HEIG–VD
Corrigé Etudiant

Probabilités et Statistique
2021-2022

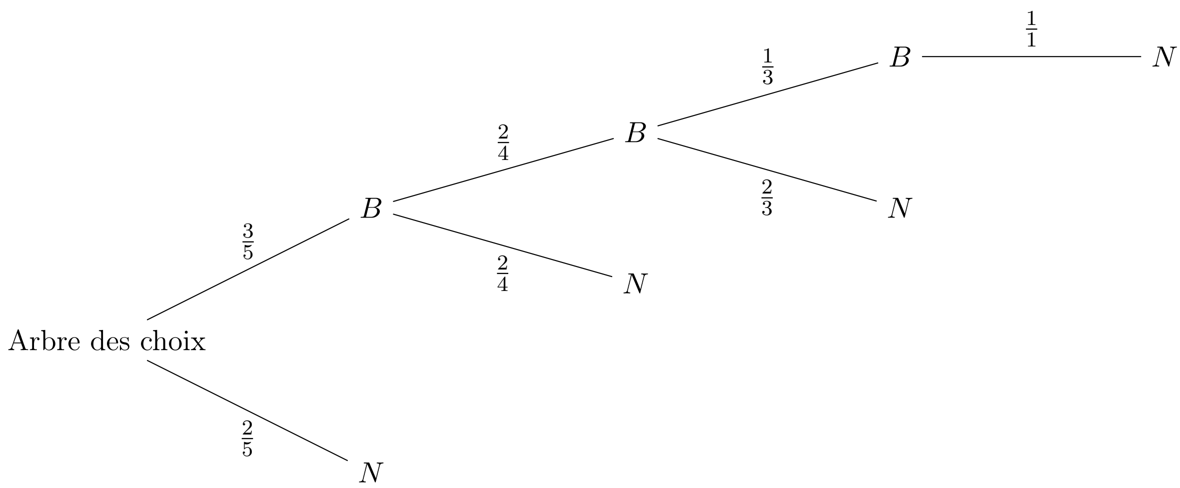
Thème : variables aléatoires — Solution de la Série 3

Exercice 1

Donnée: *Une boîte contient deux billes noires et trois billes blanches. On tire une bille à la fois, sans remise, jusqu'à ce que la première bille noire apparaisse. Dès que la bille noire est tirée, aucune autre bille est extraite de la boîte. Désignons par X le nombre de tirage nécessaires. En utilisant un arbre binaire,*

a) *Déterminer l'ensemble fondamental Ω de l'expérience aléatoire à l'aide des événements :*

— N Une bille noire est tirée
— B Une bille blanche est tirée



Nous arrivons à l'arbre si dessus. En effet avec cet arbre, nous avons toutes les possibilités que peut prendre X la probabilité est notée sur chaque branche. Ainsi lors du premier tirage, nous avons intuitivement $\frac{3}{5}$ "chances" de tirer une boule blanche, et nous avons une probabilité de tiré une boule noir égal à $\frac{2}{5}$

b) *Calculer la loi de probabilité de X*

X = xi	1	2	3	4
P(X=xi)	$\frac{2}{5}$	$\frac{3}{5} \cdot \frac{1}{2}$	$\frac{3}{5} \cdot \frac{1}{2} \cdot \frac{2}{3}$	$\frac{3}{5} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{1}$

On suit simplement la solution à travers l'arbre en multipliant les probabilités

Exercice 2

Donnée: *Des signaux binaires (0 ou 1) sont transmis d'un point A à un point B par un canal de communication. Cependant, une erreur de transmission peut toujours résulter de perturbations aléatoires agissant sur le canal. Supposons qu'un signal émis en A soit correctement enregistré en B avec probabilité 0.8 et ce* *indépendamment d'un signal à l'autre. Pour essayer d'améliorer la qualité de transmission, on se propose d'émettre chaque signal trois fois de suite, i.e. 000 au lieu de 0 et 111 au lieu de 1.*

a) *Déterminer les réceptions possibles par émission par block de trois signaux identiques*

pour le signal 1, on envoie 111. ce dernier peut être faux sur n'importe quel bit donc on a comme réception possible : 000, 001, 010, 011, 100, 101, 110, 111

b) *Supposons que le signal 0 a été émis en A par le block 000 et définisson par X la variable aléatoire qui indique le nombre d'erreurs possibles à la récéption du block. Déterminer les valeurs que peut prendre la variable aléatoire X*

