U.F.R. de Mathématiques
Licence de Mathématiques
S6, M66, année 2013-2014

Probabilités M66
Corrigé du D.S. du 28 mars 2014

Ex 1. *Interprétation du graphique d'une fonction de survie*

La variable aléatoire X a pour fonction de survie $G : t \mapsto G(t) = P(X > t)$ représentée figure 1.

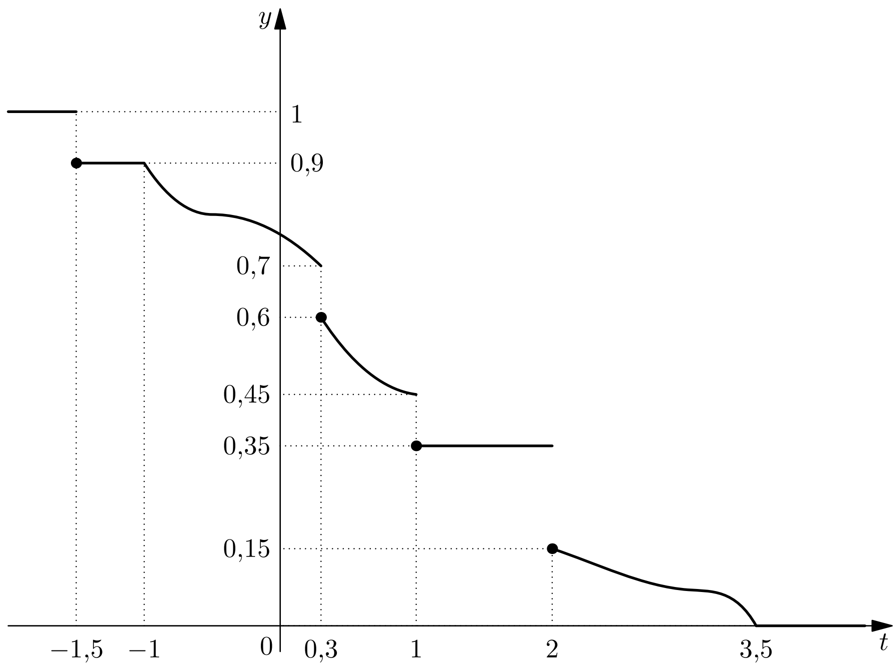

Figure 1 – Fonction de survie G de la v.a. X

1) La fonction de survie $G = 1 - F$ où F est la fonction de répartition est comme F, continue à droite et limitée à gauche en tout point. Il est facile d'adapter les formules du cours pour F exprimant les probabilités d'appartenance de X à des intervalles de tout type. Ainsi on a

$$P(X \in]a, b]) = F(b) - F(a) = (1 - G(b)) - (1 - G(a)) = G(a) - G(b)$$

et on adapte de manière analogue les formules pour les trois autres intervalles d'extrémités a et b. Notons aussi que

$$\forall c \in \mathbb{R}, \quad P(X = c) = F(c) - F(c-) = (1 - G(c)) - (1 - G(c-)) = G(c-) - G(c).$$

n°	Question	Réponse	n°	Question	Réponse
1	$P(X \leqslant -1) = ?$	0,1	2	$P(X = -1{,}5) = ?$	0,1
3	$P(X = 0{,}35) = ?$	0	4	$P(X \geqslant 0{,}3) = ?$	0,7
5	$P(X > 2) = ?$	0,15	6	$P(X \in [1; 1{,}9]) = ?$	0,1
7	$P(X \in]1; 2]) = ?$	0,2	8	$P(\lvert X\rvert > 1) = ?$	0,45

2) La variable aléatoire X n'est pas à densité car sinon sa fonction de répartition et donc aussi sa fonction de survie serait continue sur $\mathbb{R}$.

3) La variable aléatoire X n'est pas discrète car la somme des amplitudes des sauts de sa fonction de survie (ou de sa fonction de répartition) est égale à 0,5. Si X était discrète, sa loi serait discrète et cette somme devrait valoir 1.

4) On voit sur le graphique que $P(X \in [-1{,}5; 3{,}5]) = 1$. Donc X est bornée p.s., ce qui implique l'intégrabilité.

