

TDN°1: Variables aléatoires & Information

CORRIGÉ

A- Variables aléatoires

- 1- On lance deux fois un dé à six faces. Soit la variable aléatoire discrète X représentant le nombre de 6 obtenus. Déterminer la fonction de répartition de F_X de la variable X .

Solution :

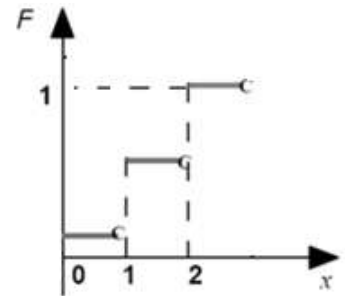
Tous les résultats possibles de l'expérience: $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$

L'ensemble des valeurs possibles de X est $\{0,1,2\}$

$N = 2^2 = 4 \rightarrow$ (2 résultats, 2 tirages). Nous savons que $\sum_{i=1} P_i = 1$

Nombre de 6	$P(X = x_i)$	F_X
0	1/4	1/4
1	2/4	3/4
2	1/4	1

$$F_X(x) = \sum_{i=1} P(X \leq x_i) \rightarrow P \text{ accumulative}$$



- 2- Considérons l'expérience aléatoire qui consiste à lancer trois pièces de monnaie. On introduit une variable aléatoire X « nombre de piles »
- Déterminer tous les résultats possibles de l'expérience
 - Déterminer les probabilités P de la variable aléatoire X
 - Tracer la fonction $P(X = x_i)$
 - Tracer la fonction répartition F_X

Solution :

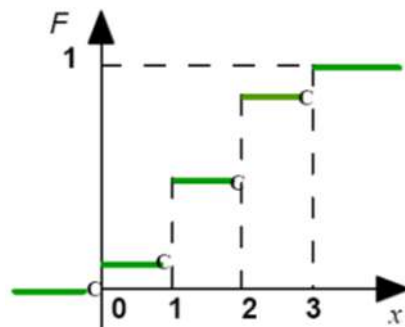
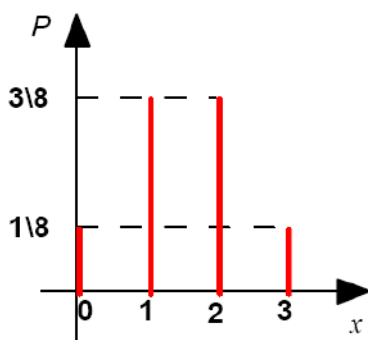
- a- $\Omega = \{(F, F, F), (F, F, P), (F, P, F), (P, F, F), (F, P, P), (P, F, P), (P, P, F), (P, P, P)\}$

- b- $N = 2^3 = 8 \rightarrow$ (3 pièces, 2 possibilités)

Pour déterminer la probabilité d'occurrence de la variable aléatoire X , nous relierons à chaque résultat le nombre de piles : par exemple 0 pile $\rightarrow$ 1 seule combinaison, 1 pile $\rightarrow$ 3 combinaisons possible

- c- Diagramme $P(X)$: vérifiant que $\sum_{i=1} P_i = 1$
- d- Fonction répartition F_X : $F_X(x) = \sum_{i=1} P(X \leq x_i)$

Nombre de piles	$P(X = x_i)$	F_X
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1



3- On a placé dans une urne cinq boules indiscernables: trois noires et deux blanches. On tire au hasard une à une toutes les boules de cette urne, et on appelle R le rang de la première boule blanche tirée.

- a- Déterminer les probabilité de R .
- b- Déterminer la fonction de répartition de la variable aléatoire R
- c- Calculer l'espérance de R

Solution :

a- Les résultats possibles du tirage sont : $\Omega = \{(1), (2), (3), (4)\}$

$N = 5 \rightarrow$ (5 boules)

Voici les possibilités

- Le rang vaut 1 si on tire une blanche au premier tirage soit 2 chances sur 5
- Le rang vaut 2 si on tire une noire au premier tirage puis une blanche au second tirage soit 3 chances sur 5 pour la boule noire et 2/4 pour une boule blanche, etc...

