sorte de symétrie dans le comportement des membres du jury en posant

$$\begin{aligned} \theta &= \mathrm{P}(\text{le membre dit coupable} \mid \text{l'accusé est coupable}) \\ &= \mathrm{P}(\text{le membre dit innocent} \mid \text{l'accusé est innocent}). \end{aligned}$$

Ex. 4 Il s'agit d'appliquer une règle de trois. En effet, après la première capture, on a une proportion de poissons marqués de $\frac{1492}{N}$ où N désigne le nombre total de poissons dans le lac. Après remise, on en retrouve une proportion de $\frac{340}{2560}$. On obtient donc l'estimation suivante de $N = \frac{1492 * 2560}{340} = 11233$. Donc si p.ex. on tire une troisième fois des poissons, plus précisément si on tire avec remise n_3 poissons, le nombre de poissons marqués parmi cet échantillon va suivre une loi binomiale $Bin(n_3, \hat{p} = \frac{340}{2560})$.

WWW.TALIB24.COM

Ex. 5 La probabilité qu'il y a au moins une erreur sur cette page se calcule comme

$$\mathrm{P}(\text{Au moins une erreur}) = 1 - \mathrm{P}(\text{Aucune erreur}) = 1 - \frac{0.5^0 \exp(-0.5)}{0!} = 1 - \exp(-0.5)\,.$$

Ex. 6 On cherche p tel que la probabilité que plus de deux tiers des composants de l'installation à deux composants fonctionnent soit supérieure à la probabilité que plus de deux tiers des composants de l'installation à quatre composants fonctionnent. Autrement dit, on cherche p tel que

$$\begin{aligned}
\binom{2}{2}p^2 &> \binom{4}{3}p^3(1-p) + \binom{4}{4}p^4\,,\\
\Leftrightarrow\quad 3p^4 - 4p^3 + p^2 &> 0\,,\\
\Leftrightarrow\quad 3p^2 - 4p + 1 &> 0\,,\\
\Leftrightarrow\quad (p-1)(3p-1) &> 0\,,\\
\Leftrightarrow\quad (3p-1) &< 0\,,\\
\Leftrightarrow\quad p &< \frac{1}{3}\,.
\end{aligned}$$

Ex. 7 a) On applique la méthode d'estimation par maximum de vraisemblance. Dans cette méthode, on ne considère plus la fonction de probabilité comme une fonction de la variable d'intérêt (ici, X) mais comme une fonction du paramètre (ici, λ). A toute valeur du paramètre, cette fonction lui associe un score (une valeur) qui traduit la vraisemblance que ce que l'on observe provient du modèle choisi avec la valeur du paramètre en question. Dès lors, maximiser cette fonction apparaît comme une méthode sensée d'estimer la valeur du paramètre. Autrement dit, ici, on cherche la valeur $\hat{\lambda}_{\mathrm{MV}}$ tel que $\log \mathrm{P}_\lambda(X=2) < \log \mathrm{P}_{\hat{\lambda}_{\mathrm{MV}}}$, $\forall \lambda > 0$. On peut écrire

$$\log \mathrm{P}_\lambda(X=2) = \log\left(\mathrm{e}^{-\lambda}\frac{\lambda^2}{2!}\right) \propto -\lambda + 2\log(\lambda)\,.$$

La somme d'une fonction linéaire et d'une fonction logarithmique étant concave, pour trouver l'argument qui la maximise, on peut la dériver et chercher la valeur pour laquelle cette dernière s'annule. On obtient

$$-1 + \frac{2}{\hat{\lambda}_{\mathrm{MV}}} = 0\,.$$

Autrement dit, la valeur qui maximise la fonction de vraisemblance est $\hat{\lambda}_{\mathrm{MV}} = 2$. De même si l'on cherche la valeur du paramètre qui maximise $\mathrm{P}(X=2)\mathrm{P}(X=1)$, on trouve $\hat{\lambda}_{\mathrm{MV}} = (2+1)/2$. On peut généraliser ces résultats en disant que si l'on a observé n fois de manière indépendante une variable aléatoire Poisson(λ), alors

$$\hat{\lambda}_{\mathrm{MV}} = \bar{X}\,.$$

b) On approxime le processus $Bin(n,p)$ par un processus de Poisson($\lambda = np$). Puisque l'estimation de p après un jour vaut $\hat{p} = 2/50 = 0.04$, on obtient l'estimation $\hat{\lambda}_{\mathrm{MV}} = 2$. Autrement dit, on estime le nombre de personnes malades dans toute la ville à $\frac{N}{n}\mathrm{E}(X) = \frac{N}{n}\lambda = 4'000$.

c) On a que $\hat{\lambda}_{\mathrm{MV}} = \bar{X} = 1.4$. Par conséquent, $\hat{\mathrm{P}}(X=0) = \mathrm{e}^{-\hat{\lambda}_{\mathrm{MV}}} \approx 0.25$.

WWW.TALD24.COM

Ex. 8 a) Comme dans l'exercice précédent, il s'agit d'appliquer la méthode du maximum de vraisemblance. Il faut résoudre l'équation

$$\left(\frac{\mathrm{d}}{\mathrm{d}p}\left(\log \mathrm{P}(X=3)\right)\right)\Bigg|_{p=\hat{p}_{\mathrm{MV}}} = 0\,.$$

On trouve $\hat{p}_{\mathrm{MV}} = 3/n$.

b) Par déduction du point a), on a $\hat{p}_{\mathrm{MV}} = 2/10$. Par conséquent,

$$\begin{aligned}\mathrm{P}_{\hat{p}_{\mathrm{MV}}}(X \geq 2) = 1 - \mathrm{P}_{\hat{p}_{\mathrm{MV}}}(X < 2) &= 1 - \mathrm{P}_{\hat{p}_{\mathrm{MV}}}(X=0) - \mathrm{P}_{\hat{p}_{\mathrm{MV}}}(X=1)\\ &= 1 - \binom{10}{0}\hat{p}_{\mathrm{MV}}^0(1-\hat{p}_{\mathrm{MV}})^{10-0} - \binom{10}{1}\hat{p}_{\mathrm{MV}}^1(1-\hat{p}_{\mathrm{MV}})^{10-1}\\ &\approx 0.62\,.\end{aligned}$$

Ex. 9 Puisque $\mathrm{P}(P_1 + \cdots + P_N = 0) = \mathrm{e}^{-\lambda} \approx 10^{-66}$, on a que $\lambda \approx 66 * \log(10) \approx 152$. Par conséquent, $\mathrm{E}(P_1 + \cdots + P_N) = \lambda = 152$. On en déduit que

$$\hat{V} = 2 \times 0.046 \times 152 \approx 14mm^3\,.$$

Ex. 10 Puisque $X_1 \sim \mathrm{Poisson}(\lambda_1)$ et $X_2 \sim \mathrm{Poisson}(\lambda_2)$, indépendantes, on a que

$$\begin{aligned}\mathrm{P}(X_1 + X_2 = x) &= \sum_{y=0}^{x} \mathrm{P}(X_1 = y, X_2 = x - y)\\ &= \sum_{y=0}^{x} \mathrm{P}(X_1 = y)\mathrm{P}(X_2 = x - y)\\ &= \sum_{y=0}^{x} \mathrm{e}^{-\lambda_1}\frac{\lambda_1^y}{y!}\mathrm{e}^{-\lambda_2}\frac{\lambda_2^{x-y}}{(x-y)!}\\ &= \mathrm{e}^{-(\lambda_1+\lambda_2)}\sum_{y=0}^{x}\frac{\lambda_1^y}{y!}\frac{\lambda_2^{x-y}}{(x-y)!}\\ &= \mathrm{e}^{-(\lambda_1+\lambda_2)}\frac{1}{x!}\sum_{y=0}^{x}\frac{x!}{(x-y)!y!}\lambda_1^y\lambda_2^{x-y}\\ &= \mathrm{e}^{-(\lambda_1+\lambda_2)}\frac{1}{x!}(\lambda_1+\lambda_2)^x\\ &= \mathrm{e}^{-\bar{\lambda}}\frac{1}{x!}\bar{\lambda}^x\end{aligned}$$

où $\bar{\lambda} = \lambda_1 + \lambda_2$. Autrement dit la somme de deux variables aléatoires indépendantes de Poisson est encore une variable aléatoire de Poisson avec pour paramètre la somme des deux paramètres.

Ex. 11 a) Dans cette expérience on approche la probabilité qu'il n'y aie pas de réaction dans un puit ($= (1-\lambda)^N$) par le nombre de puits sans réaction sur le nombre de puits total ($= n_0/n$). On veut donc

$$\frac{1}{2} = \frac{n_0}{n} = (1-\lambda)^N\,.$$

WWW.TALIB24.COM

Puisque λ est supposément très faible, on peut développer la fonction $(1-\lambda)^N$ à l'ordre 1 autour de zéro :

$$(1-\lambda)^N \approx 1 - N\lambda .$$

Ainsi, on trouve la relation $N\lambda \approx 1/2$. Typiquement, si on s'attend à un taux de concentration de l'ordre de 10^{-6}, il faudra mettre dans chaque puits $N \approx (1/2) \cdot 10^6$ cellules.

b) Tableau des taux de mutation moyens ($\times 10^{-6}$) :

	0	1.25	2.5
TK6	4.02	8.56	9.58
MT1	85.88	93.18	142.94

c) On a que $n_0 = n(1-\lambda)^N$. Le tableau suivant fourni le nombre moyen de puits sans mutation au cours des différentes expériences :

	0	1.25	2.5
TK6	67	43	38
MT1	0.02	0.01	0.0001

WWW.TALIB24.COM

Fonction de répartition, variables conjointes, indépendance

Ex. 1 On jette trois pièces équilibrées. On désigne par X le nombre de PILE obtenus.

a) Calculez la loi de probabilité de la variable aléatoire X (*i.e.* $f_X(k)$).

b) Calculez la fonction de répartition de X, (*i.e.* F_X) et représentez-la graphiquement.

Ex. 2 On tire aléatoirement un nombre réel, X entre 0 et 1 et on calcule sa racine carrée qu'on note Y. Trouvez les fonctions de répartition de X et de Y.

Ex. 3 La fonction de répartition d'une variable aléatoire Y est donnée par

$$F_Y(y) = \begin{cases} 0, & \text{si } y < 0, \\ y/2, & \text{si } 0 \leq y < 1, \\ 2/3, & \text{si } 1 \leq y < 2, \\ 11/12, & \text{si } 2 \leq y < 3, \\ 1, & \text{si } y \geq 3. \end{cases}$$

a) Représentez graphiquement F_Y.

b) Calculez

i) $\mathrm{P}(Y < 3)$.

ii) $\mathrm{P}(Y > \frac{1}{2})$.

iii) $\mathrm{P}(Y = 1)$.

iv) $\mathrm{P}(2 < Y \leq 4)$.

Ex. 4 Soit $Y = 3 + 5X$ une transformation linéaire de X.

a) Exprimez la fonction de répartition de Y en fonction de celle de X.

b) Exprimez $\mathrm{E}(Y)$ et $\mathrm{Var}(Y)$ en fonction de $\mathrm{E}(X)$ et $\mathrm{Var}(X)$.

c) Calculez la covariance entre X et Y ainsi que la variance de $Y - X$.

Ex. 5 Soit X une variable aléatoire normalement distribuée de paramètres $\mathrm{E}(X) = \mu$ et $\mathrm{Var}(X) = \mathrm{E}\big((X-\mu)^2\big) = \sigma^2$. On peut démontrer que

$$\mathrm{E}\big((X-\mu)^3\big) = 0 \qquad \text{et} \qquad \mathrm{E}\big((X-\mu)^4\big) = 3\sigma^4.$$

En toute généralité, l'asymétrie et le coefficient d'aplatissement d'une variable aléatoire Y sont définis comme suit.

$$\text{asymétrie} = \frac{\mathrm{E}((Y-\mu)^3)}{\sigma^3} \qquad \text{et} \qquad \text{coefficient d'aplatissement} = \frac{\mathrm{E}((Y-\mu)^4)}{\sigma^4}.$$

Calculez l'asymétrie et le coefficient d'aplatissement d'une variable aléatoire normale et d'une variable aléatoire exponentielle de paramètre λ.

WWW.TALIB24.COM

Ex. 6 Une solution contient un million de cellules par millilitre. Parmi ces cellules, une fraction τ est mutée. Pour estimer la valeur de τ, on étale un millilitre de cette solution dans quatre boîtes de Pétri en présence d'une condition sélective qui permet uniquement aux cellules mutées de se diviser. Après un certain temps, on comptabilise le nombre de colonies dans les diverses boîtes. On trouve les valeurs $38, 95, 112$ et 54. Proposez une estimation de τ.

Ex. 7 Soit T le temps que vit un poisson d'une certaine espèce. On suppose que la fonction de répartition de T est donné par

$$F_T(t) = \mathrm{P}(T \leq t) = \begin{cases} 0, & \text{si } t < 0\,; \\ \left(\frac{t}{10}\right)^2, & \text{si } 0 \leq t \leq 10\,; \\ 1, & \text{si } t > 10\,. \end{cases}$$

a) Déterminez la densité de T.

b) Avec quelle probabilité le poisson va mourir quand il sera entre trois et quatre ans ? Avec quelle probabilité le poisson atteint l'âge de huit ans ?

Ex. 8 Un homme et une femme se donnent rendez-vous. L'heure d'arrivée de chacune de ces deux personnes sur les lieux du rendez-vous est une variable aléatoire uniforme entre midi et une heure. Ces deux variables sont indépendantes. Quelle est la probabilité que la première arrivée doive attendre plus de dix minutes ?