Ex 2. *Partie fractionnaire de l'inverse*

On note $[x]$ la partie entière du réel x, autrement dit $[x]$ est l'unique entier $n \in \mathbb{Z}$ tel que $n \leqslant x < n+1$. Soit f la fonction :

$$f : x \longmapsto \frac{1}{(1+x)\ln 2}\mathbf{1}_{]0,1[}(x).$$

1) Il est clair que f est borélienne positive. Pour que ce soit une densité, il suffit donc de vérifier que son intégrale sur $\mathbb{R}$ vaut 1 :

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = \int_0^1 \frac{1}{(1+t)\ln 2}\,\mathrm{d}t = \left[\frac{\ln(1+t)}{\ln 2}\right]_0^1 = \frac{\ln 2 - \ln 1}{\ln 2} = 1.$$

La fonction de répartition s'obtient en intégrant la densité :

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{x} \frac{1}{(1+t)\ln 2}\mathbf{1}_{]0,1[}(t)\,\mathrm{d}t.$$

Pour $x < 0$, cette intégrale vaut 0, pour $x \geqslant 1$, elle vaut $\int_0^1 f(t)\,\mathrm{d}t$, c'est à dire 1, pour $0 \leqslant x \leqslant 1$, elle vaut :

$$\int_0^x f(t)\,\mathrm{d}t = \left[\frac{\ln(1+t)}{\ln 2}\right]_0^x = \frac{\ln(1+x)}{\ln 2}.$$

En rassemblant ces trois cas en une seule formule :

$$\forall x \in \mathbb{R}, \quad F(x) = \frac{\ln(1+x)}{\ln 2}\mathbf{1}_{[0,1[}(x) + \mathbf{1}_{[1,+\infty[}(x).$$

2) Soit X une variable aléatoire réelle dont la loi a pour densité f. On va calculer :

$$P\Big(k \leqslant \frac{1}{X} \leqslant k+t\Big) \qquad k \in \mathbb{N}^*,\ t \in]0,1[.$$

La fonction $x \mapsto 1/x$ étant décroissante sur $]0,+\infty[$, on a :

$$\forall k \in \mathbb{N}^*,\ \forall t \in]0,1[, \quad k \leqslant \frac{1}{X} \leqslant k+t \iff \frac{1}{k+t} \leqslant X \leqslant \frac{1}{k}.$$

Ainsi les deux évènements définis par les deux membres de cette équivalence sont confondus, d'où :

$$P\left(k \leqslant \frac{1}{X} \leqslant k+t\right) = P\left(\frac{1}{k+t} \leqslant X \leqslant \frac{1}{k}\right) = P\left(X \leqslant \frac{1}{k}\right) - P\left(X < \frac{1}{k+t}\right).$$

La loi de X étant à densité, $P(X=x)=0$ pour tout $x \in \mathbb{R}$, d'où $P(X<x)=P(X \leqslant x)$. En appliquant ceci à $x = 1/(k+t)$, on obtient :

$$\begin{aligned} P\left(X \leqslant \frac{1}{k}\right) - P\left(X < \frac{1}{k+t}\right) &= F\left(\frac{1}{k}\right) - F\left(\frac{1}{k+t}\right) \\ &= \frac{1}{\ln 2}\left(\ln\left(\frac{k+1}{k}\right) - \ln\left(\frac{k+1+t}{k+t}\right)\right). \end{aligned}$$

Finalement, pour tout $k \in \mathbb{N}^*$ et tout $t \in]0,1[$:

$$P\left(k \leqslant \frac{1}{X} \leqslant k+t\right) = \frac{1}{\ln 2}\big(\ln(k+1) - \ln k + \ln(k+t) - \ln(k+1+t)\big) \tag{1}$$