x_i	1	2	3	4
P_i	$\frac{2}{5}$ = 0.4	$\frac{3}{5} \times \frac{2}{4}$ = 0.3	$\frac{3}{5} \times \frac{2}{4} \times \frac{2}{3}$ = 0.2	$\frac{3}{5} \times \frac{2}{4} \times \frac{1}{3} \times \frac{2}{2}$ = 0.1

b- $F_X = \{(0.4), (0.7), (0.9), (1)\}$

c- L'espérance de R est simplement la valeur moyenne (pondéré par les possibilités d'occurrence)

$$E[R] = \sum_{i=1}^N x_i P_i = 1 * 0.4 + 2 * 0.3 + 3 * 0.2 + 4 * 0.1 = 2$$

4- Dans une population de canards, lors d'une alerte, l'ensemble des individus quittent leur lieu de repos. Ainsi à $t = 0$, la surface de l'étang est désertée et la probabilité qu'un canard regagne l'étang entre les temps t_1 et t_2 (en minutes) et donne par

$$f(t) = 2 e^{-t} - 2 e^{-2t}$$

- a- Calculer le temps nécessaire pour que le 50 % de la colonie réintègre l'étang, sachant que la densité de probabilité représente l'évolution de la recolonisation de l'étang en fonction du temps.
- b- Donner l'expression analytique et tracer la fonction de répartition en fonction de la variable temps, sachant qu'au bout de 7 minutes tous les canards ont regagné l'étang
- c- Calculer le temps moyen de recolonisation de l'étang

Solution :

a- Nous allons évaluer la fonction densité des probabilités à sa valeur de 0.50

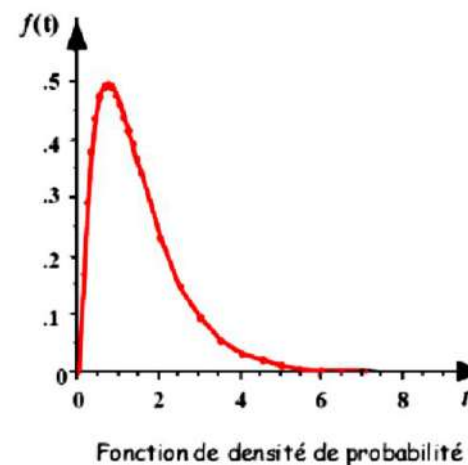
$$f(t) = 2 e^{-t} - 2 e^{-2t} = 0.50$$

$$1 - e^{-t} = 0.50 \quad 0.50 = e^{-t} \quad \ln(0.5) = -t \quad t = 0.70 \text{ min}$$

b- Pour établir la fonction répartition, il faut intégrer la densité des probabilités :

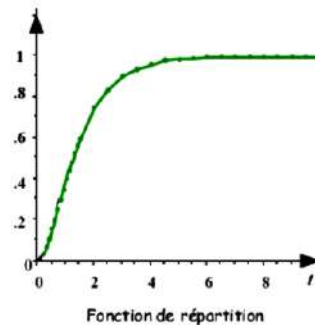
$$F_t(t) = \int_0^t 2 e^{-t} - 2 e^{-2t} dt = 2 \left(\int_0^t e^{-t} dt \right) - \int_0^t 2 e^{-2t} dt$$

$$F_t(t) = 2 \left(1 - e^{-t} \right) - \left(1 - e^{-2t} \right) = \boxed{1 - 2e^{-t} + e^{-2t}}$$



Nous cherchons quelques valeurs pour pouvoir tracer :

$$\begin{aligned} F_t(t=0) &= 0 \\ F_t(t=1) &= 0.4 \\ F_t(t=2) &= 0.75 \\ F_t(t=3) &= 0.90 \\ F_t(t=4) &= 0.96 \\ F_t(t=7) &= 0.998 \end{aligned}$$



- c- Le calcul du temps moyen nous conduit à calculer l'espérance de la variable aléatoire temps :

$$E(t) = \langle t \rangle = \int_0^{+\infty} t(2e^{-t} - 2e^{-2t}) dt = 2\left(\int_0^{+\infty} te^{-t} dt\right) - \int_0^{+\infty} t2e^{-2t} dt$$

$$\langle t \rangle = 2 * \left(\frac{1}{\lambda=1}\right) - \frac{1}{\lambda=2}$$

$$\langle t \rangle = 2(1) - \frac{1}{2} = \frac{3}{2} = 1.5 \text{ minutes}$$

- 5- Dans une substance radioactive, par exemple le radium, la désintégration des noyaux se fait de façon spontanée. Le temps d'attente entre deux désintégrations est modélisé par une loi exponentielle de forme $f(t) = e^{-\lambda t}$ (t en années), où λ est la constante de désintégration. On désigne aussi, la période du radium comme le laps de temps nécessaire à la désintégration de la moitié de sa masse.
- a- Donner l'expression analytique de la période du radium en fonction de la constante de désintégration
- b- Calculer la vie moyenne du radium, sachant que sa période est de 1610 ans