Ex. 9 Supposons qu'un guichet de la poste sert en moyenne deux clients par cinq minutes. Soient X_1 et X_2 les temps de service de chacun. On suppose que ces temps sont indépendants et identiquement distribués selon une loi exponentielle $\mathcal{E}(\lambda)$. Posons $T = X_1 + X_2$.

a) Calculez l'espérance et la variance de T. Quelle valeur pourrait-on choisir pour λ ?

b) Déterminez la densité de T. Esquissez les densités de T et de X_1.

Ex. 10 On considère le couple de variables aléatoires (X, Y). La densité conjointe est donnée par

$$f(x,y) = \begin{cases} c \cdot xy, & \text{si } x, y \in [0,1]\,, \\ 0, & \text{sinon,} \end{cases}$$

où c est une constante.

a) Déterminez c.

b) Est-ce que les variables X et Y sont indépendantes ?

c) Esquissez les courbes de niveaux de la densité $f(x,y)$.

WWW.TALIB24.COM

Ex. 11 Un biologiste veut cultiver des plantes issues d'une expérience de croisements. Pour pouvoir planifier l'étude, il considère des variables aléatoites $X_1, \ldots, X_n$, où X_i compte le nombre de germes présents sur le $i^{\text{ième}}$ terrain d'essai après un certain temps d'attente. Supposons que ces variables sont indépendantes et identiquement distribuées selon une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = 3$.

a) Soit $n = 5$. Quelle est la probabilité que sur les cinq terrains ensembles on compte exactement quinze germes ?

b) Soit maintenant n quelconque. Quelle est la probabilité (en fonction de n) de n'avoir aucun germe ?

c) Que doit valoir n pour avoir au moins deux germes avec une probabilité d'au moins 99% ?

Ex. 12 a) Un généticien étudie deux gènes situés sur deux chromosomes différents. Chacun de ces gènes a deux allèles : A, a et B, b. Il décide de croiser deux plantes de génotype $AaBb$. Pour décrire les génotypes des descendants, on peut introduire les variables aléatoires X à valeurs dans $\{AA, Aa, aa\}$ et Y à valeurs dans $\{BB, Bb, bb\}$. Donnez la loi conjointe de (X, Y).

b) Selon la Croix-Rouge des Etats-Unis, la distribution (en pourcent) des groupes sanguins dans son pays est la suivante :

	R+	R-
A	34	6
B	9	2
AB	3	1
O	38	7

On se demande si les deux variables *groupe* (A, B etc.) et *Rhésus* (+ et -) sont distribuées de manière indépendante. Comparez la distribution observée avec la distribution théorique sous l'hypothèse de l'indépendance.

WWW.TALIB24.COM

Corrigé : **Collection 06 : Fonction de répartition, variables conjointes, indépendance**

Ex. 1 **a)** Notre expérience consiste à jeter trois pièces équilibrées. X est une variable aléatoire et peut prendre les valeurs $0, 1, 2, 3$ avec les probabilités :

$$\mathrm{P}(X=0) = 1/8,\ \ \mathrm{P}(X=1) = 3/8,\ \ \mathrm{P}(X=2) = 3/8,\ \ \mathrm{P}(X=3) = 1/8\,.$$

Cette variable, X, suit une loi binomiale $Bin(3, 1/2)$.

b) La fonction de répartion est donnée par

$$F(x) = \mathrm{P}(X \leq x) = \begin{cases} 0\,, & \text{si } x < 0\,, \\ 1/8\,, & \text{si } 0 \leq x < 1\,, \\ 4/8\,, & \text{si } 1 \leq x < 2\,, \\ 7/8\,, & \text{si } 2 \leq x < 3\,, \\ 1\,, & \text{si } x \geq 3\,. \end{cases}$$

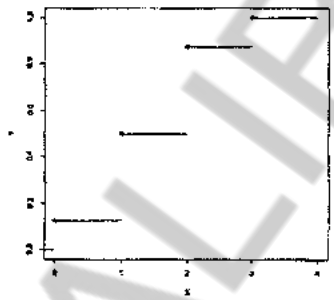


FIG. 1 – La fonction de répartition de X.

Ex. 2 X est une variable aléatoire uniforme et $Y = \sqrt{X}$ est une transformation de X. Les fonctions de répartition de X et de Y sont données par

$$F_X(x) = \mathrm{P}(X \leq x) = \begin{cases} 0\,, & \text{si } x < 0\,, \\ x\,, & \text{sinon}\,, \\ 1\,, & \text{si } x \geq 1\,. \end{cases}$$

$$F_Y(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(\sqrt{X} \leq y) = \mathrm{P}(X \leq y^2) = \begin{cases} 0\,, & \text{si } y < 0\,, \\ y^2\,, & \text{sinon}\,, \\ 1\,, & \text{si } y \geq 1\,. \end{cases}$$

Ex. 3 **a)** Cf Fig. 2.

b) Les probabilités sont égales à

i) $\mathrm{P}(Y < 3) = \frac{11}{12}$,

ii) $\mathrm{P}(Y > \frac{1}{2}) = \frac{3}{4}$,

WWW.TALIB24.COM

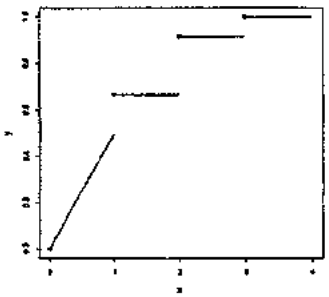


FIG. 2 – La fonction de répartition de X.

iii) $P(Y = 1) = \frac{1}{6}$,

iv) $P(2 < Y \leq 4) = \frac{1}{12}$.

Ex. 4 Soient X et Y deux variables aléatoires et $Y = 3 + 5X$ une transformation linéaire de X.

a) La fonction de répartition de Y en fonction de celle de X vaut :

$$F_Y(y) = P(Y \leq y) = P(3 + 5X \leq y) = P\Big(X \leq \frac{y-3}{5}\Big) = F_X\Big(\frac{y-3}{5}\Big).$$

b) L'espérance et la variance de la variable aléatoire Y valent $E(Y) = 5E(X) + 3$ et $Var(Y) = 25Var(X)$.

c) On calcule $Cov(X, Y) = 5Var(X)$ et $Var(Y - X) = Var(Y) + Var(X) - 2Cov(Y, X) = 16Var(X)$.

Ex. 5 Pour une variable aléatoire X de loi normale $\mathcal{N}(\mu, \sigma^2)$ il s'ensuit directement que l'asymétrie vaut 0 et le coefficient d'aplatissement vaut 3.

Soit Y une variable aléatoire exponentielle de paramètre λ. L'asymétrie vaut,

$$E((Y-\mu)^3) = E\big((Y - \frac{1}{\lambda})^3\big) = \int_0^\infty (y - \frac{1}{\lambda})\lambda e^{-\lambda y} dy.$$

On fait une intégration par parties et on trouve,

$$E\big((Y - \frac{1}{\lambda})^3\big) = \frac{2}{\lambda^3} \qquad E\big((Y - \frac{1}{\lambda})^4\big) = \frac{9}{\lambda^4}$$

Par conséquent l'asymétrie est égale à 2. De la même façon, on calcule le coefficient d'aplatissement. il vaut 9.

Ex. 6 Une méthode simple pour estimer le taux de mutation (noté τ) est :

$$\hat{\tau} = \frac{\sum_{i=1}^n c_i}{nN},$$

où n est le nombre de boîte de petri, N est le nombre de cellules dans chaque boîte et c_i est le nombre de colonies qui se sont développées dans la i^e boîte. Par conséquent, $\hat{\tau} = 74.75 \times 10^{-6}$.

Ex. 7 **a)** On dérive la fonction de répartition par rapport à t pour trouver la densité $f_T(t) = 2t/10^2$ si $0 \leq t \leq 10$ (et $f_T(t) = 0$ sinon).

WWW.TALIB24.COM

b) On calcule

$$P(3 \leq T \leq 4) = \int_3^4 f_T(t)\,dt = F_T(4) - F_T(3) = 0.07\,,$$

et

$$P(T \geq 8) = 1 - F_T(8) = 0.36\,.$$

Ex. 8 On a les variables aléatoires X, Y indépendantes et de loi uniforme sur $[0, 60]$. On calcule l'intégral

$$\iint_{\{|x-y|>10\}} f(x,y)\,dx\,dy\,,$$

ce qui donne

$$P(|X - Y| > 10) = \frac{50^2}{60^2} = \frac{25}{36}\,.$$

Ex. 9 a) On a $E(T) = E(X_1) + E(X_2) = 2/\lambda$ et comme X_1 et X_2 sont indépendantes $Var(T) = Var(X_1) + Var(X_2) = 2/\lambda^2$. Comme le temps moyen pour deux clients est cinq minutes il est naturel de poser

$$E(T) = 5 \implies \hat{\lambda} = 2/5\,.$$

b) On peut utiliser la formule de la convolution,

$$f_T(t) = \int f_{X_1}(x) f_{X_2}(t-x)\,dx = \int_0^t \lambda e^{-\lambda x} \lambda e^{-\lambda(t-x)}\,dx = \lambda^2 e^{-\lambda t} t\,.$$

Les densités de X_1 et de T sont données dans la Figure 3.

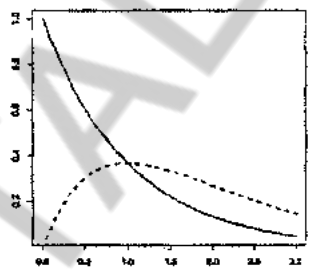


FIG. 3 – La densité de X_1 et la densité de T (traitillé) où l'on a choisi $\lambda = 1$.

Ex. 10 a) Comme l'intégrale sous la densité doit donner un, on trouve $c = 4$.

b) On peut écrire

$$f(x,y) = 2x \cdot 2y\,,$$

ce qui montre que $f(x,y) = f(x)f(y)$. Donc X et Y sont indépendantes.

c) Cf. Fig. 4.

WWW.TALIB24.COM

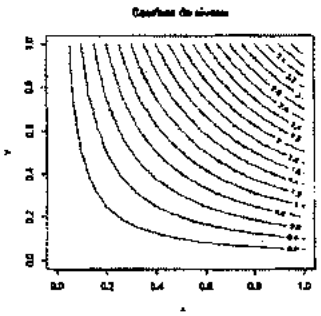


FIG. 4 – La densité conjointe $f(x, y)$.

Ex. 11 **a)** On a vu dans l'exercice 10 de la collection 5 que la somme de variables indépendentes de loi de Poisson va encore suivre une loi de Poisson. Plus précisément, $S_5 = X_1 + \cdots + X_5 \sim \mathcal{P}(5\lambda)$. Comme $\lambda = 3$,

$$\mathrm{P}(S_5 = 15) = e^{-15}\frac{15^{15}}{15!} \approx 0.1.$$

b) On a $S_n = X_1 + \cdots + X_n \sim \mathcal{P}(3n)$ et donc $\mathrm{P}(S_n = 0) = e^{-3n}$.

c) Or, $\mathrm{P}(S_n \geq 2) = 1 - e^{-3n}(1 + 3n)$. Calculons cette fonction pour quelques n :

n	$\mathrm{P}(S_n \geq 2)$
1	0.8
2	0.982
3	0.999

Il faut donc considérer au moins $n = 3$ terrains d'essai.

Ex. 12 **a)** Chacun des parents donne un des gamètes AB, Ab, aB, ab avec probabilité $1/4$. On fait un tableau de croisements pour en déduire la distribution de (X,Y) :

	AA	Aa	aa
BB	1/16	1/8	1/16
Bb	1/8	1/4	1/8
bb	1/16	1/8	1/16

b) Les deux traits sont indépendants si on a $\mathrm{P}(X = i, Y = j) = \mathrm{P}(X = i)\mathrm{P}(Y = j)$ pour toues les valeurs possibles de X et Y. On peut comparer la distribution réelle,

Loi observée :

	R+	R-	
A	0.34	0.06	0.4
B	0.09	0.02	0.11
AB	0.03	0.01	0.04
O	0.38	0.07	0.45
	0.84	0.16	1

avec la distribution théorique sous l'hypothèse de l'indépendance.

WWW.TALIB24.COM

Loi sous l'indép. :

	R+	R-
A	0.336	0.064
B	0.0924	0.0176
AB	0.0336	0.0064
O	0.378	0.072.

On voit que les deux distributions sont très proches et les deux traits semblent d'être distribués de manière indépendants.

WWW.TALIB24.COM

Puisque λ est supposément très faible, on peut développer la fonction $(1-\lambda)^N$ à l'ordre 1 autour de zéro :

$$(1-\lambda)^N \approx 1 - N\lambda\,.$$

Ainsi, on trouve la relation $N\lambda \approx 1/2$. Typiquement, si on s'attend à un taux de concentration de l'ordre de 10^{-6}, il faudra mettre dans chaque puits $N \approx (1/2)\cdot 10^6$ cellules.

b) Tableau des taux de mutation moyens ($\times 10^{-6}$) :

	0	1.25	2.5
TK6	4.02	8.56	9.58
MT1	85.88	93.18	142.94

c) On a que $n_0 = n(1-\lambda)^N$. Le tableau suivant fourni le nombre moyen de puits sans mutation au cours des différentes expériences :

	0	1.25	2.5
TK6	67	43	38
MT1	0.02	0.01	0.0001

WWW.TALIB24.COM

Espaces de probabilité

Exercice 3.1

On pose une question au hasard à une personne dans une population comptant une proportion p de tricheurs.