3) Soit $Y = \frac{1}{X} - \left[\frac{1}{X}\right]$. Cherchons sa fonction de répartition G définie par $G(t) = P(Y \leqslant t)$. D'abord, il est clair qu'avec probabilité 1, $0 \leqslant Y \leqslant 1$. Donc $G(t) = 0$ pour tout $t < 0$ et $G(t) = 1$ pour tout $t \geqslant 1$. Ainsi $F(t) = G(t)$ dans ces deux cas. Calculons $G(t)$ pour $0 < t < 1$. On note pour cela que :

$$\begin{aligned} \{Y \leqslant t\} &= \left\{\frac{1}{X} - \left[\frac{1}{X}\right] \leqslant t\right\} \\ &= \bigcup_{k \in \mathbb{N}^*} \left(\left\{\left[\frac{1}{X}\right] = k\right\} \cap \left\{\frac{1}{X} - k \leqslant t\right\}\right) \\ &= \bigcup_{k \in \mathbb{N}^*} \left(\left\{k \leqslant \frac{1}{X} < k+1\right\} \cap \left\{\frac{1}{X} \leqslant k+t\right\}\right) \\ &= \bigcup_{k \in \mathbb{N}^*} \left\{k \leqslant \frac{1}{X} \leqslant k+t\right\}. \end{aligned}$$

On a ainsi décomposé l'évènement $\{Y \leqslant t\}$ en une réunion dénombrable d'évènements disjoints. Par conséquent,

$$P(Y \leqslant t) = \sum_{k=1}^{+\infty} P\left(k \leqslant \frac{1}{X} \leqslant k+t\right).$$

Pour calculer cette série, on utilise (1) en revenant aux sommes partielles et en remarquant que chaque terme se simplifie avec son successeur :

$$\begin{aligned}\sum_{k=1}^{m} P\left(k \leqslant \frac{1}{X} \leqslant k+t\right) &= \frac{1}{\ln 2}\left(\sum_{k=1}^{m}\big(\ln(k+1)-\ln k\big)+\sum_{k=1}^{m}\big(\ln(k+t)-\ln(k+1+t)\big)\right)\\ &= \frac{1}{\ln 2}\big(\ln(m+1)+\ln(1+t)-\ln(m+1+t)\big).\end{aligned}$$

Pour trouver la limite quand m tend vers $+\infty$ de cette expression, on écrit :

$$\ln(m+1+t)=\ln(m+1)+\ln\left(1+\frac{t}{m+1}\right),$$

d'où l'on déduit :

$$\lim_{m\to+\infty}\frac{1}{\ln 2}\big(\ln(m+1)+\ln(1+t)-\ln(m+1+t)\big)=\frac{\ln(1+t)}{\ln 2}.$$

Ainsi les fonctions F et G sont égales et Y a même loi que X.

4) X est une variable aléatoire vérifiant $P(X\in[0,1])=1$, dont est intégrable comme variable aléatoire p.s. bornée. Y a même loi que X, donc Y est aussi intégrable et $\mathbf{E}Y=\mathbf{E}X$. Enfin, l'espérance de X se calcule facilement comme suit.

$$\mathbf{E}X=\int_0^1\frac{x}{(1+x)\ln 2}\,\mathrm{d}x=\frac{1}{\ln 2}\int_0^1\left(1-\frac{1}{1+x}\right)\mathrm{d}x=\frac{1}{\ln 2}-\frac{1}{\ln 2}\left[\ln(1+x)\right]_0^1=\frac{1}{\ln 2}-1.$$

Ex 3. *Sur la « loi du tiers »*

Sur `http://www.pronostic-sportif-gratuit.com/casino/loi-du-tiers.php`, on peut lire ce qui suit [1].

> *« Cette loi est sans doute la plus importante, car elle se vérifie de manière constante sur des cycles à court terme. Les numéros de la Roulette sont au nombre de 37, Zéro inclus. [...]*
>
> *« Bien que la probabilité de sortie de chaque numéro en 37 boules soit de 1 chance sur 37, les 37 numéros de la roulette n'apparaissent pas tous au cours d'un cycle de 37 boules. Environ 12 numéros (1/3) sortent plusieurs fois, environ 12 autres numéros (1/3) sortent une seule fois, et environ 12 autres numéros encore (1/3) ne sortent pas du tout. C'est cette particularité qui a donné son nom à cette loi. »*

Le but de cet exercice est d'essayer d'y voir plus clair sur ce sujet. On se limitera à la question des numéros qui ne sortent pas du tout. Pour cela, on va généraliser le problème en considérant une suite de n tirages avec remise d'une boule dans une urne contenant exactement n boules numérotées de 1 à n. La roulette correspond au cas $n=37$ (avec un décalage d'une unité sur la numérotation).

1. À la date du 19 mars 2014.