On a envoyé 000, nous voulons savoir combien il peut y avoir d'erreur sur ce message. Chaque bit peut etre faux (un 1 à la place du 0). Nous avons donc : les signaux 100, 010 et 001 qui comptabilise une erreur. Nous avons les signaux 110, 101 et 011 qui comptabilisent 2 erreurs, 111 qui comptabilise 3 erreurs et 000 qui n'a pas d'erreur

La valeur aléatoire peut donc prendre les valeurs 0,1,2 et 3

c) *Calculer la loi de probabilité de X*

Pour trouver la loi de proba, nous devons trouver la probabilité que X prenne la valeur x. X suit une lois binomiale, en effet elle compte le nombre de succès (chapitre 6 slide 6) parmis n épreuves. La lois de proba est $P(X = k) = \binom{n}{k} \cdot p^x(1-p)^{n-x}$ Nous avons donc 3 épreuves (une pour chaque bit), avec une proba de réussir de 0.8 pour chacune

$X = x$	0	1	2	3
$P(X = x)$	$\binom{3}{0}(0.8)^3$=0.512	$\binom{3}{1}0.8^2 0.2^1$=0.384	$\binom{3}{2}0.8^1 0.2^2 = 0.096$	$\binom{3}{3}0.8^0 0.2^3 = 0.008$

Intuitivement, pour le cas $X = 1$ nous devons avoir 1 erreur et deux correctes. Nous voulons donc l'intersection de ces trois évènements, comme ils sont indépendants, nous multiplions leur probabilités, cependant il faut prendre en compte qu'il y a trois évènements qui donne 1 erreur et 2 correctes puisque l'erreur peut être sur n'importe quel bit. Nous devons donc mutliplier cette proba par trois (le binom $\binom{3}{1}$ représente exactement la multiplicité d'avoir 1 faux et deux justes)

d) *Le décodage à la réception se réalise suivant le principe majoritaire. Par exemple, 010 est décodé en 0. Calculer la probabilité d'erreur pour le signal 0 émis par le bloc 000. Est-elle inférieure à 0.2, probabilité sans répétition du signal ? La qualité de transmission est-elle amélioré*

Nous devons avoir au moins 2 erreur sur notre transmission pour que l'interpretation sois fausse donc

$$P(X \geq 2) = P(X = 2) + P(X = 3) = 0.096 + 0.008 = 0.104$$

puisque X est discrète, en choisissant $x = 2$ et $x = 3$ nous avons toutes les valeurs que X peut prendre donc nous respectons $P(X \geq 2)$

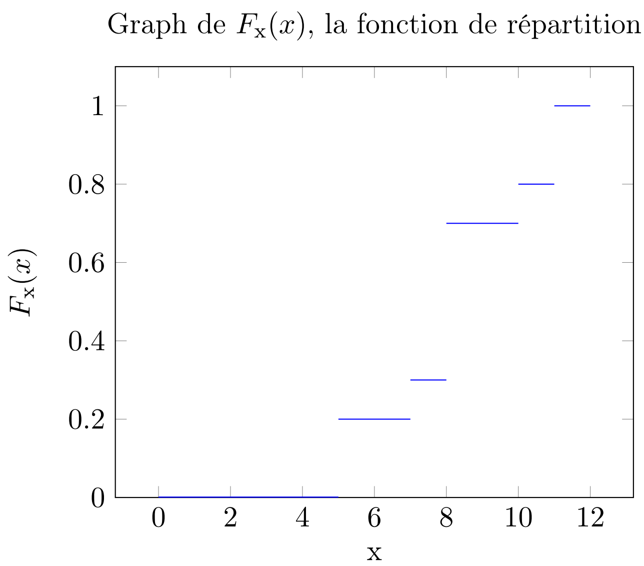
Notre probabilité de 0.104 est plus petite que la probabilité sans répétition (0.2 puisque un signal a une proba de 0.8 d'arriver juste) Le signal est donc amélioré.

Exercice 3

Donnée: *La fonction de répartition d'une variable aléatoire X est*

$$F_{\mathrm{x}}(x) = \begin{cases} 0 & \text{si } x < 5 \\ 0.2 & \text{si } 5 \leq x < 7 \\ 0.3 & \text{si } 7 \leq x < 8 \\ 0.7 & \text{si } 8 \leq x < 10 \\ 0.8 & \text{si } 10 \leq x < 11 \\ 1 & \text{si } x \geq 11 \end{cases}$$

a) *Représenter garaphiquement la fonction de répartition de* X



b) *Déterminer les réalisations de la variable aléatoire* X *ainsi que sa loi de probabilité*