On admet que si cet individu est tricheur, il connaît d'avance la réponse ; sinon il a une chance sur 12 de répondre correctement.

a) *Quelle est la probabilité que le candidat choisi connaisse la bonne réponse ?*

b) *Sachant que le candidat a répondu correctement, quelle est la probabilité qu'il ait triché ? (Résoudre de deux façons différentes.)*

Dessinez une abaque de la probabilité que le candidat a d'être tricheur s'il a bien répondu en fonction de la proportion de tricheurs dans la population.

Solution

On sait que : $P(B/T) = 1$; $P(T) = p$; $P(B/\overline{T}) = \frac{1}{12}$; $P(\overline{T}) = (1-p)$.

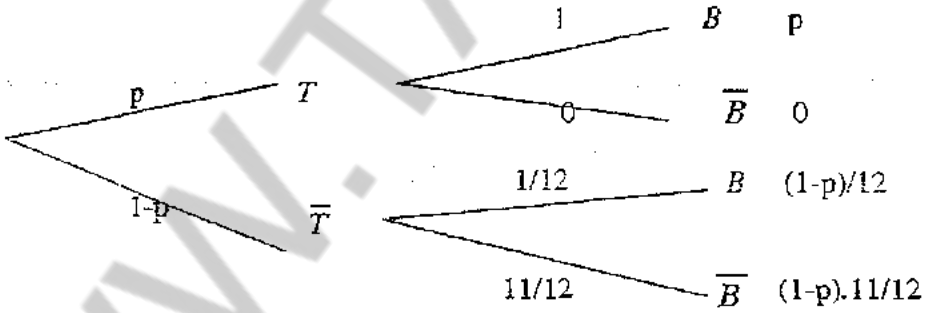
On cherche : a) $P(B)$ et b) $P(T/B)$.

a) $P(B) = P(B/T).P(T) + P(B/\overline{T}).P(\overline{T}) = 1.p + \frac{1}{12}.(1-p) = \frac{1+11p}{12}$.

b) – Par la formule de Bayes :

$$P(T/B) = \frac{P(B/T).P(T)}{P(B/T).P(T) + P(B/\overline{T}).P(\overline{T})} = \frac{P(B/T).P(T)}{P(B)} = \frac{1.p}{\frac{1+11p}{12}} = \frac{12p}{1+11p}.$$

– Par un diagramme en arbre :



Exercice 3.2

Une urne contient 9 jetons numérotés de 1 à 9. On en sort deux sans en noter les numéros

a) *Quelle est la probabilité qu'un troisième jeton tiré de l'urne soit pair ?*

b) *Quelle est la probabilité que les trois jetons soient impairs si l'un des trois l'est ?*

WWW.TALIB24.COM

Solution

Une urne contient 9 jetons numérotés de 1 à 9. On en sort deux sans en noter les numéros .

c) *Quelle est la probabilité qu'un troisième jeton tiré de l'urne soit pair ?*

P.N. La difficulté de ce problème est sa modélisation plus complexe. Particulièrement celle des conditions. Le choix qui a été fait infra a été de les décrire en extension sur base de deux événements élémentaires :

$$JP = \text{« Tirer un jeton pair. »} \text{ et } JI = \text{« Tirer un jeton impair. »} = \overline{JP}.$$

Pour décrire complètement les conditions auxquelles est soumis le résultat de l'épreuve : « *Tirer le 3e jeton.* », nous définissons les événements composés suivants :

$A1=JP,JP$; $A2=JI,JI$; $A3=JP,JI$; $A4=JI,JP$. Avec $\{Ai\ ;\ i=1, 2, 3, 4\}$ un SCE.
et B = «*Le 3è jeton tiré est pair.* ».

Donc (LPT) $P(B)=\sum_{i=1}^{4}P(B\cap Ai)=\sum_{i=1}^{4}P(B/Ai).P(Ai)$.

Or (Laplace)

$P(A1)=\frac{4}{9}\times\frac{3}{8}$ et $P(B/A1)=\frac{2}{7}$; $P(A2)=\frac{5}{9}\times\frac{4}{8}$ et $P(B/A2)=\frac{4}{7}$;

$P(A3)=\frac{4}{9}\times\frac{5}{8}$ et $P(B/A3)=\frac{3}{7}$; $P(A4)=\frac{5}{9}\times\frac{4}{8}$ et $P(B/A4)=\frac{3}{7}$.

Donc $P(B)=\frac{4.3.2+5.4.4+4.5.3.2}{9.8.7}=\frac{3+10+15}{9.7}=\frac{28}{9.7}=\frac{4}{9}$.

d) *Quelle est la probabilité que les trois jetons soient impairs si l'un des trois l'est ?*

Soit A = « *Les trois jetons sont impairs.* » = (JI, JI, JI).
Soit B = « *Un des trois jetons est impair.* » = $(JI\cap I\cap I)\cup(I\cap JI\cap I)\cup(I\cap I\cap JI)$; $I=JP\cup JI$

Or $\overline{B}=(JP,JP,JP)$ donc $P(B)=1-P(\overline{B})$ et $P(\overline{B})=\frac{4.3.2}{9.8.7}=\frac{3}{63}=\frac{1}{21}$

On cherche $P(A/B)=\frac{P(A\cap B)}{P(B)}$

or $A\cap B$ = A puisque $A\Rightarrow B$ ($A\subset B$) et $P(A)=\frac{5}{9}.\frac{4}{8}.\frac{3}{7}=\frac{5}{42}$.

Donc $P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)}{P(B)}=\frac{P(A)}{1-P(\overline{B})}=\frac{\frac{5}{42}}{\frac{20}{21}}=\frac{1}{2}.\frac{1}{4}=\frac{1}{8}$.

or $A\cap B$ = A puisque $A\Rightarrow B$ ($A\subset B$) et $P(A)=\frac{5}{9}.\frac{4}{8}.\frac{3}{7}=\frac{5}{42}$.

Donc $P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)}{P(B)}=\frac{P(A)}{1-P(\overline{B})}=\frac{\frac{5}{42}}{\frac{20}{21}}=\frac{1}{2}.\frac{1}{4}=\frac{1}{8}$.

WWW.TALIB24.COM

Exercice 3.3

En Europe occidentale, 5% des garçons et 0,25% des filles naissent daltoniens. 51% des naissances concernent des garçons.

a) *Quelle est la proportion de garçons dans la population des bébés daltoniens ?*

b) *Quelle est la proportion d'enfants daltoniens dans le total ?*

Solution

En Europe occidentale, 5% des garçons et 0,25% des filles naissent daltoniens. 51% des naissances concernent des garçons.

a) *Quelle est la proportion de garçons dans la population des bébés daltoniens ?*

Soit D = « *Etre daltonien* » ; G = « *Etre un garçon* » et F = « *Etre une fille* ».

On cherche $P(G/D)$

Par la formule de Bayes : $P(G/D) = \dfrac{P(D/G).P(G)}{P(D/G).P(G)+P(D/F).P(F)}$

Or $P(G) = 0,51$; donc $P(F) = 0,49$ puisque (G,F) est un SCE ;[2]

$P(D/G) = 0,05$ et $P(D/F) = 0,0025$.

Donc $P(G/D) = \dfrac{0,05.0,51}{0,05.0,51+0,0025.0,49} = \dfrac{0,0255}{0,026725} = 0,954$.

b) *Quelle est la proportion d'enfants daltoniens dans le total ?*

N.B. : (G,F) est un SCE.

On cherche $P(D) = P(D/G).P(G) + P(D/F).P(F) = 0,026725 \approx 2,5\%$.

Exercice 3.4

La compagnie d'assurance SECOURT-SERTAIN propose des primes réduites à ses clients à faible risque. Elle a donc réparti la population en deux classes : ceux qui sont enclins aux accidents et ceux qui ne le sont pas. Ses statistiques montrent qu'un individu enclin aux accidents a une probabilité de 0,4 d'en avoir un au cours d'une année, alors que cette probabilité tombe à 0,2 pour les gens à faible risque. On suppose que 30% de la population appartient à la classe à haut risque.

a) *Quelle est la probabilité qu'un nouvel assuré soit victime d'un accident durant la première année de son contrat ?*

b) *Quelle est la probabilité qu'il fasse partie de la classe à haut risque s'il est victime d'un accident dans l'année qui suit la conclusion de son contrat ?*

c) *Quelle est la probabilité conditionnelle pour un nouveau client d'avoir un accident durant la deuxième année de son contrat s'il a eu un accident durant la première année ?*

Solution

a) *Quelle est la probabilité qu'un nouvel assuré soit victime d'un accident durant la première année de son contrat ?*

Soit :

AI = « *Le nouvel assuré aura un accident au cours de la première année.* » ;

HR = « *Le nouvel assuré fait partie de la classe à haut risque.* » ;

WWW.TALIB24.COM

FR = « Le nouvel assuré fait partie de la classe à faible risque. ».

$P(A1) = P(A1/HR).P(HR) + P(A1/FR).P(FR) = (0,4.0,3) + (0,2.0,7) = 0,12 + 0,14 = 0,26.$

b) ***Quelle est la probabilité qu'il fasse partie de la classe à haut risque s'il est victime d'un accident dans l'année qui suit la conclusion de son contrat ?***

$P(HR/A1) = P(HR \cap A1) / P(A1) = P(A1/HR).P(HR) / P(A1) = (0,4.0,3) / 0,26 = 6/13.$

c) ***Quelle est la probabilité conditionnelle pour un nouveau client d'avoir un accident durant la deuxième année de son contrat s'il a eu un accident durant la première année ?***

Soit *A2 = « Le nouvel assuré aura un accident au cours de la deuxième année ».*

On peut conditionner sur le fait que le client peut être à haut risque ou non.

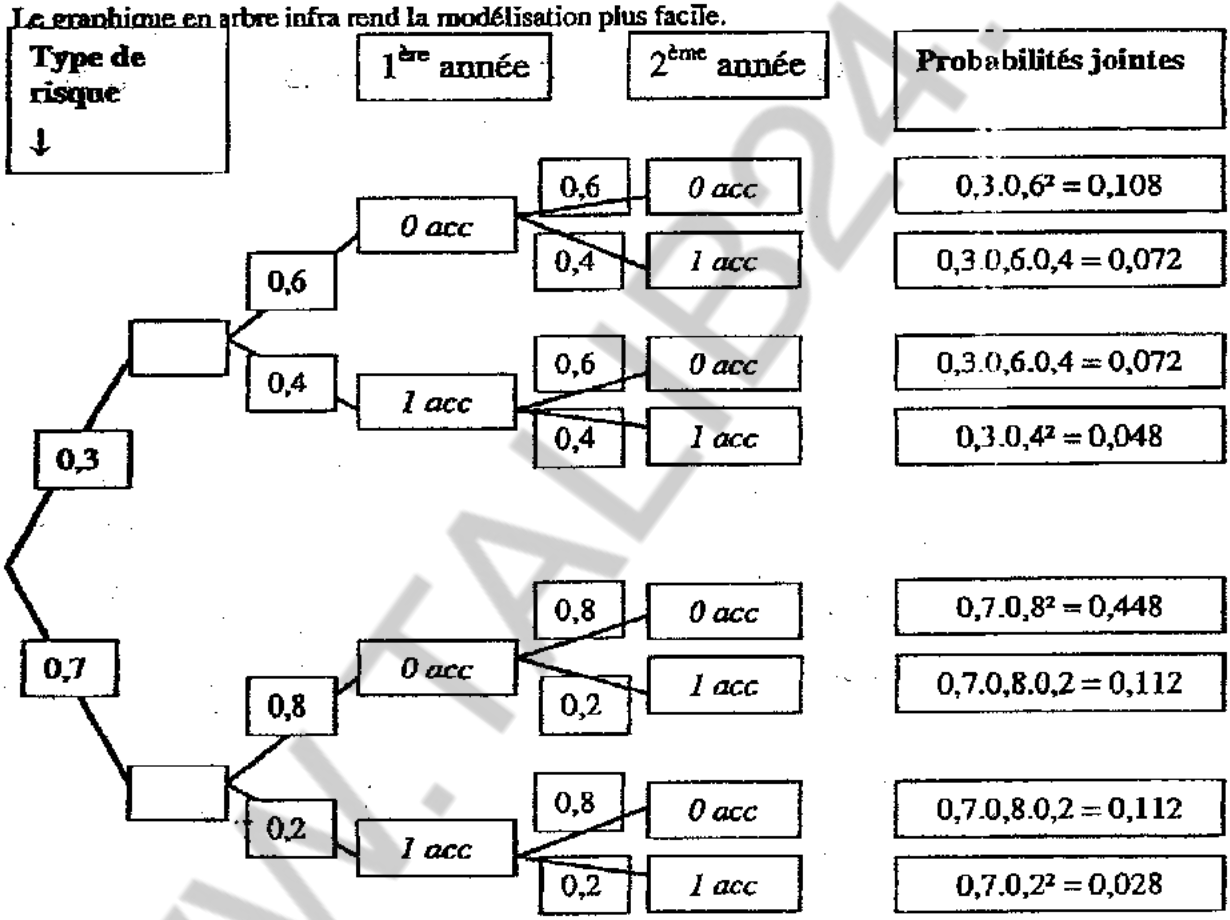
$P(A2/A1) = P(A2/HR \cap A1).P(HR/A1) + P(A2/FR \cap A1).P(FR/A1) =$

$4/10 . 6/13 + 2/10 . 7/13 =$

$24/130 + 14/130 = 38/130$

$\cong 0,292.$

Le graphique en arbre infra rend la modélisation plus facile.



