TD de Probabilités et Statistique
Correction du TD 3

Exercice 1 : Vers la simulation de variables aléatoires

Soit X une variable aléatoire réelle et F sa fonction de répartition. On suppose F continue et strictement croissante. Quelle est la loi de $Y = F(X)$?

Fonction de répartition : Elle est définie pour tout $x \in \mathbb{R}$ par $F(x) = \mathbb{P}[X \leq x] = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$. Où f_X est la densité de X.

Pour commencer, Y est bien une variable aléatoire. En effet, F est une fonction de $\mathbb{R}$ dans $[0,1]$, et X est une v.a. qui prend des valeurs sur $\mathbb{R}$. Donc $Y = F(X)$ est une v.a. à valeur dans $[0,1]$.

Notion F_Y la fonction de répartition de Y, et essayons de la déterminer. Par définition, $\forall y \in \mathbb{R}$, $F_Y(y) = \mathbb{P}[Y \leq y]$. On a donc,

1. $\forall y \leq 0$, $F_Y(y) = \mathbb{P}[Y < y] = 0$,
2. $\forall y \geq 1$, $F_Y(y) = \mathbb{P}[Y < y] = 1$.

Enfin, la fonction de répartition F de X est strictement croissante et continue. Elle est donc inversible sur $\mathbb{R}$. Notons F^{-1} son inverse.

Soit $y \in [0,1]$, on a $F_Y(y) = \mathbb{P}[Y < y] = \mathbb{P}[F(X) < y] = \mathbb{P}\left[X < F^{-1}(y)\right] = F(F^{-1}(y)) = y$.

Finalement, $F_Y(y) = \begin{cases} 0 & y \leq 0 \\ y & y \in [0,1] \\ 1 & y \geq 1 \end{cases}$. Donc Y suit la loi uniforme sur $[0,1]$.

Remarque : Pour une loi de probabilité continue, pour tout $x \in \mathbb{R}$, $\mathbb{P}[X = x] = 0$. Donc $\mathbb{P}[X < x] = \mathbb{P}[X \leq x]$.

Utilité d'un tel résultat. Pour simuler (informatiquement) une variable aléatoire dont on connait la fonction de répartition F continue et strictement croissante, il nous suffit de générer une variable aléatoire Y qui suit la loi uniforme sur $[0,1]$ puis de faire $X = F^{-1}(Y)$.

Exercice 2 : Lois du Sup et de l'Inf

Dans ce qui suit, X et Y sont des variables aléatoires indépendantes. On pose $Z = \sup(X,Y)$ et $T = \inf(X,Y)$. On demande de déterminer les lois de Z et de T dans les cas suivants :

1. *X et Y suivent la loi uniforme sur $\{1,\ldots,n\}$,*
2. *X et Y suivent des lois exponentielles de paramètres respectifs $\lambda > 0$ et $\mu > 0$.*

1) Les lois de X et Y sont discrètes, on va donc chercher $\mathbb{P}[Z=k]$ pour les entiers $k \in \{1,\dots,n\}$. La variable $Z=\sup(X,Y)$ est à valeur dans le même ensemble que X et Y, donc $\{1,\dots,n\}$.
Puis pour $k \in \{1,\dots,n\}$, $\mathbb{P}[Z=k] = \mathbb{P}[(X=k \text{ et } Y \leq k) \text{ ou } (Y=k \text{ et } X<k)]$.

Les évènement $(X=k \text{ et } Y \leq k)$ et $(Y=k \text{ et } X<k)$ sont disjoints (ils ne peuvent pas se produire en même temps, grâce à l'inégalité stricte en rouge), on peut donc decomposer l'union en somme :

$$\mathbb{P}[Z=k] = \mathbb{P}[(X=k \text{ et } Y \leq k) \text{ ou } (Y=k \text{ et } X<k)] = \mathbb{P}[(X=k \text{ et } Y \leq k)] + \mathbb{P}[(Y=k \text{ et } X<k)].$$

En utilisant l'indépendance de X et Y, on a
— $\mathbb{P}[(X=k \text{ et } Y \leq k)] = \mathbb{P}[X=k]\,\mathbb{P}[Y \leq k] = \frac{1}{n}\frac{k}{n}$
— $\mathbb{P}[(Y=k \text{ et } X<k)] = \mathbb{P}[Y=k]\,\mathbb{P}[X<k] = \frac{1}{n}\frac{k-1}{n}$

Finalement,

$$\mathbb{P}[Z=k] = \frac{2k-1}{n^2}$$

(et on peut vérifier que $\sum_{k=1}^{n} \mathbb{P}[Z=k]=1$).