1) Calculons la probabilité de l'évènement

$$A_n = \{\text{en } n \text{ tirages, chaque } i \in [\![1,n]\!] \text{ sort exactement une fois}\}.$$

Puisque les tirages se font avec remise, le modèle standard qui s'applique ici est $\Omega_n = [\![1,n]\!]^n$, ensemble de tous les n-uplets d'entiers pris entre 1 et n, muni de la tribu $\mathcal{P}(\Omega_n)$ de toutes les parties de Ω et de l'équiprobabilité P_n sur Ω_n. Un évènement élémentaire quelconque $\omega = (x_1, \dots, x_n)$ a donc une probabilité $1/n^n$ de se réaliser. On arriverait au même résultat en considérant que les tirages sont indépendants et qu'à chaque tirage, l'entier i a une chance sur n d'être choisi. Puisque P_n est l'équiprobabilité sur Ω,

$$P_n(A_n) = \frac{\operatorname{card} A_n}{\operatorname{card} \Omega_n}.$$

Le cardinal de A_n est exactement le nombre de permutations sur l'ensemble $[\![1,n]\!]$ (nombre de bijections de cet ensemble dans lui-même), c'est-à-dire $n!$, d'où

$$P_n(A_n) = \frac{n!}{n^n}.$$

Applications numériques :

$$P_6(A_6) \simeq 0{,}015\,432\,1.$$

$$P_{37}(A_{37}) \simeq 1{,}303\,986\,462 \times 10^{-15}.$$

Ces probabilités sont très faibles. Déjà pour un dé, la probabilité d'obtenir 6 numéros différents en 6 lancers ne vaut guère plus de 1,5%. Pour la roulette, on a moins d'un chance sur 100 000 milliards d'observer la sortie de chacun des 37 numéros au cours de 37 parties consécutives. Le caractère rarissime de cet évènement est bien expliqué par le calcul des probabilités.

Pour aller un peu plus loin, on peut remarquer que $P_n(A_n)$ tend vers 0 quand n tend vers l'infini et donner un équivalent grâce à la formule de Stirling : $n! \sim \sqrt{2\pi} n^{n+1/2} \mathrm{e}^{-n}$, d'où

$$P_n(A_n) \sim \sqrt{2\pi n}\,\mathrm{e}^{-n}, \quad n \to +\infty.$$

Dans la suite, étant donné n fixé, on note X_i le nombre d'apparitions de la boule n° i en n tirages.

2) La suite des tirages avec remise constituant une suite d'épreuves répétées indépendantes, X_i suit la loi binomiale de paramètres n (nombre d'épreuves répétées indépendantes) et $p = 1/n$ (probabilité d'un succès, ici la sortie du n° i, lors d'une épreuve).

$$\forall k \in [\![1,n]\!], \quad P_n(X_i = k) = C_n^k \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k}.$$

3) Les X_i ne peuvent être indépendantes car

$$\forall \omega \in \Omega_n, \quad \sum_{i=1}^{n} X_i(\omega) = n,$$

ce qui entraîne pour $1 \leqslant j \leqslant n$,

$$P_n(X_n = k \mid X_1 + \cdots + X_{n-1} = j) = \begin{cases} 1 & \text{si } k = n - j, \\ 0 & \text{sinon.} \end{cases} \qquad (\star)$$

Or si les X_i étaient indépendantes, le vecteur $(X_1, \dots, X_{n-1})$ et la variable aléatoire X_n devraient être indépendants, ce qui entraînerait l'indépendance des évènements $\{X_n = k\}$ et $\{X_1 + \cdots + X_{n-1} = j\}$ et donc l'égalité $P_n(X_n = k \mid X_1 + \cdots + X_{n-1} = j) = P_n(X_n = k)$, en contradiction avec $(\star)$.

4) Le vecteur aléatoire $(X_1, \dots, X_n)$ suit la loi multinomiale de paramètres n (nombre d'épreuves répétées indépendantes) et $(\frac{1}{n}, \dots, \frac{1}{n})$, vecteur des probabilités des n résultats possibles lors d'une épreuve.