On cherche $P(A2/A1) = P(A2 \cap A1)/P(A1) =$ **(d'après les probabilités jointes du diagramme)** $(0,048 + 0,028) / [(0,072+0,048) + (0,112 + 0,028)] = 0,076/0,260 = 0,292.$

WWW.TALEB24.COM

Exercice 3.5

Chacun de deux petits meubles identiques a deux tiroirs. Le meuble A contient une pièce d'argent dans chaque tiroir, le meuble B ayant une pièce d'argent dans un tiroir et une pièce d'or dans l'autre. On désigne un des deux meubles au hasard, on ouvre l'un de ses deux tiroirs et on y découvre une pièce d'argent.

a) Quelle est la probabilité que ce soit le meuble A qui a été choisi ?

b) Quelle est la probabilité qu'il y ait une pièce d'argent dans l'autre tiroir ?

c) Quelle est la probabilité qu'il y ait une pièce d'or dans l'autre tiroir ?

Solution

Soit $A1$ = « *Découvrir une pièce d'argent dans un des deux tiroirs.* ».
Soit $A2$ = « *Découvrir une pièce d'argent dans l'autre tiroir.* ».
Soit $O2$ = « *Découvrir une pièce d'or dans l'autre tiroir.* ».
Soit A = « *Choisir le meuble A.* ».
Soit B = « *Choisir le meuble B.* ».
On cherche $P(A/A1) = P(A\cap A1)/P(A1) =$

$$\frac{P(A1/A).P(A)}{P(A1/A).P(A)+P(A1/B).P(B)} = \frac{1.\frac{1}{2}}{1.\frac{1}{2}+\frac{1}{2}.\frac{1}{2}} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}.$$

b) *Quelle est la probabilité qu'il y ait une pièce d'argent dans l'autre tiroir ?*
On cherche $P(A2/A1) = P(A2/A\cap A1)/P(A/A1) + P(A2/B\cap A1)/P(B/A1) =$

$$1.\frac{2}{3}+0.\frac{1}{3}=\frac{2}{3}.$$

c) *Quelle est la probabilité qu'il y ait une pièce d'or dans l'autre tiroir ?*
On cherche $P(O2/A1) = P(O2/A\cap A1)/P(A/A1) + P(O2/B\cap A1)/P(B/A1) =$

$$0.\frac{2}{3}+1.\frac{1}{3}=\frac{1}{3}.$$

Exercice 3.6

On sait que globalement 98% des bébés survivent à l'accouchement. Cependant 15% des naissances se font par césarienne, dans ce dernier cas, le taux de survie des nouveaux nés est de 96.%.

a) Si, dans la population, une femme enceinte est choisie au hasard n'accouche pas par césarienne, quelle est la probabilité que son bébé survive ? (deux méthodes de solution).

b) si, dans la population des bébés nés ce jour et survivants, on en choisit un au hasard, quelle est la probabilité qu'il soit ne par césarienne ?

Solution

On sait que globalement 98% des bébés survivent à l'accouchement. Cependant 15% des naissances se font par césarienne, dans ce dernier cas, le taux de survie des nouveaux nés est de 96.%.

a) si, dans la population, une femme enceinte est choisie au hasard n'accouche pas par césarienne, quelle est la probabilité que son bébé survive ? (deux méthodes de solution).

Soit S = « *Le bébé survit.* » et C = «*L'accouchement se fait par césarienne.* ».

WWW.TALIB24.COM

$P(S) = 0{,}98$ $\quad P(\overline{S}) = 0{,}02.$ $\quad P(S/C) = 0{,}96.$

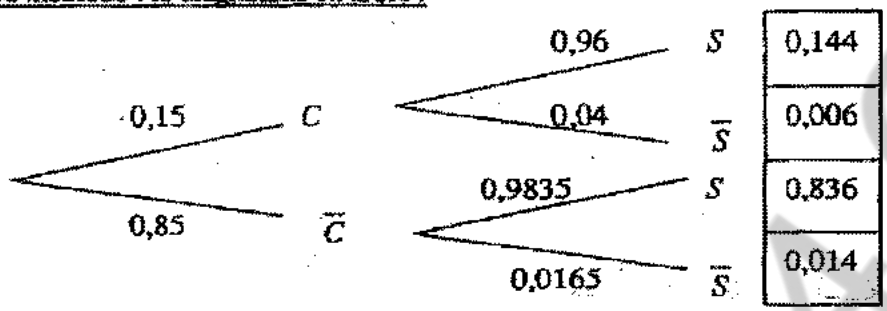
$P(C) = 0{,}15$ $\quad P(\overline{C}) = 0{,}85.$ $\quad P(\overline{S}/C) = 0{,}04.$

On cherche $P(S/\overline{C})$; or par la loi des probabilités totales :

$P(S) = P(S/C).P(C) + P(S/\overline{C}).P(\overline{C}) \Rightarrow 0{,}98 = 0{,}96 . 0{,}15 + P(S/\overline{C}) . 0{,}85.$

$$\Rightarrow P(S/\overline{C}) = \frac{0{,}98 - 0{,}96 \times 0{,}15}{0{,}85} = 0{,}9835.$$

<u>Autre méthode : le diagramme en arbre :</u>



Calcul des probabilités jointes et conditionnelles :

$P(C \cap S) = 0{,}144 = 0{,}15 * 0{,}96.$ $\quad P(C \cap \overline{S}) = 0{,}006 = 0{,}15 * 0{,}04.$

$P(\overline{C} \cap S) = 0{,}836 = 0{,}98 - 0{,}144.$ $\quad P(S/\overline{C}) = 0{,}9835 = 0{,}836/0{,}85.$

$P(\overline{S}/\overline{C}) = 0{,}0165 = 1 - 0{,}9835.$ $\quad P(\overline{C} \cap \overline{S}) = 0{,}014 = 0{,}85 * 0{,}0165.$

b) si, dans la population des bébés nés ce jour et survivants, on en choisit un au hasard, quelle est la probabilité qu'il soit ne par césarienne ?

On cherche $P(C/S) = \dfrac{P(S/C).P(C)}{P(S)} = \dfrac{0{,}96.0.15}{0{,}98} = 0{,}14694.$

Exercice 3.7

500 couples mariés dont chacun des deux conjoints travaillent ont vu leur déclaration d'impôt de l'an dernier analysée à des fins d'études statistiques. Dans le rapport de cette analyse, on trouve le tableau de contingence suivant :

Répartition des revenus des conjoints (nombre de couples)

	Mari	
Femme	< 25.000 Dh	≥ 25.000 Dh
< 25.000 Dh	212	198
≥ 25.000 Dh	36	54

donc dans cette population de 500 couples, il y a 54 couples ou a la fois la femme et le mari gagnent chacun au moins 25.000 Dh, etc.

A) si, dans la population de ces couples, on en choisit un au hasard, quelle est la probabilité que le mari gagne au moins 25.000 Dh ?

WWW.TALIB24.COM

B) si, dans la population de ces couples, on en choisit un au hasard, quelle est la probabilité conditionnelle que la femme gagne au moins 25.000 Dh sachant que son mari gagne au moins la même somme ?

C) si, dans la population de ces couples, on en choisit un au hasard, quelle est la probabilité conditionnelle que la femme gagne au moins 25.000 Dhs sachant que son mari gagne moins que cette somme ?

Solution

A) si, dans la population de ces couples, on en choisit un au hasard, quelle est la probabilité que le mari gagne au moins 25.000 € ?

il s'agit de la probabilité marginale que le mari gagne au moins 25.000 €, via l'approche fréquentiste, on peut établir quelle vaut (198 + 54)/ 500 = 252/500 = 0,504.

B) si, dans la population de ces couples, on en choisit un au hasard, quelle est la probabilité conditionnelle que la femme gagne au moins 25.000 € sachant que son mari gagne au moins la même somme ?

selon la meme approche, la probabilité s'établit à 54/(198 + 54) = 54/252 = 3/14 = 0,214.

C) si, dans la population de ces couples, on en choisit un au hasard, quelle est la probabilité conditionnelle que la femme gagne au moins 25.000 € sachant que son mari gagne moins que cette somme ?

selon la même approche, la probabilité s''tablit à 36/(212 + 36) = 36/248 = 9/62 = 0,145.

WWW.TALIB24.COM

SERIE 4 : PROBABILITE ET STATISTIQUE ** VARIABLES ALEATOIRES

EXERCICE 1--*Loi d'un dé truqué*----------

On considère un dé cubique truqué, de telle sorte que la probabilité d'obtenir la face numéro k est proportionnelle à k (on suppose que les faces sont numérotées de 1 à 6. Soit X la variable aléatoire associée au lancer de ce dé.

1. Déterminer la loi de probabilité de X, calculer son espérance.

2. On pose Y = 1/X. Déterminer la loi de Y, et son espérance et sa variance.

EXERCICE 2----------

Un garagiste dispose de deux voitures de location. Chacune est utilisable en moyenne 4 jours sur 5. Il loue les voitures avec une marge brute de 300 dh par jour et par voiture. On considère X la variable aléatoire égale au nombre de clients se présentant chaque jour pour louer une voiture. On suppose que $X(\Omega) = \{0, 1, 2, 3\}$ avec $P(X = 0) = 0,1$; $P(X = 1) = 0,3$; $P(X = 2) = 0,4$; $P(X = 3) = 0,2$.

1. On note Z le nombre de voitures disponibles par jour. Déterminer la loi de Z. On pourra considérer dans la suite que X et Y sont indépendantes.

2. On note Y la variable aléatoire : "nombre de clients satisfaits par jour". Déterminer la loi de Y

3. Calculer la marge moyenne par jour et la dispersion autour de cette moyenne .

EXERCICE 3----------

Les vaches laitières sont atteintes par une maladie M avec la probabilité $p = 0,15$. *Pour dépister la maladie M dans une étable de n vaches, on fait procéder à une analyse de lait. Deux méthodes sont possibles :*

Première méthode : On fait une analyse sur un échantillon de lait de chaque vache.

Deuxième méthode : On effectue d'abord une analyse sur un échantillon de lait provenant du mélange des n vaches. Si le résultat est positif, on effectue une nouvelle analyse, cette fois pour chaque vache.

On voudrait connaître la méthode la plus économique (= celle qui nécessite en moyenne le moins d'analyse). Pour cela, on note X_n *la variable aléatoire du nombre d'analyses réalisées dans la deuxième étape. On pose* $Y_n = X_n/n$.

1. Déterminer la loi de Y_n *et montrer que son espérance vaut :* $1 + 1/n - (0.85)^n$.

2. Etudier la fonction $f(x) = ax + \ln x$, *pour* $a = \ln(0,85)$. *Donner la liste des entiers n tels que* $f(n) > 0$.

3. Montrer que $f(n) > 0$ *équivaut à* $E(Y_n) < 1$. *En déduire la réponse (en fonction de n) à la question posée).*

EXERCICE 4----------

Une urne contient 2 boules noires et 8 boules blanches.

1°) Un joueur tire successivement 5 boules en remettant la boule dans l'urne après chaque tirage. Si il tire une boule blanche il gagne 2 points dans le cas contraire il perd trois points. Soit X le nombre de points obtenus par le joueur en une partie.

a) Déterminer la loi de probabilité de la v.a.r. X.

b) Calculer $E(X)$ et $V(X)$.

2°) Le joueur tire 5 boules simultanément, les 10 boules de l'urne étant numérotées de 1 à 10.

WWW.TALIB24.COM

a) Soit Y le plus grand des numéros tirés. Déterminer la loi de probabilité de Y et calculer E(Y).
b) Soit T le nombre de boules blanches obtenues. Après ce premier tirage le joueur remet les boules noires obtenues et effectue un nouveau tirage simultané de 5 boules.
On appelle Z le nombre de boules blanches obtenues lors de ce second tirage. Déterminer les lois de T et de Z. Calculer E(T).

EXERCICE 5

Deux urnes U_1 et U_2 contiennent chacune des boules blanches et des boules noires. U_1 contient 2 boules blanches et 2 boules noires, U_2 contient 1 boule blanche et 3 boules noires. On effectue une suite de tirages avec remise de la boule tirée en procédant comme suit :
Le premier tirage s'effectue dans U_1. Si au n-ième tirage on obtient une boule blanche alors le (n+1)-ème tirage s'effectue dans U_1. Si au n-ième tirage on obtient une boule noire alors le (n+1)-ème tirage s'effectue dans U_2.
On désigne par :
p_n la probabilité d'obtenir une boule blanche au n-ième tirage ;
X_n la variable aléatoire qui vaut 1 Si la boule obtenue au n-ième tirage est blanche, 0 sinon.
S_n est le nombre total de boules blanches obtenues au bout de n tirages.
1°) Calculer p_1, p_2.
2°) Déterminer une relation entre p_{n+1} et p_n ; en déduire l'expression de p_n en fonction de n, et la limite de p_n quand n tend vers $+\infty$.
3°) Pour n supérieur ou égal à 1, donner la loi de X_n. Préciser $E(X_n)$ et $V(X_n)$.
4°) Les variables aléatoires X_1 et X_2 sont-elles indépendantes ?
5°) Exprimer S_n en fonction des X_k, et de n ; En déduire E(Sn).

EXERCICE 6

On admettra que pour $|x| < 1$ et k dans $\mathbb{N}^*$:

$$\sum_{n=0}^{+\infty} \frac{(n+k)!}{n!} x^n = \frac{k!}{(1-x)^{k+1}}$$

Soit a un réel tel que $0 < a < 1$.
1°) Soit X une variable aléatoire à valeurs entières dont la loi est donnée par :
$\forall n \in \mathbb{N}$ $P(X = n) = (1-a)^n a$.

a) Vérifier qu'il s'agit bien d'une loi de probabilité.
b) Calculer l'espérance et la variance de X.