Pense-bête : Pour couper une union $\cup$ (ou) en "+", il faut que les événements soient disjoints. Pour couper une intersection $\cap$ (et) en $\times$ il faut que les variables aléatoires soient indépendantes.

Avec le même raisonnement on trouve pour T, pour $k \in \{1,\dots,n\}$,

$$\mathbb{P}[T=k] = \frac{2n-2k+1}{n^2}.$$

(bonus) : X et Y suivent la loi uniforme sur $[0,1]$ Cette fois les lois sont continues on va donc chercher leur fonction de répartition. L'idée reste la même. Soit $t \in [0,1]$,

$$\mathbb{P}[Z<t] = \mathbb{P}[(X<t)\cap(Y<t)] \overset{\text{(independance)}}{=} \mathbb{P}[X<t]\,\mathbb{P}[Y<t] = t^2.$$

Donc la fonction de répartition de Z est

$$F_Z(t) = \begin{cases} 0 & t \leq 0 \\ t^2 & t \in [0,1] \\ 1 & t \geq 1 \end{cases}.$$

On peut aussi calculer sa densité en dérivant : $F_Z'(t) = 2t\,\mathbf{1}_{[0,1]}(t)$.

De même on peut montrer que la fonction de répartition de T est,

$$F_T(t) = \begin{cases} 0 & t \leq 0 \\ 1-(1-t)^2 & t \in [0,1] \\ 1 & t \geq 1 \end{cases},$$

de densité $F_T'(t) = 2(1-t)\,\mathbf{1}_{[0,1]}(t)$.

2) Cette fois X et Y suivent la loi exponentielle (continue) de paramètre λ, resp. μ. Leurs fonctions de densité sont $f_X(t) = \lambda e^{-\lambda t}\mathbf{1}_{\mathbb{R}_+}(t)$ resp. $f_Y(t) = \mu e^{-\mu t}\mathbf{1}_{\mathbb{R}_+}(t)$. Leurs fonctions de répartitions sont $F_X(t) = (1-e^{-\lambda t})\mathbf{1}_{\mathbb{R}_+}(t)$, resp. $F_Y(t) = (1-e^{-\mu t})\mathbf{1}_{\mathbb{R}_+}(t)$

Comme avant, on a donc, pour tout $t \geq 0$,

$$\mathbb{P}[Z<t] = \mathbb{P}[(X<t)\cap(Y<t)] \overset{\text{(independance)}}{=} \mathbb{P}[X<t]\,\mathbb{P}[Y<t] = \left(1-e^{-\lambda t}\right)\left(1-e^{-\mu t}\right).$$

De même, pour tout $t \geq 0$,

$$\mathbb{P}[T>t] = \mathbb{P}[(X>t)\cap(Y>t)] \overset{\text{(independance)}}{=} \mathbb{P}[X>t]\,\mathbb{P}[Y>t] = (1-F_X(t))(1-F_Y(t)) = e^{-(\lambda+\mu)t}.$$

Donc $\mathbb{P}[T \leq t] = (1-e^{-(\lambda+\mu)t})\mathbf{1}_{\mathbb{R}_+}(t)$. La variable T suit la loi exponentielle de paramètre $\lambda+\mu$.