Solution :

- a- Sachant que le radium se désintègre d'un 50%, nous pouvons établir que $f(t) = 0.5$ et donner une expression pour t

$$f(t) = 0.5 = e^{-\lambda t} \rightarrow \ln\left(\frac{1}{2}\right) = -\lambda t \ln(e) \rightarrow -\ln(2) = -\lambda t$$

$$\boxed{t = \frac{\ln(2)}{\lambda}}$$

- b- Si $t = 1610$ ans il suffit de calculer la constante de désintégration à partir de cette période :

$$\lambda = \frac{\ln(2)}{t} = 4.30 \cdot 10^{-4}$$

Ainsi, comme la vie moyenne est associé au moment de premier ordre de la fonction densité de probabilité, donc nous avons :

$$\langle t \rangle = \frac{1}{\lambda} = 2325 \text{ ans}$$

B- Information & statistique

- 6- Soit un jeu de 64 cartes différentes dont la probabilité d'apparaître est la même pour toutes les cartes
- a- On veut choisir une carte dans ce jeu de 64 cartes, quelle est la probabilité P_i d'obtenir une carte parmi ce jeu de cartes ? En déduire la quantité d'information H_1 nécessaire pour obtenir la carte désirée en fonction de P_i . Calculer sa valeur.
 - b- On veut maintenant choisir deux cartes dans ce jeu de 64 cartes quelle est la quantité d'information H_1 nécessaire pour obtenir les deux cartes désirées ? Calculer la valeur.

Solution :

- a- $N = 64$ cartes différentes $\rightarrow$ 64 états équiprobables

○ Probabilité P_i d'obtenir une carte parmi ce jeu $\rightarrow P_i = \frac{1}{N} = \frac{1}{64} = 0.15625$

○ Quantité d'information $\rightarrow I = \log_2 \left(\frac{N_i}{N_f} \right)$ avec $N_i = 64$ états initialement possibles

$$N_f = 1 \text{ état final possible}$$

Selon la relation entre n et P :

$$N_i = \frac{1}{P_i} \quad \text{et} \quad N_f = \frac{1}{P_f}$$

$$P_i = \frac{1}{64} \quad \text{et} \quad P_f = \frac{1}{1} = 1$$

$$I = \log_2 \left(\frac{P_f}{P_i} \right) = -\log_2 \left(\frac{P_i}{P_f} \right)$$

Pour ce cas en particulier on a :

$$I = -\log_2(P_i) = -\log_2 \left(\frac{1}{64} \right) = \log_2(64) = \log_2(2^6) = 6 \log_2(2) = \boxed{6 \text{ bits}}$$

- b- Pour choisir deux cartes l'état finalement sera maintenant $N_f = 2$, ainsi la quantité d'information devient :

$$I = -\log_2 \left(\frac{P_i}{P_f} \right) = -\log_2 \left(\frac{2}{64} \right) = -\log_2 \left(\frac{1}{32} \right) = \log_2(2^5) = 5 \log_2(2) = \boxed{5 \text{ bits}}$$

- 7- Soit deux paquets distincts de cartes, chacun d'eux contenant 32 cartes différentes. Soit I_1 ($i = 1,2$) la quantité d'information nécessaire pour obtenir une carte dans le paquet $n^{\circ}i$.
- a- Quelle est la probabilité P d'obtenir une carte dans chacun des deux paquets de 32 cartes ?
 - b- En déduire que l'information totale I nécessaire pour choisir une carte de chaque paquet est la somme de I_1 et I_2 . Calculer sa valeur.

Solution :

- a- Le nombre total de possibilités N est : $N = N_1 N_2$ avec $N_1 = N_2 = 32 \rightarrow$ ainsi $N = 2^5 2^5 = 2^{10}$
La probabilité d'obtenir une carte dans chacun des 2 paquets :

$$P = \frac{1}{N} = \frac{1}{2^{10}}$$

b- L'information totale : $I = \log_2 \left(\frac{N_i}{N_f} \right)$ avec :

$$N_i = 2^{10} \quad \text{et} \quad N_f = 1$$

$$I = \log_2 \left(\frac{N_i}{N_f} \right) = \log_2(N) = \log_2(N_1 N_2) = \log_2(N_1) + \log_2(N_2) = I_1 + I_2$$

$$I_1 + I_2 = \log_2(2^5) + \log_2(2^5) = 5 \log_2(2) + 5 \log_2(2) = \boxed{10 \text{ bits}}$$

$$I_{tot} = \log_2(N) = \log_2(2^{10}) = \boxed{10 \text{ bits}}$$

8- On dispose de deux paquets de 64 cartes.