Pour rappel nous avons

$$F_{\mathrm{x}}(x) = P(X \leq x) \tag{1}$$

. Autrement dit nous avons

$$F_{\mathrm{x}}(x) = \sum_{x_i \leq x} P(X = x_i)$$

Faisons un exemple avec $x = 7$

$$\begin{aligned} F_x(7) &= \sum_{x_i \leq 7} P(X = x_i) \\ &= P(X = 5) + P(X = 7) \\ P(X = 7) &= F_x(7) - P(X = 5) \\ &= 0.3 - 0.2 \\ &= 0.1 \end{aligned}$$

$X = x$	5	7	8	10	11
$P(X = x)$	0.2	0.1	0.4	0.1	0.2

c) *Calculer les probabilité* $P(X \leq 6.5)$, $P(8 < X \leq 10)$,$P(8 \leq X \leq 10)$,$P(X > 10)$,

Nous avons $F_x(X \leq 6.5)$ puisque (1) et donc

$$\begin{aligned} F_x(X \leq 6.5) &= \underbrace{P(X \leq 5)}_{0} + \underbrace{P(5 < X \leq 6.5)}_{0.2} \\ &= 0.2 \end{aligned}$$

d) *Sans calcul, déterminer l'espérance de X*

Par définition l'espérance

$$\mathbb{E}(X) \text{ est égal à } \sum_{i=1}^{n} x_i \cdot P(X = x_i)$$

Mais on peut la trouver facilement en reprenant la lois de probabilité de b), en effet on peut voir que les valeur sont symétrique autour de 8, avec les mêmes probabilités Donc on aura la moyenne pondérée égale à 8

Exercice 4

Donnée: *Imaginons qu'on souhaite transmettre les réalisations* $x_1, x_2, ..., x_n$ *d'une variable aléatoire discrète X d'un point d'observation A à un point de réception B à l'aide d'un canal de communication ne pouvant transférer que des 0 ou des 1. Ainsi, les valeurs prises par X devront être codées en chaînes formées uniquement de 0 et de 1 avant d'être transmises. Pour éviter toute ambiguïté, on exige qu'un code ne puisse pas être une extension d'un autre. Comme exemple, supposons que les réalisations de X sont* x_1, x_2, x_3, x_4. *Comme code possible, on peut envisager*

$$x_1 \leftrightarrow 00, x_2 \leftrightarrow 01, x_3 \leftrightarrow 10, x_4 \leftrightarrow 11 \tag{1}$$

Ainsi, si X prend la valeur x_1, le message envoyé en B sera 00, il vaudra 01 si X $= x_2$ et ainsi de suite. Un autre code possible est

$$x_1 \leftrightarrow 0, x_2 \leftrightarrow 10, x_3 \leftrightarrow 110, x_4 \leftrightarrow 111 \tag{2}$$

En revanche le codage

$$x_1 \leftrightarrow 0, x_2 \leftrightarrow 1, x_3 \leftrightarrow 00, x_4 \leftrightarrow 01 \tag{3}$$

n'est pas admis étant donné que x_3 et x_4 sont des extensions de x_1 Un objectif du codage consiste tout naturellement à minimiser le nombre espéré de bits nécessaires pour transmettre l'information. Ainsi, un code est dit plus efficace qu'un autre si son nombre espéré de bits est plus petit que celui nécessaire à l'autre code.

a) *Supposons qe la lois de probabilité est*

$X = x$	x_1	x_2	x_3	x_4
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

et considérons les codes données ci-dessus par (1) et (2) Pour cette distribution de X quel est le code le plus efficace ?

$$\mathbb{E}_{nbBits}(\text{code1}) = 2 * \frac{1}{2} + 2 * \frac{1}{4} + 2 * \frac{1}{8} + 2 * \frac{1}{8} = 2$$

Donc nous pouvons nous attendre à avoir besoins de 2 bits pour envoyé le x_i en utilisant le code (1) ce qui fait sens, puisque nous codons toutes les réalisations de X avec deux bits

$$\mathbb{E}_{nbBits}(\text{code2}) = 1 * \frac{1}{2} + 2 * \frac{1}{4} + 3 * \frac{1}{8} + 3 * \frac{1}{8} = 1.75$$

Si nous utilisons le code 2 cependant, nous avons en moyenne besoins de 1.75 bits pour coder les réalisations x_i de X

b) *Considérons en toute généralité une variable aléatoire discrète X prenant ses valeurs dans l'ensemble $\{x_1, x_2, ..., x_n\}$ avec probabilité correspondantes $p_1, p_2 ... p_n$. La grandeur*

$$H(X) = -\sum_{i=i}^{n} p_i \cdot \log_2(p_i)$$

est appelée, en théorie de l'information, entropie de la variable aléatoire X. Par convention, si l'une des probabilités p_i est nul on pose $0 \cdot log_x(0) = 0$. L'entropie représente en quelque sorte la quantité d'incertitude relative à X. Autrement dit, on considère H(X) comme l'information lié à l'overvation de X. Caluler l'entropie de X selon la distribution donné en a)