2°) Une urne contient des boules blanches et noires en proportion p et q avec $p + q = 1$.
On effectue des tirages avec remise : le nombre de tirages suit la loi de X. Y est la variable aléatoire égale au nombre de boule blanches obtenues.

a) Calculer pour tous les entiers k et n la probabilité conditionnelle $P(Y = k / X = n)$, ainsi que $P(Y = k \cap X = n)$.

b) En déduire la loi de Y et calculer l'espérance de Y.

WWW.TALIB24.COM

SERIE 4 : PROBABILITE ET STATISTIQUE ** VARIABLES ALEATOIRES

CORRIGE

EXERCICE 1

1. X prend ses valeurs dans $\{1, \ldots, 6\}$. Par hypothèse, il existe un réel a tel que $P(X = k) = ka$. Maintenant, puisque P_X est une loi de probabilité, on a :

$$\sum_{k=1}^{6} P(X = k) = 1 \iff a\frac{6 \times 7}{2} = 1 \implies a = 1/21.$$

On a donc :

k	1	2	3	4	5	6
$P(X = k)$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

On vérifie aisément en appliquant la formule que $E(X) = \frac{13}{3}$.

2. On a $Y = k \iff X = 1/k$. Y prend donc ses valeurs dans $\{1, 1/2, 1/3, 1/4, 1/5, 1/6\}$, et la loi est donnée par :

k	1/1	1/2	1/3	1/4	1/5	1/6
$P(Y = k)$	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

Le calcul de l'espérance n'est pas plus difficile, et donne :

$$E(Y) = \frac{2}{7}.$$

Attention à l'erreur suivante : ce n'est pas parce que $Y = 1/X$ que $E(Y) = 1/E(X)$!!!.

EXERCICE 2

1. Z est élément de $\{0, 1, 2\}$. On a :

$$P(Z = 2) = \frac{4}{5} \times \frac{4}{5} = \frac{16}{25}$$

(les deux voitures sont disponibles). D'autre part,

$$P(Z = 0) = \frac{1}{5} \times \frac{1}{5} = \frac{1}{25}$$

(les deux voitures sont simultanément indisponibles). Enfin, on obtient :

$$P(Z = 1) = 1 - P(Z = 0) - P(Z = 1) = \frac{8}{25}.$$

2. Remarquons que Y est à valeurs dans $\{0, 1, 2\}$. On calcule sa loi en utilisant la formule des probabilités totales. L'événement $Y = 0$ se produit si $X = 0$ ou bien si $X \geq 1$ et $Z = 0$. Ces deux événements étant disjoints, on a :

$$P(Y = 0) = P(X = 0) + P(X \geq 1 \cap Z = 0) = P(X = 0) + P(X \geq 1)P(Z = 0)$$

WWW.TALIB24.COM

(la disponibilité des voitures étant supposée indépendante de l'arrivée des clients). D'où :

$$P(Y=0) = 0,1 + 0,9 \times \left(\frac{1}{5}\right)^2 = 0,136.$$

De même, l'événement $Y = 1$ se produit si $X = 1$ et $Z \geq 1$ ou bien si $X \geq 2$ et $Z = 1$. On en déduit :

$$P(Y=1) = P(X=1)P(Z \geq 1) + P(X=2)P(Z=1) = 0,48.$$

Enfin, l'événement $Y = 2$ est réalisé si $X \geq 2$ et $Z = 2$. Ceci donne :

$$P(Y=2) = P(X \geq 2)P(Z=2) = 0,6 \times \left(\frac{4}{5}\right)^2 = 0,384.$$

3. La marge brute vaut $300Y$. La marge brute moyenne par jour est en euros :

$$E(300Y) = 300(0 \times 0,136 + 1 \times 0,48 + 2 \times 0,384) = 374,4.$$

EXERCICE 3

1. Y_n ne prend que deux valeurs, $1/n$ et $1 + 1/n$. On a en outre :

$$(Y_n = 1/n) \iff \text{aucune vache n'est malade}$$

d'où $P(Y_n = 1/n) = 0,85^n$. On en déduit - la loi de Y est une loi de probabilité - $P(Y = 1 + 1/n) = 1 - (0,85)^n$. Le calcul de l'espérance donne :

$$E(Y_n) = \frac{0,85^n}{n} + \frac{n+1}{n}(1 - 0,85^n) = 1 + \frac{1}{n} - 0,85^n.$$

2. f est dérivable sur $]0, +\infty[$, et $f'(x) = \frac{1+ax}{x}$. $f'(x)$ est donc du signe de $1 + ax$, ce qu permet de dire que f est croissante sur $]0, -1/a[$, et décroissante ensuite. La limite de en $+\infty$ est $-\infty$, il en est de même en 0. En calculant les valeurs successives de $f(n)$, o a $f(17) > 0,07$ et $f(18) < -0,03$. 17 est donc la plus grande valeur entière pour laquell $f(n)$ est positive. En outre, $f(1) < 0$ alors que $f(2) > 0$. L'ensemble d'entiers recherch est donc $\{2, \ldots, 17\}$.

3. On a :

$$\begin{aligned} E(Y_n) < 1 &\iff 1 + \frac{1}{n} - 0,85^n < 1 \\ &\iff 0,85^n > \frac{1}{n} \\ &\iff n \ln(0,85) > -\ln n. \end{aligned}$$

Par suite, $E(Y_n) < 1 \iff f(n) > 0$. L'étude précédente montre que les entiers pour lesquels $f(n) > 0$ est $\{2, \ldots, 17\}$. On a intérêt à choisir la deuxième méthode si, seulement si, il y a de 2 à 17 vaches dans l'étable !

WWW.TALIB24.COM

EXERCICE 4

1) (a) La variable aléatoire X est régie par une loi binomiale. En effet,

- 5 expériences identiques et indépendantes (car les tirages se font avec remise) sont effectuées.
- chaque expérience a deux issues : « la boule tirée est blanche » avec une probabilité $p = \frac{2}{10} = \frac{1}{5}$ et « la boule tirée n'est pas blanche » avec une probabilité $1 - p = \frac{4}{5}$.

La variable aléatoire X est régie par une loi binomiale de paramètres $n = 5$ et $p = \frac{1}{5}$. On sait alors que

$$X(\Omega) = [\![0,5]\!] \text{ et } \forall k \in [\![0,5]\!],\ p(X = k) = \binom{5}{k}\left(\frac{1}{5}\right)^k \cdot \left(\frac{4}{5}\right)^{5-k}.$$

Plus explicitement,

- $p(X = 0) = \left(\frac{4}{5}\right)^5 = \frac{1024}{3125} = 0,32768.$
- $p(X = 1) = 5 \times \frac{1}{5} \times \left(\frac{4}{5}\right)^4 = \frac{256}{625} = 0,4096.$
- $p(X = 2) = 10 \times \left(\frac{1}{5}\right)^2 \times \left(\frac{4}{5}\right)^3 = \frac{128}{625} = 0,2048.$
- $p(X = 3) = 10 \times \left(\frac{1}{5}\right)^3 \times \left(\frac{4}{5}\right)^2 = \frac{32}{625} = 0,0512.$
- $p(X = 4) = 5 \times \left(\frac{1}{5}\right)^4 \times \frac{4}{5} = \frac{4}{625} = 0,0064.$
- $p(X = 5) = \left(\frac{1}{5}\right)^5 = \frac{1}{3125} = 0,00032.$

L'espérance de X est $E(X) = np = 5 \times \frac{1}{5} = 1$ et la variance de X est $V(X) = np(1-p) = 5 \times \frac{1}{5} \times \frac{4}{5} = \frac{4}{5} = 0,8.$

(b) $Y = 2X - 3(5 - X) = 5X - 15$. Par suite, $Y(\Omega) = \{5k - 15,\ k \in [\![0,5]\!]\} = \{-15, -10, -5, 0, 5, 10\}$. Ensuite,

$$\forall k \in [\![0,5]\!],\ p(Y = 5k - 15) = p(X = k) = \binom{5}{k}\left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{5-k}.$$

Ensuite, $E(Y) = 5E(X) - 15 = 5 \times 1 - 15 = -10$ et $V(Y) = V(5X - 15) = 5^2 V(X) = 25 \times \frac{4}{5} = 20.$

2) (a) $X(\Omega) = \{0, 1, 2\}$. La loi de probabilité ne change pas si on suppose les tirages simultanés.

Le nombre de tirages simultanés de 5 boules parmi 10 est $\binom{10}{5}$.

Soit $k \in [\![0,2]\!]$. Au cours d'un tirage de 5 boules, on obtient k boules blanches si et seulement si on tire k boules parmi les 2 blanches et $5 - k$ boules parmi les 8 noires. Il y a donc $\binom{2}{k} \times \binom{8}{5-k}$ tirages où on obtient k boules blanches. Donc,

$$\forall k \in [\![0,2]\!],\ p(X = k) = \frac{\binom{2}{k} \times \binom{8}{5-k}}{\binom{10}{5}}.$$

Plus explicitement,

- $p(X = 0) = \dfrac{\dfrac{8 \times 7 \times 6 \times 5 \times 4}{5 \times 4 \times 3 \times 2}}{\dfrac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2}} = \dfrac{5 \times 4}{10 \times 9} = \dfrac{2}{9}.$
- $p(X = 1) = \dfrac{2 \times \dfrac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2}}{\dfrac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2}} = \dfrac{2 \times 5 \times 5}{10 \times 9} = \dfrac{5}{9}.$

WWW.TALIB24.COM

$$\bullet\ p(X=2)=\frac{\dfrac{8\times 7\times 6}{3\times 2}}{\dfrac{10\times 9\times 8\times 7\times 6}{5\times 4\times 3\times 2}}=\frac{5\times 4}{10\times 9}=\frac{2}{9}.$$

L'espérance de X est $E(X)=0\times\frac{2}{9}+1\times\frac{5}{9}+2\times\frac{2}{9}=1$ et la variance de X est

$$V(X)=E\left(X^2\right)-\left(E(X)\right)^2=0\times\frac{2}{9}+1\times\frac{5}{9}+4\times\frac{2}{9}-1^2=\frac{4}{9}.$$

(b) Comme à la question 1). $Y(\Omega)=\{5k-15,\ k\in[\![0,2]\!]\}=\{-15,-10,-5\}$ et

$$\forall k\in[\![0,2]\!],\ p(Y=5k-15)=p(X=k)=\frac{\binom{2}{k}\times\binom{8}{5-k}}{\binom{10}{5}}.$$

Ensuite, $E(Y)=5E(X)-15=-10$ et $V(Y)=5^2E(X)=\frac{100}{9}$.

WWW.TALIB24.COM

SÉRIE 5 : LOIS CLASSIQUES USUELLES

ÉNONCÉ 1

On considère le jeu suivant : le joueur lance d'abord un dé non truqué. S'il obtient 1, 2 ou 3, il gagne l'équivalent en dhs (c'est-à-dire 1 e s'il obtient 1, par exemple). Sinon, il perd 2 dhs. On note X la variable aléatoire correspondant au gain du joueur (négatif en cas de perte).

1.- Donnez la loi de X et sa fonction de répartition F_X.

2.- Calculez l'espérance de X ainsi que sa variance.

On modifie le jeu de la façon suivante : les gains restent les mêmes pour les résultats 1, 2 ou 3, mais si le joueur obtient autre chose, il relance le dé. S'il obtient 3 ou moins, il gagne 3 dhs, sinon il perd 5 dhs.

3.- Décrivez formellement l'univers du nouveau jeu.

4.- Donnez la loi de Y et calculez son espérance.

5.- Quelle variante du jeu est la plus avantageuse pour le joueur ?

ÉNONCÉ 2

Pour améliorer la sûreté de fonctionnement d'un serveur informatique, on envisage d'introduire de la redondance, c'est-à-dire d'avoir plusieurs exemplaires des composants importants. On peut réaliser les opérations suivantes :

- *On utilise trois alimentations de 300W chacune : le serveur peut continuer à tourner avec une alimentation en panne : il consomme au maximum 500W.*
- *On place les quatre disques durs en configuration RAID 5 : le serveur peut continuer à tourner avec un disque dur en panne.*

On suppose que la probabilité de panne d'une alimentation est p et que celle d'une panne de disque dur est q. On suppose en outre que tous les composants sont indépendants.

1.- Soit un serveur avec alimentations redondantes : calculer la probabilité de panne du serveur en supposant qu'aucun autre composant que les alimentations ne peut tomber en panne.

2.- Soit un serveur avec disques durs RAID 5 : calculer la probabilité de panne du serveur en supposant qu'aucun autre composant que les disques dur ne peut tomber en panne.

3.- Si p = q, quelle solution de redondance est la plus intéressante ?

Cet exercice consiste avant tout en la modélisation d'un problème réel au moyen d'outils probabilistes classiques. Il faut donc reconnaître dans l'énoncé les hypothèses qui permettent de se ramener à des modèles connus.

ÉNONCÉ 3

Un couple de variables aléatoires (X; Y) tel que $X(\Omega)=\{-2;0;1\}$ et $Y(\Omega)=\{-1;1;2\}$ dont la loi jointe est donnée par le tableau suivant :

$\mathbf{P}(X=x, Y=y)$	$y=-1$	$y=1$	$y=2$
$x=-2$	0,2	0,2	α
$x=0$	0,1	0,1	0,05
$x=1$	0,2	0	0,1

1.- Donner l'unique valeur possible pour α en justifiant brièvement la réponse.