- a- Quelle est la quantité d'information I nécessaire pour obtenir une carte dans le premier paquet et deux cartes dans le deuxième paquet ?
- b- Quelle est la quantité d'information I nécessaire pour obtenir quatre cartes dans le premier paquet et deux cartes dans le deuxième paquet ?

Solution :

a- La quantité d'information en fonction des états initiale et finale :

$$I = \log_2 \left(\frac{N_i}{N_f} \right)$$

Avec :

$$N_i = 2^6 2^6 = 2^{12} \quad \text{et} \quad N_f = N_{f1} N_{f2} = 1 \times 2 = 2$$

$$I = \log_2 \left(\frac{2^{12}}{2} \right) = \log_2(2^{11}) = 11 \log_2(2) = \boxed{11 \text{ bits}}$$

b- Dans ce cas la quantité d'information sera pour $N_f = N_{f1} N_{f2} = 4 \times 2 = 8$

$$I = \log_2 \left(\frac{2^{12}}{8} \right) = \log_2 \left(\frac{2^{12}}{2^3} \right) = \log_2(2^9) = 9 \log_2(2) = \boxed{9 \text{ bits}}$$

9- On veut choisir une carte parmi 2 cartes dont la probabilité d'apparaître est la même pour les deux cartes. Quelle est la quantité d'information nécessaire pour l'obtenir, que peut-on conclure ?

Solution : Dans ce cas on a :

$$N_i = 2 \quad \text{et} \quad N_f = 1$$

$$H = \log_2 \left(\frac{N_i}{N_f} \right) = \log_2 \left(\frac{2}{1} \right) = \boxed{2 \text{ bits}} \rightarrow \text{Information qu'il faut acquérir pour lever le doute entre deux symboles équiprobables.}$$

10- Soit un dé normal à 6 faces, chaque face portant un numéro différent entre 1 et 6. Chaque face a la même probabilité P d'apparaître.

- a- On lance le dé. Déterminer l'expression littérale de l'information moyenne apportée par chacun des symboles du dé normal à six faces en fonction de P . Calculer sa valeur.
- b- Le dé est truqué. Ainsi en lançant le dé on est sûrs d'obtenir une des faces. Déterminer l'information moyenne apportée par chacun des symboles.

- c- On lance un dé partiellement truqué. Ainsi en lançant le dé, la face 1 a une probabilité d'apparaître de 50% et les autres faces de 10%. Déterminer l'information moyenne apportée par chacun des symboles. Conclure.

Solution :

- a- La probabilité pour chaque face est de $P_i = \frac{1}{6} = P = 0.166$

L'information moyenne est donnée par théorème de Shannon

$$I = - \sum_{i=1}^6 P_i \log_2(P_i)$$

Alors à partir de cette expression :

$$I = - \log_2(P) \left(\sum_{i=1}^6 P_i \right) = - \log_2(P)$$

$$I = - \log_2\left(\frac{1}{6}\right) = \log_2(6) = \frac{\ln(6)}{\ln(2)} = \boxed{2.58 \text{ bits/symb}}$$

- b- Pour ce cas, quand le dé est truqué : $P_i = 1$ et $P_j = 0$ ($i \neq j$)

$I = 0 \rightarrow$ On sait déjà quelle face va apparaître !

- c- A partir de l'expression de l'information (on a différentes probabilités) on doit additionner sur toutes ses probabilités :

$$I = - \sum_{i=1}^6 P_i \log_2(P_i) = -[0.5 \log_2(0.5) + 5 \times 0.1 \log_2(0.1)]$$

$$I = - \left[\frac{1}{2} \log_2\left(\frac{1}{2}\right) + 5 \times \frac{1}{10} \log_2\left(\frac{1}{10}\right) \right] = - \left[-\frac{1}{2} \log_2(2) - 5 \times \frac{1}{10} \log_2(10) \right]$$

$$I = \frac{1}{2} \log_2(2) + 5 \times \frac{1}{10} \log_2(10) = 0.5 + 0.5 \times 3.32 = 2.16 \text{ bits/symb}$$

$\boxed{2.16 < 2.58} \rightarrow$ Puisque l'incertitude sur le résultat est inférieure que pour un système équiprobable.

- 11- L'ADN est conformé par des longues séquences de 4 bases phosphatées A, T, C et G . Dans l'ADN des primates, les 4 bases ont la même fréquence d'apparition. Par contre, dans le matériel génétique d'une bactérie les concentrations de 4 bases sont radicalement différentes : il y a une fréquence d'apparition $P(G) = P(C) = 0.38$ et $P(A) = P(T) = 0.12$. Calculer l'information moyenne apportée par la séquence des primates et la comparer à celle de la bactérie.