Appliquons la définition :

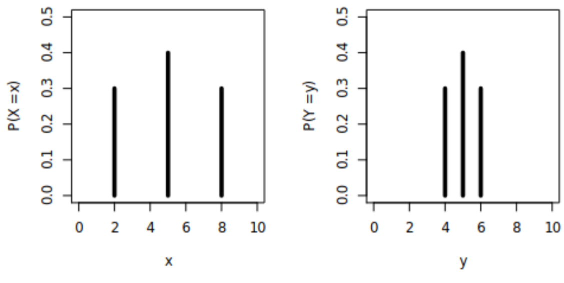
$$\begin{aligned} H(X) &= -\sum_{i=i}^{n} p_i \cdot \log_2(p_i) && (1) \\ &= \frac{1}{2} \cdot log_2(\frac{1}{2}) + \frac{1}{4} \cdot log_2(\frac{1}{4}) + \frac{1}{8} \cdot log_2(\frac{1}{8}) + \frac{1}{8} \cdot log_2(\frac{1}{2}) && (2) \\ &= 1.75 && (3) \end{aligned}$$

c) *Selon le théorème du codage sans bruit, tout codage nécessite un nombre espéré de bits au minimum égal à l'entropie de X*

Le code (2) est donc un code qui permet de transmettre l'information de la manière la plus optimale possible

Exercice 5

Donnée: *Considérons deux variables aléatoires X et Y dont les histogrammes se trouvent dans la figure ci-dessous.*



En n'effectuant aucun calcul, l'écart-type de X est-il plus grand que celui de Y? L'écart type est plus grand dans la figure x car les données sont plus écartées que dans la figure y

Exercice 6

Donnée: *La distribution de Benford prétend que pour des mesures d'une certaine quantité physique, la probabilité de tomber sur un 1 est de l'ordre de 30 % alors que celle de tomber sur un 9 est de l'ordre de 4.6 %. Cette distribution n'est applicable que pour un grand nombre d'objets mesurables. Elle s'applique par exemple aux longueurs de tous les fleuves du monde, aux prix des articles d'un supermarché ou encore aux volumes des tables de logarithmes d'une encyclopédie. En effet, les premières pages de chaque volume sont* *toujours plus usées que les suivantes, celles du premier volume étant plus utilisées que celles du second et ainsi de suite. Il en sera de même pour les nombres d'une comptabilité. Ainsi, pour détecter d'éventuelles fraudes dans les déclarations d'impôts de grandes entreprises, le fisc américain a utilisé la distribution de Benford. En effet, dans une comptabilité de grande envergure, qui bien souvent est formée de nombreuses pages de chiffres, les comptables essayaient pour déjouer le fisc de créer des données truquées en tirant des nombres au hasard. L'idée n'était pas très bonne. En se livrant à une petite analyse statistique à l'aide de la distribution de Benford, il était possible d'identifier des fraudes dans les déclarations d'impôts de grandes sociétés. La distribution de probabilités de Benford se trouve dans la table ci-dessous.*

valeur	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
probabilité	*0.301*	*0.176*	*0.125*	*0.097*	*0.079*	*0.067*	*0.058*	*0.051*	*0.046*

a) *calculer l'espérance mathématique d'une variable aléatoire X issue d'une distribution de Benford*

L'espérance est la moyenne pondérée càd définie comme

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \cdot p_i$$

avec p_i la probabilité que la variable aléatoire prenne la valeur x_i

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{n} x_i \cdot p_i \\ &= 1 \cdot 0.301 + 2 \cdot 0.176 + 3 \cdot 0.125 + 4 \cdot 0.097 + 5 \cdot 0.079 + 6 \cdot 0.067 + 7 \cdot 0.058 + 8 \cdot 0.051 + 9 \cdot 0.046 \\ &= 3.441\end{aligned}$$

b) *Determiner la probabilité que $P(X < 7 | X \geq 4)$*

Nous pouvons réécrire la probabilité conditionel de la manière suivante

$$\begin{aligned}P(X < 7 | X \geq 4) =& \frac{P(X < 7 \cap X \geq 4)}{P(X \geq 4)} \\ =& \frac{0.097 + 0.079 + 0.067}{0.097 + 0.079 + 0.067 + 0.058 + 0.051 + 0.046}\end{aligned}$$