2.- Calculer les lois marginales de X et de Y .

3.- Montrer que X et Y ne sont pas indépendantes.

WWW.TAHEB24.COM

4.- Calculer la loi conditionnelle de X sachant Y = 1. En déduire E(X/Y=1).
5.- Calculer l'espérance conditionnelle de X sachant que Y ≠ 2.
6.- Calculer E(X.Y) et en déduire COV(X;Y).
7.- On pose Z = X + Y. Calculer la loi de Z, puis E(Z) et V (Z).

ÉNONCÉ 4

Soit la fonction f définie par :

$$f(x) = \begin{cases} ax^2 + b & \text{si } x \in [0,1] \\ 0 & \text{sinon} \end{cases}$$

1.- À quelles conditions sur a et b la fonction f est-elle la densité d'une variable aléatoire continue ?
2.- On suppose que a et b vérifient les conditions déterminées à la question précédente. Soit X une variable aléatoire de densité f.
On suppose que P(X ≥ 0.5) = 1/8. En déduire les valeurs de a et de b.
3.- Calculer E(X) et V (X) en utilisant les valeurs de a et b obtenues précédement.

ÉNONCÉ 5

On suppose que la durée de vie d'un disque dur est une loi exponentielle. Le fabricant veut garantir que le disque dur a une probabilité inférieure à 0.001 de tomber en panne sur un an.
Quelle durée de vie moyenne minimale doit avoir le disque dur ?

ÉNONCÉ 6

D'après une étude récente, la taille des femmes marocaines est distribuée selon une loi normale de moyenne m = 1,58 et d'écart-type σ = 0;06. Pour produire un stock de vêtements, un fabricant souhaite utiliser cette loi.
1.- Il commence par déterminer un intervalle de la forme [m-a , m+a] (donc symétrique autour de la moyenne) contenant en moyenne 90 % (environ) des tailles des femmes marocaines : calculer a.
2.- Il en déduit trois tailles, S, M et L, correspondant respectivement aux intervalles [m-a , m-a/3], [m-a/3 , m+a/3] et [m+a/3 , m+a].
Calculer le pourcentage de la production qui doit être affecté à chaque taille.

ÉNONCÉ 7

Soient a et p des réels strictement positifs et $f_{a,p}$ la fonction définie par :

$$f_{a,p}(x) = \frac{pa^p}{x^{p+1}} \mathbb{1}_{]a,+\infty[}(x).$$

a) Vérifier que $f_{a,p}$ est la densité d'une variable aléatoire X et déterminer, lorsqu'elles existent, son espérance et sa variance.
Soit Z une variable aléatoire dont la loi conditionnelle sachant (X = x) est, pour tout x > a, la loi uniforme sur]0, x[.
b) Déterminer la densité du couple (X,Z) (On représentera en particulier le domaine sur lequel elle n'est pas nulle).
c) Déterminer et reconnaître la loi conditionnelle de X sachant (Z = z) en distinguant les cas z ∈]0, a[et z ∈]a,+1[. En déduire E[X/Z=z] puis E[X/Z].
d) Calculer COV(X,Z) et déterminer la régression linéaire de Z en X.

WWW.FMPB24.COM

SERIE 5 : PROBABILITE ET STATISTIQUE ** : VARIABLES ALEATOIRES
CORRIGE

EXERCICE 1

Correction (P1)

1) L'expérience aléatoire est un lancer de dé dont l'univers est donc

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

muni d'une probabilité uniforme (car le dé n'est pas truqué). La variable aléatoire X est donnée par le tableau suivant :

ω	1	2	3	4	5	6
$X(\omega)$	1	2	3	-2	-2	-2

On en déduit donc que $X(\Omega) = \{-2, 1, 2, 3\}$. La loi de X est donnée par $\mathbb{P}(\{X(\omega) = x\}) = \mathbb{P}(X^{-1}(\{x\}))$ pour les $x \in X(\Omega)$. En utilisant le fait que la probabilité $\mathbb{P}$ est uniforme, on obtient les résultats résumés dans le tableau suivant, dont la dernière ligne donne la loi de X :

x	-2	1	2	3
$X^{-1}(\{x\})$	$\{4,5,6\}$	$\{1\}$	$\{2\}$	$\{3\}$
$\mathbb{P}_X(x)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Une fois la loi de X déterminée, les questions suivantes sont essentiellement des questions de cours, associées à la réalisation de quelques calculs.

La fonction de répartition F_X définie par $F_X(x) = \mathbb{P}(X \leq x)$ est alors

$$F_X(x) = \begin{cases} 0 & \text{pour } x < -2, \\ \frac{1}{2} & \text{pour } -2 \leq x < 1, \\ \frac{2}{3} & \text{pour } 1 \leq x < 2, \\ \frac{5}{6} & \text{pour } 2 \leq x < 3, \\ 1 & \text{pour } x \geq 3. \end{cases}$$

2) Pour calculer $\mathbb{E}(X)$, on applique la définition de l'espérance, à savoir

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x \mathbb{P}_X(x).$$

On obtient ainsi

$$\mathbb{E}(X) = -2 \times \frac{1}{2} + 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} = 0.$$

Pour calculer $V(X)$, on utilise $V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. On a

$$\mathbb{E}(X^2) = (-2)^2 \times \frac{1}{2} + 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} = \frac{13}{3},$$

et donc $V(X) = \frac{13}{3} \simeq 4{,}33$.

WWW.TALIB24.COM

La deuxième partie de l'exercice est plus délicate car l'expérience ne semple pas symétrique. La solution la plus simple pour l'analyser est de faire comme si elle était symétrique : on considère deux lancers de dé, le deuxième étant inutile pour la définition de Y dans certaines situations. Cette approche s'apparente à celle utilisée pour faciliter le traitement de certains problèmes dans lesquels on fait comme si les objets étudiés (des jetons par exemple) étaient discernables alors qu'ils ne le sont pas.

3) On choisit comme univers celui d'une expérience dans laquelle on lance toujours deux fois le dé. On a donc

$$\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\},$$

muni de la probabilité uniforme.

La variable aléatoire Y est alors donnée par le tableau suivant en fonction du résultat de l'expérience $\omega = (u,v)$:

(u,v)	1	2	3	4	5	6
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	3	3	3	-5	-5	-5
5	3	3	3	-5	-5	-5
6	3	3	3	-5	-5	-5

4) On constate que $Y(\Omega) = \{-5,1,2,3\}$. Pour tout $y \in Y(\Omega)$, on doit déterminer $Y^{-1}(\{y\})$ pour calculer la probabilité de cet événement et ainsi la loi de Y. D'après le tableau, on a

$$\begin{aligned}
Y^{-1}(\{-5\}) &= \{4,5,6\} \times \{4,5,6\},\\
Y^{-1}(\{1\}) &= \{1\} \times \{1,2,3,4,5,6\},\\
Y^{-1}(\{2\}) &= \{2\} \times \{1,2,3,4,5,6\},\\
Y^{-1}(\{3\}) &= (\{3\} \times \{1,2,3,4,5,6\}) \cup (\{4,5,6\} \times \{1,2,3\}).
\end{aligned}$$

En utilisant le fait que la probabilité sur Ω est uniforme, on obtient le tableau suivant qui résume la loi de Y :

y	-5	1	2	3
$\mathbb{P}_Y(y)$	$\frac{9}{36} = \frac{1}{4}$	$\frac{6}{36} = \frac{1}{6}$	$\frac{6}{36} = \frac{1}{6}$	$\frac{6}{36} + \frac{9}{36} = \frac{5}{12}$

Comme dans la première partie de l'exercice, une fois la loi de Y déterminée, le reste n'est que calcul...

5) On a

$$\mathbb{E}(Y) = -5 \times \frac{1}{4} + 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{5}{12} = \frac{1}{2}.$$

Comme $\mathbb{E}(Y) > \mathbb{E}(X)$, on gagne en moyenne plus à la deuxième variante qui est donc plus avantageuse pour le joueur.

EXERCICE 2

Cet exercice consiste avant tout en la modélisation d'un problème réel au moyen d'outils probabilistes classiques. Il faut donc reconnaître dans l'énoncé les hypothèses qui permettent de se ramener à des modèles connus.

Correction

1) Comme un composant possède ici deux états (en panne ou pas), on considère cet état comme une variable de Bernoulli, l'état de panne correspondant au « succès » de la variable.

WWW.TALEB24.COM

Donc, l'état de l'alimentation i est représenté par une variable A_i distribuée selon $B(p)$ une variable aléatoire Bernoulli qui vaut 1 si l'alimentation est en panne.

Cette modélisation est très fréquente et très classique. À chaque fois qu'on rencontre une grandeur aléatoire qui peut prendre deux valeurs, on est en fait en présence d'une variable de Bernoulli. La valeur à représenter par 1 dépend de ce qu'on veut faire dans la suite. Ici, on doit compter le nombre de composants en panne pour savoir si l'ordinateur continue de fonctionner. Il est donc naturel de représenter par 1 la panne plutôt que le bon fonctionnement.

L'alimentation globale est constituée de trois alimentations A_1, A_2 et A_3. Comme les alimentations sont des variables de Bernoulli indépendantes de même paramètre p, le nombre d'alimentation en panne, qui est la somme $A_1 + A_2 + A_3$ est une variable de loi Binomiale $B(3, p)$.

De nouveau, nous utilisons ici une modélisation très classique. Pour justifier l'utilisation d'une loi Binomiale, il faut rappeler les propriétés importantes suivantes : on compte les succès de variables de Bernoulli *indépendantes* et de *même paramètre*. Si les variables ne sont pas indépendantes ou si elles ont des paramètres différents, le résultat n'est pas une loi Binomiale. Notons aussi que compter les succès ou faire la somme des variables est exactement la même chose et qu'on peut donc utiliser la justification la plus adaptée au contexte.

Le serveur est en panne si deux au moins des alimentations sont en panne. On note A l'évènement « panne de l'alimentation globale ». On a donc

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A_1 + A_2 + A_3 \geq 2), \\
&= \mathbb{P}(A_1 + A_2 + A_3 = 2) + \mathbb{P}(A_1 + A_2 + A_3 = 3), \\
&= C_3^2 p^2 (1-p) + C_3^3 p^3, \\
&= p^2(3 - 2p),
\end{aligned}$$

en utilisant les propriétés de la loi Binomiale.

La deuxième question est exactement la même que la première mais avec des paramètres numériques légèrement différents...

Comme pour les alimentations, on modélise les disques durs par des variables D_j de Bernoulli $B(q)$, indépendantes. Le nombre de disques en panne est alors une variable de loi Binomiale $B(4, q)$. Le serveur est en panne

WWW.FALIBA.COM

si au moins deux des disques sont en panne, l'évènement correspondant D étant de probabilité

$$\begin{aligned}
\mathbb{P}(D) &= \mathbb{P}\left(\sum_{j=1}^{4} D_j \geq 2\right), \\
&= \mathbb{P}\left(\sum_{j=1}^{4} D_j = 2\right) + \mathbb{P}\left(\sum_{j=1}^{4} D_j = 3\right) + \mathbb{P}\left(\sum_{j=1}^{4} D_j = 4\right) \\
&= C_4^2 q^2 (1-q)^2 + C_4^3 q^3 (1-q) + C_4^4 q^4, \\
&= q^2(6(1-q)^2 + 4q(1-q) + q^2), \\
&= q^2(3q^2 - 8q + 6).
\end{aligned}$$

Pour choisir la meilleure solution, on compare les probabilités de panne en supposant que $p = q$. On a ainsi

$$\begin{aligned}
\mathbb{P}(A) - \mathbb{P}(D) &= p^2(3-2p) - p^2(3p^2 - 8p + 6), \\
&= p^2(-3p^2 + 6p - 3).
\end{aligned}$$

Le signe de cette différence est alors celui de $f(p) = -3p^2 + 6p - 3$. La dérivée de f est $f'(p) = 6(1-p)$ qui s'annule en $p = 1$ et qui est donc positive sur l'intervalle $[0,1]$. La fonction est donc croissante sur cet intervalle. Or, $f(1) = 0$ et $f(0) = -3$. Donc, sur l'intervalle $[0,1]$, $f(p) \leq 0$. La probabilité de défaillance des alimentations est donc toujours inférieure à celle des disques durs et il est donc plus intéressante d'utiliser une redondance sur les alimentations.

On pourrait considérer que l'approche choisie pour évaluer l'intérêt de la redondance est un peu naïve. En effet, un ordinateur classique comprend nécessairement une alimentation et un stockage pérenne de type disque dur. De ce fait, on devrait normalement étudier l'ordinateur comme un ensemble comportant un sous-système redondant (par exemple les disques durs) et un sous-système non redondant (ici l'alimentation), puis calculer la probabilité de panne et comparer les deux solutions.