Solution :

Le matériel génétique des primates possède des séquences où les 4 bases sont équiprobables :

$$I = - \log_2(P) \left(\sum_{i=1}^4 P_i \right) = - \log_2(P)$$

$$I_{\text{primate}} = - \log_2 \frac{1}{4} = \log_2(2^2) = \boxed{2 \text{ bits}}$$

Le matériel génétique des bactéries les fréquences des bases comportent des différentes probabilités donc, nous avons

$$I = - \sum_{i=1}^4 P_i \log_2(P_i)$$

$$I_{bacterie} = 2 \times 0.38 \log_2\left(\frac{1}{0.38}\right) + 2 \times 0.12 \log_2\left(\frac{1}{0.12}\right) = 2 \times 0.38 (1.39) + 2 \times 0.12 (3.05)$$

$I_{bacterie} = 1.788 \text{ bits}$

12- Soient n cartes sur lesquelles sont inscrites sur la face cachée un nombre (tous les nombres sont différents). L'objectif est de réordonner les cartes. On doit payer 1€ à chaque fois qu'on veut connaître la réponse à la question « *Est-ce que le nombre sur cette carte est supérieur à celle-là ?* » On obtient juste cette information sans connaître le nombre inscrit sur la carte. Au final on reçoit $3 \times n$ € moins le nombre des questions posées lorsqu'on a reconstitué l'ordre des cartes.

- a- Quel est le nombre de fois en moyenne où l'on doit poser la question pour restituer l'ordre des cartes ?
- b- Avons-nous statistiquement plus de chance de gagner de l'argent si $n = 2, 3, 6$ ou 10 ?
- c- Quelle valeur de n permet de gagner statistiquement le plus d'argent ?

Solution :

a- Il y a $n!$ manières d'ordonner les cartes, par exemple :

- $n = 2 \rightarrow 2! = 1 \times 2 = 2$
- $n = 3 \rightarrow 3! = 1 \times 2 \times 3 = 6$
- $n = 4 \rightarrow 4! = 1 \times 2 \times 3 \times 4 = 24$

L'incertitude sur l'ordre dans lequel ces cartes sont rangées (ou information à obtenir) est égale à $I = \log_2(n!)$. Ce nombre correspond donc au nombre des questions en moyenne à poser avant d'établir l'ordre des cartes (sans connaître le nombre écrit sur chaque carte)

b- Pour $n = 2$ il y a $2! = 2$ manières d'ordonner les cartes, donc l'incertitude sur l'ordre correspond à

$$\log_2(n!) = \log_2(2) = 1 \text{ question (bit)}$$

Que veut dire qu'il faudra juste une question pour connaître l'ordre de 2 cartes. Donc : $3 \times 2 \text{ €} - 1 \text{ €} = 5 \text{ €}$

- Pour $n = 3$ il y a $3! = 6$ manières d'ordonner les cartes donc l'incertitude sur l'ordre correspond à

$$\log_2(6) = 2.58 \text{ bit}$$

Il faudra en moyenne 2.5 questions pour connaître l'ordre des cartes. On gagnera, $3 \times 3 \text{ €} - 2.5 \text{ €} = 6.5 \text{ €}$

- Pour $n = 6$ il y a $6! = 720$ manières d'ordonner les cartes donc l'incertitude sur l'ordre correspond à

$$\log_2(720) = 9.5 \text{ bit}$$

Il faudra en moyenne 9.5 questions pour connaître l'ordre des cartes. On gagnera, $3 \times 6 \text{ €} - 9.5 \text{ €} = 8.5 \text{ €}$

- Pour $n = 10$ il y a $10! = 3.6 \times 10^6$ manières d'ordonner les cartes donc l'incertitude sur l'ordre correspond à

$$\log_2(3.6 \times 10^6) = 21.8 \text{ bit}$$

Il faudra en moyenne 22 questions pour connaître l'ordre des cartes. On gagnera, $3 \times 10 \text{ €} - 21.8 \text{ €} = 8.2 \text{ €}$

c- La valeur de n que permet d'obtenir plus d'argent : on utilise un graphique pour explorer la fonction

$$f(n) = 3 \times n - \log_2(n!)$$

$$f(n) = 3 \times n - \frac{\ln(n!)}{\ln(2)}$$

➔ On voit le maximum de cette fonction en $n = 7$ et $n = 8$