On considère maintenant la variable aléatoire

$$S_n = \sum_{i=1}^{n} \mathbf{1}_{\{X_i = 0\}}.$$

5) S_n représente le nombre aléatoire de numéros non sortis au cours des n tirages. En effet en tant que somme de termes valant zéro ou 1, $S_n(\omega)$ est égale au nombre de ses termes égaux à 1. Or le terme $\mathbf{1}_{\{X_i=0\}}(\omega)$ vaut 1 si et seulement si $X_i(\omega) = 0$, c'est-à-dire si et seulement si le numéro i n'a fait aucune apparition en n tirages.

6) Pour calculer $\mathbf{E}\left(\frac{S_n}{n}\right)$, en utilisant la linéarité de l'espérance, la formule $\mathbf{E}\mathbf{1}_B = P(B)$ et le fait que les X_i ont même loi, on obtient :

$$\mathbf{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\mathbf{E}S_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\mathbf{1}_{\{X_i=0\}} = \frac{1}{n}\sum_{i=1}^{n}P_n(X_i = 0) = \frac{1}{n}nP_n(X_1 = 0) = P_n(X_1 = 0).$$

Comme X_1 suit la loi $\mathrm{Bin}(n, 1/n)$, on en déduit que

$$\mathbf{E}\left(\frac{S_n}{n}\right) = \left(1 - \frac{1}{n}\right)^n.$$

Il est bien connu [2] que

$$\lim_{n\to+\infty}\left(1 - \frac{1}{n}\right)^n = \mathrm{e}^{-1} \simeq 0{,}368.$$

Dans le cas de la roulette, on trouve :

$$\mathbf{E}\left(\frac{S_{37}}{37}\right) = \left(\frac{36}{37}\right)^{37} \simeq 0{,}362\,85.$$

2. Si vous l'avez oublié, montrez que $\ln\left((1 - 1/n)^n\right)$ tend vers -1.

7) Pour interpréter le résultat précédent, on va utiliser la version suivante de la loi forte des grands nombres : si $(Y_j)_{j\geqslant 1}$ est une suite de variables aléatoires indépendantes, de même loi et bornées,

$$P\left(\lim_{N\to+\infty}\frac{1}{N}\sum_{j=1}^{N}Y_j=\mathbf{E}Y_1\right)=1.$$

On travaille toujours avec n fixé et on effectue N séries de n tirages avec remise dans l'urne. Pour la j^{e} de ces séries, on note Y_j la proportion de numéros non sortis au cours des n tirages. D'après la loi forte des grands nombres, pour N grand, la fréquence observée de ces proportions de numéros non sortis en n tirages, c'est-à-dire $N^{-1}\sum_{j=1}^{N}Y_j$ tend vers $\mathbf{E}Y_1=\mathbf{E}(S_n/n)=(1-1/n)^n$ avec probabilité 1. En pratique, cela signifie que pour N grand, cette fréquence est proche de $(1-1/n)^n$. Pour la roulette, cette fréquence doit donc être proche de 0,362 85, ce qui n'est pas très loin de 1/3.

Remarque complémentaire. On peut utiliser la même méthode pour trouver l'espérance de la proportion de numéros sortant exactement une fois en n tirages. Cette proportion s'écrit S'_n/n, où

$$S'_n=\sum_{i=1}^{n}\mathbf{1}_{\{X_i=1\}}$$

d'où

$$\mathbf{E}\left(\frac{S'_n}{n}\right)=P_n(X_1=1)=n\frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}=\left(1-\frac{1}{n}\right)^{n-1}.$$

Pour la roulette on obtient ainsi

$$\mathbf{E}\left(\frac{S'_{37}}{37}\right)=\left(\frac{36}{37}\right)^{36}\simeq 0{,}372\,93.$$

On en déduit immédiatement l'espérance de la proportion de numéros sortis au moins 2 fois :

$$\mathbf{E}\left(\frac{S''_{37}}{37}\right)=1-\mathbf{E}\left(\frac{S'_{37}}{37}\right)-\mathbf{E}\left(\frac{S_{37}}{37}\right)\simeq 0{,}264\,22.$$

Finalement, on peut corriger la loi du tiers pour la roulette sous la forme suivante : « sur un grand nombre de séries de 37 coups, la proportion de numéros qui ne sortent pas lors d'une série est d'environ 36, 3 %, celle des numéros sortant exactement une fois en 37 coups est d'environ 37, 3 % et celle des numéros sortant au moins 2 fois est d'environ 26, 4 % ».