Il s'avère en fait que l'approche proposée dans l'exercice donne un résultat identique à celui de l'approche plus complexe. Considérons le cas d'un ordinateur avec une alimentation redondante et un unique disque dur. On s'intéresse à l'évènement « l'ordinateur ne tombe pas en panne », ce qui s'écrit $\overline{A} \cap \overline{D}$ avec les notations de la correction. En supposant les sous-systèmes indépendants, on a

$$\mathbb{P}(\overline{A} \cap \overline{D}) = \mathbb{P}(\overline{A})\mathbb{P}(\overline{D}) = (1 - \mathbb{P}(A))(1 - \mathbb{P}(D)).$$

En utilisant les résultats de la correction pour A et les hypothèses pour D, on a donc

$$\mathbb{P}(\overline{A} \cap \overline{D}) = (1 - p^2(3-2p))(1-q).$$

Un raisonnement similaire conduit la probabilité suivant dans le cas d'un serveur aux disques durs redondants et à l'alimentation unique :

$$\mathbb{P}(\overline{A} \cap \overline{D}) = (1-p)(1 - q^2(3q^2 - 8q + 6)).$$

Si on suppose maintenant que $p = q$, on voit apparaître dans les deux expressions un facteur $(1-p)$ commun qu'on peut donc supprimer dans l'analyse. On se retrouve ainsi à comparer $\mathbb{P}(\overline{A})$ et $\mathbb{P}(\overline{D})$ dans le cas de redondance pour les deux sous-systèmes. Or, il s'agit de la comparaison des probabilités des complémentaires des évènements étudiés dans la correction. On obtient donc exactement les mêmes conclusion que dans cette correction : il est plus intéressant de rendre les alimentations redondantes.

WWW.TAHAR24.COM

EXERCICE 3

Correction

1) On sait que $\mathbb{P}(X \in X(\Omega), Y \in Y(\Omega)) = 1$ et que

$$\mathbb{P}(X \in X(\Omega), Y \in Y(\Omega)) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y).$$

Donc la somme des probabilités contenues dans le tableau doit être égale à 1. On constate que la somme est ici $0{,}95 + \alpha$. Donc $\alpha = 0{,}05$.

La justification proposée est assez détaillée. On peut se contenter de dire que la somme des probabilités doit être égale à 1, puisque c'est une propriété majeure des tableaux utilisés pour donner les lois de couple de variables aléatoires discrètes.

2) La loi marginale de X est donnée par

$$\mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y).$$

Un simple calcul donne les résultats résumés dans le tableau suivant

x	-2	0	1
$\mathbb{P}(X = x)$	0,45	0,25	0,3

De la même façon, on obtient la loi de Y donnée par

y	-1	1	2
$\mathbb{P}(Y = y)$	0,5	0,3	0,2

On se contente ici de justifier les calculs dans un des deux cas, par symétrie. Il est utile de vérifier que la somme des probabilités de la loi obtenue pour chaque variable est bien 1.

3) Les variables X et Y sont indépendantes si et seulement si pour tout x et y, $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$. Or, on remarque dans le tableau que $\mathbb{P}(X = 1, Y = 1) = 0$, alors que $\mathbb{P}(X = 1) \times \mathbb{P}(Y = 1) = 0{,}3 \times 0{,}3 \neq 0$. Les deux variables ne sont donc pas indépendantes.

Pour les questions d'indépendance, on doit toujours rappeler la définition puis essayer de l'appliquer. Quand on cherche comme ici à montrer que les variables ne sont pas indépendantes, il suffit de trouver un contre-exemple, c'est-à-dire une valeur pour chaque variable telles que la probabilité du couple soit différente du produit des probabilités de chaque variable. Si on cherche à montrer l'indépendance, on doit au contraire vérifier que toutes les paires de valeurs satisfont l'égalité demandée.

4) On utilise la définition de la loi conditionnelle. $\mathbb{P}(X = x | Y = 1) = \frac{\mathbb{P}(X=x, Y=1)}{\mathbb{P}(Y=1)}$, ce qui conduit immédiatement à la loi résumée dans le tableau

WWW.MLIB24.COM

suivant

x	-2	0	1
$\mathbb{P}(X=x\|Y=1)$	$\frac{0,2}{0,3}=\frac{2}{3}$	$\frac{0,1}{0,3}=\frac{1}{3}$	$\frac{0}{0,3}=0$

L'espérance conditionnelle est obtenue en appliquant la définition c'est-à-dire

$$\begin{aligned}\mathbb{E}(X|Y=1) &= \sum_{x\in X(\Omega)} x\mathbb{P}(X=x|Y=1),\\ &= -2\times\frac{2}{3}+0\times\frac{1}{3}+1\times 0,\\ &= -\frac{4}{3}.\end{aligned}$$

On se contente ici d'appliquer les définitions et de faire des calculs élémentaires. La question suivante est légèrement plus complexe puisqu'il faut considérer un conditionnement qui ne s'exprime pas sous la forme $Y=y$.

5) On doit d'abord calculer $\mathbb{P}(X=x|Y\neq 2)=\frac{\mathbb{P}(X=x,Y\neq 2)}{\mathbb{P}(Y\neq 2)}$. Or, $\{Y\neq 2\}=\{Y=-1\}\cup\{Y=1\}$, avec une union disjointe. Donc

$$\mathbb{P}(X=x,Y\neq 2)=\mathbb{P}(X=x,Y=-1)+\mathbb{P}(X=x,Y=1),$$

ainsi que $\mathbb{P}(Y\neq 2)=\mathbb{P}(Y=-1)+\mathbb{P}(Y=1)$, c'est-à-dire

$$\mathbb{P}(X=x|Y\neq 2)=\frac{\mathbb{P}(X=x,Y=-1)+\mathbb{P}(X=x,Y=1)}{\mathbb{P}(Y=-1)+\mathbb{P}(Y=1)}.$$

En appliquant cette formule, on obtient la loi conditionnelle résumée dans le tableau suivant

x	-2	0	1
$\mathbb{P}(X=x\|Y\neq 2)$	$\frac{0,4}{0,8}=\frac{1}{2}$	$\frac{0,2}{0,8}=\frac{1}{4}$	$\frac{0,2}{0,8}=\frac{1}{4}$

On en déduit l'espérance conditionnelle souhaitée par le calcul suivant

$$\begin{aligned}\mathbb{E}(X|Y\neq 2) &= \sum_{x\in X(\Omega)} x\mathbb{P}(X=x|Y\neq 2),\\ &= -2\times\frac{1}{2}+0\times\frac{1}{4}+1\times\frac{1}{4}= -3/4\end{aligned}$$

$\mathbb{P}(X=x|Y\neq 2)$ est égale à la probabilité $1-\mathbb{P}(X=x|Y=2)$. Or, c'est en général faux. Dans le cas présent, si on considère par exemple $x=-2$, on a

$$\begin{aligned}1-\mathbb{P}(X=-2|Y=2) &= 1-\frac{\mathbb{P}(X=-2,Y=2)}{\mathbb{P}(Y=2)},\\ &= 1-\frac{0,05}{0,2},\\ &= \frac{3}{4},\end{aligned}$$

alors que $\mathbb{P}(X=-2|Y\neq 2)=\frac{1}{2}$. En fait, on a bien

$$1-\mathbb{P}(X=x|Y=2)=\mathbb{P}(X\neq x|Y=2),$$

mais ce n'est pas utile pour calculer $\mathbb{P}(X=-2|Y\neq 2)$!

6) On calcule $\mathbb{E}(XY)$ en appliquant la propriété classique : pour toute fonction ϕ, on a

$$\mathbb{E}(\phi(X,Y))=\sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)}\phi(x,y)\mathbb{P}(X=x,Y=y).$$

Le tableau suivant donne les calculs intermédiaires $xy\mathbb{P}(X=x,Y=y)$:

$\mathbb{P}(X=x,Y=y)$	$y=-1$	$y=1$	$y=2$
$x=-2$	$0,2\times 2=0,4$	$0,2\times -2=-0,4$	$0,05\times -4=-0,2$
$x=0$	$0,1\times 0=0$	$0,1\times 0=0$	$0,05\times 0=0$
$x=1$	$0,2\times -1=-0,2$	$0\times 1=0$	$0,1\times 2=0,2$

On obtient ainsi $\mathbb{E}(XY)=-0,2$.

WWW.TALIB24.COM

Comme dans la plupart des questions de calcul, il est important [illegible] définition de ce qu'on calcule et les propriétés utilisées, comme nous l'avons fait ci-dessus. On procède de la même façon pour la suite de la question, en rappelant la définition de la covariance et de l'espérance d'une variable seule.

On sait que $Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. On calcule donc $\mathbb{E}(X)$ et $\mathbb{E}(Y)$ en utilisant les définitions de ces espérances :

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x \in X(\Omega)} x\mathbb{P}(X = x)\\ &= -2 \times 0{,}45 + 0 \times 0{,}25 + 1 \times 0{,}3 = -0{,}6\end{aligned}$$

De la même façon

$$\begin{aligned}\mathbb{E}(Y) &= -1 \times 0{,}5 + 1 \times 0{,}3 + 2 \times 0{,}2\\ &= 0{,}2.\end{aligned}$$

On obtient finalement $Cov(X,Y) = -0{,}2 - (-0{,}6) \times 0{,}2 = -0{,}08$.

7) Etudions maintenant $Z = X + Y$. Le tableau suivant donne les valeurs prises par Z en fonction de celles de X et Y :

$z = x + y$	$y = -1$	$y = 1$	$y = 2$
$x = -2$	-3	-1	0
$x = 0$	-1	1	2
$x = 1$	0	2	3

On a donc $Z(\Omega) = \{-3, -1, 0, 1, 2, 3\}$. Pour calculer la loi de $Z = f(X,Y)$, on doit déterminer $f^{-1}(z)$ pour tout $z \in Z(\Omega)$ puis calculer la probabilité de $f^{-1}(z)$ à partir de la loi jointe de X et Y, en utilisant la propriété $\mathbb{P}(Z = z) = \mathbb{P}((X,Y) \in f^{-1}(z))$. On obtient les résultats suivants :

$$\begin{aligned}\mathbb{P}(Z = -3) &= \mathbb{P}(X = -2, Y = -1) = 0{,}2\\ \mathbb{P}(Z = -1) &= \mathbb{P}((X,Y) \in \{(-2,1),(0,-1)\}) = 0{,}1 + 0{,}2 = 0{,}3\\ \mathbb{P}(Z = 0) &= \mathbb{P}((X,Y) \in \{(-2,0),(1,-1)\}) = 0{,}05 + 0{,}2 = 0{,}25\\ \mathbb{P}(Z = 1) &= \mathbb{P}(X = 0, Y = 1) = 0{,}1\\ \mathbb{P}(Z = 2) &= \mathbb{P}((X,Y) \in \{(0,2),(1,1)\}) = 0{,}05 + 0 = 0{,}05\\ \mathbb{P}(Z = 3) &= \mathbb{P}(X = 1, Y = 2) = 0{,}1\end{aligned}$$

On utilise la propriété classique qui caractérise la loi d'une variable définie comme fonction d'autres variables à partir de la fonction réciproque (ici f^{-1}) et de la loi jointe des autres variables.

Comme $Z = X + Y$ et que l'espérance est linéaire, on a

$$\mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y) = -0{,}6 + 0{,}2 = -0{,}4.$$

De plus, on sait que

$$V(X + Y) = V(X) + V(Y) + 2Cov(X,Y).$$

On calcule donc les variances de X et Y. On a

$$\begin{aligned}\mathbb{E}(X^2) &= (-2)^2 \times 0{,}45 + 0^2 \times 0{,}25 + 1^2 \times 0{,}3\\ &= 2{,}1.\end{aligned}$$

et

$$\begin{aligned}\mathbb{E}(Y^2) &= (-1)^2 \times 0{,}5 + 1^2 \times 0{,}3 + 2^2 \times 0{,}2\\ &= 1{,}8.\end{aligned}$$

Donc $V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = 1{,}74$ et $V(Y) = 1{,}56$. On obtient finalement

$$V(Z) = 1{,}74 + 1{,}56 + 2 \times (-0{,}08) = 3{,}14.$$

Notons que l'espérance et la variance de Z qui sont obtenues ici au moyen de propriétés de ces deux grandeurs pourraient aussi être calculées directement à partir de la loi de Z. Quelle que soit l'option choisie, il faut rappeler les propriétés et définitions utilisées.

WWW.TALIB24.COM

EXERCICE 4

1)

Correction

Pour que f soit une densité, il faut que pour tout x, $f(x) \geq 0$ et que $\int_{-\infty}^{\infty} f(x)dx = 1$. On commence par la première condition. Comme f est nulle en dehors de l'intervalle $[0,1]$, il suffit de considérer le cas $x \in [0,1]$ pour lequel on doit avoir $ax^2 + b \geq 0$.

Pour les deux conditions, on peut en général se focaliser uniquement sur les intervalles pour lesquels la fonction proposée n'est pas nulle.

On remarque que $f(0) = b$, et il faut donc que $b \geq 0$. Si $a = 0$, il n'y a pas d'autre condition nécessaire pour assurer la positivité.

La dérivée de f est donnée par $f'(x) = 2ax$. Sur l'intervalle $[0,1]$, le signe de f' est donc celui de a. Si $a > 0$, f est croissante sur $[0,1]$ et on a donc $f(x) \geq b$ pour tout $x \in [0,1]$. De ce fait, la condition $b \geq 0$ est encore suffisante pour assurer $f(x) \geq 0$. Si $a < 0$, f est décroissante sur $[0,1]$, et donc a donc $f(x) \geq f(1) = a + b$ pour tout $x \in [0,1]$. On doit donc avoir $a + b \geq 0$.

En résumé si $b \geq 0$ et si $a \geq -b$, alors f est positive (le cas $a \geq -b$ inclus en particulier le cas $a \geq 0$).

Calculons maintenant $\int_{-\infty}^{\infty} f(x)dx = 1$. D'après les propriétés classiques de l'intégrale, on a

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{\infty} f(x)dx.$$

Comme f est nulle en dehors de $[0,1]$, on a donc

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} f(x)dx.$$

Il faut toujours justifier (brièvement) la réduction de l'intervalle d'intégration. On peut cependant se contenter d'écrire directement la deuxième égalité, avec la justification qui la précède.

La fonction $x \to ax^2 + b$ admet comme primitive possible la fonction $x \mapsto \frac{a}{3}x^3 + bx$. On a donc

$$\int_0^1 ax^2 + b\,dx = \left[\frac{a}{3}x^3 + bx\right]_0^1 = \frac{a}{3} + b.$$

On doit donc avoir $\frac{a}{3} + b = 1$ (en plus de $b \geq 0$ et si $a \geq -b$). Comme $b = 1 - \frac{a}{3}$, on a doit donc avoir $1 - \frac{a}{3} \geq 0$, soit $a \leq 3$. De plus $a + b \geq 0$ devient $a + 1 - \frac{a}{3} \geq 0$, soit $a \geq -\frac{3}{2}$.

Finalement, f est une densité si et seulement si $a \in \left[-\frac{3}{2}, 3\right]$ et $b = 1 - \frac{a}{3}$.

2) Il n'y a pas de difficulté majeure dans l'analyse qui précède, mais il faut tout de même être attentif dans le calcul de la primitive (il faut bien faire la primitive deux termes de la somme $ax^2 + b$ et ne pas écrire seulement $\frac{1}{3}ax^3 + b$, par exemple). De même, l'analyse des conditions sur a et b doit être réalisée avec soin afin de ne pas oublier une condition lors des transformations. Ici, on a bien conservé la condition qui donne une intégrale de 1, puis on a utiliser les deux conditions d'inégalités pour obtenir deux bornes sur a. On a donc conservé trois conditions, comme il le fallait.

On sait que $\mathbb{P}(X \geq 0,5) = \int_{0,5}^{\infty} f(x)dx$. Comme f est nulle en dehors de

WWW.TALIB24.COM

$[0,1]$, et en utilisant la primitive déterminée à la question précédente, on a

$$\begin{aligned}
\mathbb{P}(X \geq 0,5) &= \int_{0,5} f(x)dx, \\
&= \int_{0,5}^{1} f(x)dx, \\
&= \left[\frac{a}{2}x^2 + bx\right]_{0,5}^{1}, \\
&= \frac{a}{2} + b - \frac{a}{2} \times \frac{1}{8} - b \times \frac{1}{2}, \\
&= 1 - \frac{a}{24} - \frac{b}{2}.
\end{aligned}$$

Une erreur courante dans l'analyse précédente consiste à confondre densité f et fonction de répartition F. On sait en effet que, par définition, si F_X est la fonction de répartition de X, $\mathbb{P}(X \geq 0,5) = 1 - F_X(0,5)$ (on a utilisé la continuité de X pour s'affranchir de problèmes aux bornes de intervalles). Si on confond f et F, on peut être tenté d'écrire (ce qui est faux !) que $\mathbb{P}(X \geq 0,5) = 1 - f(0,5) = 1 - \frac{a}{4} - b$. On obtient alors un résultat totalement différent dans la suite (et, insistons sur ce point, faux).

En utilisant les résultats de la première question, on a donc deux équations à deux inconnues :

$$\begin{aligned}
\frac{a}{2} + b &= 1, \\
1 - \frac{a}{24} - \frac{b}{2} &= \frac{1}{8}.
\end{aligned}$$

En utilisant $b = 1 - \frac{a}{2}$ dans la deuxième équation, on obtient

$$1 - \frac{a}{24} - \frac{1}{2} + \frac{a}{8} = \frac{1}{8},$$

ce qui donne $a = 2$, puis $b = 0$ (et donc $f(x) = 3x^2$ sur $[0,1]$). On remarque que les valeurs de a et b sont compatibles avec les inégalités établies à la question 1 et donc qu'on obtient bien une densité.

La dernière vérification est très importante. On pourrait en effet supposer par exemple que $\mathbb{P}(X \geq 0,5) = \frac{1}{16}$. Des calculs similaires à ceux effectués ci-dessus donnent $a = \frac{7}{6}$. Or, cette valeur n'est pas acceptable d'après les contraintes obtenues à la question 1. L'hypothèse $\mathbb{P}(X \geq 0,5) = \frac{1}{16}$ entre donc en contradiction avec l'hypothèse sur la densité de X. Ce n'est pas le cas avec l'hypothèse $\mathbb{P}(X \geq 0,5) = \frac{1}{8}$!

3) On sait que $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)dx$, ce qui donne ici (en tenant compte de la nullité de f en dehors de $[0,1]$) :

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^1 x \times 3x^2 dx, \\
&= \int_0^1 3x^3 dx, \\
&= \left[\frac{3}{4}x^4\right]_0^1 = \frac{3}{4}.
\end{aligned}$$

WWW.FLIB24.COM

en utilisant le fait que $x \mapsto \frac{3}{4}x^4$ est une primitive de $x \mapsto 3x^3$.

Pour calculer $V(X)$, on utilise la relation $V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Comme $\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$, on a

$$\begin{aligned}\mathbb{E}(X^2) &= \int_0^1 3x^4 dx, \\ &= \left[\frac{3}{5}x^5\right]_0^1 = \frac{3}{5}.\end{aligned}$$

en utilisant le fait que $x \mapsto \frac{3}{5}x^5$ est une primitive de $x \mapsto 3x^4$. On a donc

$$V(X) = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}.$$

EXERCICE 5

Correction

On appelle D la variable aléatoire donnant la durée de vie du disque dur. D suit une loi exponentielle de paramètre λ. Le fabricant veut garantir que

$$\mathbb{P}(D \leq 1) \leq 0.001.$$

Comme $\mathbb{P}(D \leq a) = F_D(a)$ par définition, on obtient l'inégalité

$$1 - \exp(-\lambda \times 1) \leq 0.001,$$

en appliquant la formule classique pour la fonction de répartition d'une variable de loi exponentielle. On a alors

$$\begin{aligned}1 - \exp(-\lambda) &\leq 0.001, \\ 0.999 &\leq \exp(-\lambda), \\ \ln(0.999) &\leq -\lambda, && \text{en utilisant la croissance de ln} \\ \lambda &\leq -\ln(0.999), \\ \frac{-1}{\ln(0.999)} &\leq \frac{1}{\lambda}, && \text{en utilisant la décroissance de } \frac{1}{x} \\ 999.5 &\leq \frac{1}{\lambda}.\end{aligned}$$

Or, comme D suit une loi exponentielle, son espérance est $\frac{1}{\lambda}$. La durée de vie moyenne du disque dur doit donc être d'au moins 999.5 ans !

EXERCICE 6

Correction

1) Soit T la variable aléatoire représentant la taille d'une femme. Par hypothèse, T suit une loi normale $\mathcal{N}(1.58; 0.06^2)$. On cherche $a > 0$ tel que

$$\mathbb{P}(T \in [m - a, m + a]) = 0.9.$$

Dans cet exercice, il n'y pas d'étape de modélisation car la loi de la variable d'intérêt est donnée explicitement, avec la valeur de ses paramètres. La traduction de l'énoncé se limite donc à exprimer mathématiquement l'évènement étudié et à donner les hypothèses sur cet évènement.

WWW.TALIB24.COM

Soit la variable $Y = \frac{T-m}{\sigma}$. On sait que Y suit une loi normale standard $\mathcal{N}(0;1^2)$. De plus, on a

$$\begin{array}{rcl} m-a \leq & T & \leq m+a, \\ -a \leq & T-m & \leq a, \\ -\frac{a}{\sigma} \leq & \frac{T-m}{\sigma} & \leq \frac{a}{\sigma}. \end{array}$$

Donc $\mathbb{P}(T \in [m-a, m+a]) = 0{,}9$ est équivalent à $\mathbb{P}(Y \in [-\frac{a}{\sigma}, \frac{a}{\sigma}]) = 0{,}9$. Cherchons donc λ tel que $\mathbb{P}(Y \in [-\lambda, \lambda]) = 0{,}9$.

On utilise ci-dessus la manipulation classique permettant de se ramener à une variable aléatoire distribuée selon une loi normale standard pour laquelle on dispose d'une table. La technique consiste à appliquer à l'evenement défini sur la variable d'origine (ici T) les transformations qui conduisent à la variable centrée et réduite (ici Y). On transforme ainsi l'évènement sur T en un évènement sur Y pour lequel on pourra appliquer la table.

On sait que

$$\mathbb{P}(Y \in [-\lambda, \lambda]) = F_Y(\lambda) - F_Y(-\lambda),$$

car Y est une variable aléatoire continue. De plus, par symétrie de la loi normale standard, on $F_Y(-\lambda) = 1 - F_Y(\lambda)$, et ainsi

$$\mathbb{P}(Y \in [-\lambda, \lambda]) = 2F_Y(\lambda) - 1.$$

De ce fait, chercher λ tel que $\mathbb{P}(Y \in [-\lambda, \lambda]) = 0{,}9$ est équivalent à chercher λ tel que $F_Y(\lambda) = \frac{1+0{,}9}{2} = 0{,}95$.

La table ne donne que la fonction de répartition de la loi normale. Il faut donc se ramener à une équation portant sur cette fonction. On utilise d'abord le fait que pour toute variable aléatoire continue (c'est faux pour une variable discrète), on a

$$\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a).$$

Ensuite, on applique la propriété de symétrie de la loi normale pour se débarrasser du terme négatif (le terme $F_Y(-\lambda)$).

La lecture de la table de la loi normale donne :

$$\begin{aligned} F_Y(1{,}64) &= 0{,}9495, \\ F_Y(1{,}65) &= 0{,}9505. \end{aligned}$$

Pour avoir un intervalle légèrement plus grand que celui recherché par le fabricant, on choisit $\lambda = 1{,}65$. Si on pose $a = \sigma\lambda = 0{,}06 \times 1{,}65 = 0{,}099$, on a donc

$$\mathbb{P}(T \in [m-a, m+a]) = \mathbb{P}(T \in [1{,}481, 1{,}679]) \simeq 0{,}9$$

Il faut être attentif à l'inversion de la relation dans le calcul final de a. On a en effet trouvé λ grâce à la table, mais ce n'est pas la grandeur recherchée puisque l'évènement étudié est $Y \in [-\frac{a}{\sigma}, \frac{a}{\sigma}]$ alors qu'on ajusté la probabilité de $Y \in [-\lambda, \lambda]$. On doit donc faire $\frac{a}{\sigma} = \lambda$, ce qui donne a.

Le reste de l'exercice est une application de la même stratégie de centrage et de réduction de T, mais avec une utilisation « inverse » de la table : l'évènement étudié est fixé et on doit calculer sa probabilité. On a intérêt à conserver la formule $a = \sigma\lambda$ pour simplifier les calculs.

WWW.TALIB24.COM

Correction

2) Etudions le premier intervalle. On a

$$\begin{array}{rcl} m-a \leq & T & \leq m-\frac{a}{3}, \\ -a \leq & T-m & \leq -\frac{a}{3}, \\ -\frac{a}{\sigma} \leq & \frac{T-m}{\sigma} & \leq \frac{a}{3\sigma}, \\ -\lambda \leq & Y & \leq -\frac{\lambda}{3}. \end{array}$$

et donc

$$\begin{aligned} \mathbb{P}\left(T \in \left[m-a, m-\frac{a}{3}\right]\right) &= \mathbb{P}\left(Y \in \left[-\lambda, -\frac{\lambda}{3}\right]\right), \\ &= F_Y\left(-\frac{\lambda}{3}\right) - F_Y(-\lambda), \\ &= 1 - F_Y\left(\frac{\lambda}{3}\right) - 1 + F_Y(\lambda), && \text{par symétrie} \\ &= 0{,}9505 - F_Y\left(\frac{1{,}65}{3}\right), && \text{selon l'analyse précédente} \\ &= 0{,}9505 - 0{,}7088, && \text{par lecture dans la table} \\ &= 0{,}2417. \end{aligned}$$

On a de la même façon

$$\begin{aligned} \mathbb{P}\left(T \in \left[m-\frac{a}{3}, m+\frac{a}{3}\right]\right) &= \mathbb{P}\left(Y \in \left[-\frac{\lambda}{3}, \frac{\lambda}{3}\right]\right), \\ &= F_Y\left(\frac{\lambda}{3}\right) - F_Y\left(-\frac{\lambda}{3}\right), \\ &= 2F_Y\left(\frac{\lambda}{3}\right) - 1, \\ &= 2 \times 0{,}7088 - 1, \\ &= 0{,}4176. \end{aligned}$$

Enfin :

$$\begin{aligned} \mathbb{P}\left(T \in \left[m+\frac{a}{3}, m+a\right]\right) &= \mathbb{P}\left(Y \in \left[\frac{\lambda}{3}, \lambda\right]\right), \\ &= F_Y(\lambda) - F_Y\left(\frac{\lambda}{3}\right), \\ &= 0{,}9505 - 0{,}7088, \\ &= 0{,}2417. \end{aligned}$$

3) ce dernier résultat étant évident par symétrie de la loi normale autour de sa moyenne.

On calcule enfin les pourcentages à partir de ces probabilités. La production totale correspond à 90% de la population et on doit donc diviser les probabilités obtenues par cette valeur. On obtient alors

$$\begin{array}{lll} \% \text{ de S} & = \frac{0{,}2417}{0{,}90} & \simeq 27\% \\ \% \text{ de M} & = \frac{0{,}4176}{0{,}90} & \simeq 46\% \\ \% \text{ de L} & = \frac{0{,}2417}{0{,}90} & \simeq 27\% \end{array}$$

L'exercice se termine par un piège classique : comme le fabricant se contente de cibler 90 % de la population, les probabilités d'obtenir les différentes tailles ne donnent pas directement les pourcentages de la production. On doit passer par un ratio pour obtenir ces pourcentages.

WWW.TALIB24.COM