Application aux files d'attente *Trois clients A, B et C se présentent à deux guichets libres. A et B entrent en service et C attend que l'un des deux guichets se libère puis entre en service. On suppose que les temps de service de A, B et C sont des variables aléatoires indépendantes de lois exponentielles de paramètres λ_A, λ_B et λ_C. Quel est le temps moyen d'attente de C ? Quel est le temps moyen passé par C dans le système (attente + service) ?*

Le temps d'attente de C est le temps que A ou B soient servis. Son temps d'attente est donc representé par $T_C = \inf(S_A, S_B)$ où S_A et S_B représentent les temps de service de A et B. On a montré à la question (2) que T_C suit la loi exponentielle de paramètre $\lambda_A + \lambda_B$ donc le temps moyen d'attente de T_C est,

$$\mathbb{E}[T_C] = \frac{1}{\lambda_A + \lambda_B}.$$

De même, $\mathbb{E}[\text{temps total passé par C}] = \mathbb{E}[\text{temps d'attente de C} + \text{temps de service de C}]$. On obtient,

$$\mathbb{E}[\text{temps total passé par C}] = \mathbb{E}[T_C] + \mathbb{E}[S_C] = \frac{1}{\lambda_A + \lambda_B} + \frac{1}{\lambda_C}.$$

Exercice 3 : Mélanges gaussiens

Au cours d'une récolte de fruits, les variétés A et B sont mélangées. Le poids d'un fruit de la variété A est une variable aléatoire X de loi gaussienne $\mathcal{N}(m_1, \sigma_1^2)$. Le poids d'un fruit de la variété B est une variable aléatoire Y de loi gaussienne $\mathcal{N}(m_2, \sigma_2^2)$. Les fruits sont mélangés dans la proportion $p \in]0,1[$ pour la variété A et $1-p$ pour la variété B. Appelons Z la variable aléatoire égale au poids d'un fruit quelconque de la récolte.

1. *Quelle est la fonction de répartition de Z ?*
2. *Déterminer $\mathbb{E}(Z)$ et $V(Z)$.*

1) Soit F_Z la fonction de répartition de Z. Pour $z \geq 0$, on cherche $F_Z(z) = \mathbb{P}[Z \leq z]$. Notons A l'événement "le fruit est de variété A", et B "le fruit est de variété B". On a (par la formule des probabilité totales),

$$\mathbb{P}[Z \leq z] = \mathbb{P}[(Z \leq z) \cap (A \cup B)] = \mathbb{P}[(Z \leq z \cap A) \cup (Z \leq z \cap B)].$$

Les événements A et B sont disjoints, donc,

$$\mathbb{P}[Z \leq z] = \mathbb{P}[(Z \leq z) \cap A] + \mathbb{P}[(Z \leq z) \cap B].$$

Enfin la (très importante) loi de Bayes, stipule que $\mathbb{P}[(Z \leq z) \cap A] = \mathbb{P}[Z \leq z \mid A]\,\mathbb{P}[A] = pF_X(z)$. Ici F_X dénote bien évidemment la fonction de répatition de X.

De même $\mathbb{P}[(Z \leq z) \cap B] = \mathbb{P}[Z \leq z \mid B]\,\mathbb{P}[B] = (1-p)F_Y(z)$. Finalement,

$$F_Z(z) = pF_X(z) + (1-p)F_Y(z).$$

2) La densité est la dérivé de la fonction de répartition. Soit f_Z la densité de Z, on a donc pour tout $z \geq 0$,

$$f_Z(z) = pf_X(z) + (1-p)f_Y(z).$$

On va ensuite utiliser la définition suivante : Pour toute fonction (sous les bonnes hypothèses), on a :

$$\mathbb{E}[g(Z)] = \int g(z) f_Z(z)\,\mathrm{d}z = \int g(z)(pf_X(z) + (1-p)f_Y(z))\,\mathrm{d}z = p\mathbb{E}[g(X)] + (1-p)\mathbb{E}[g(Y)] \qquad (1)$$

Ainsi on a, (pour $g(x) = x$),

$$\mathbb{E}[Z] = p\mathbb{E}[X] + (1-p)\mathbb{E}[Y] = pm_1 + (1-p)m_2.$$

Pour la variance, Attention, la variance n'est pas linéaire, et en particulier, en général $V(X+Y) \neq V(X) + V(Y)$ (en revanche c'est vrai si X et Y sont indépendants). Ici on ne sait pas s'il y a indépendance donc on fait le calcul.

On va utiliser la formule $V(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$.
Pour trouver $\mathbb{E}[Z^2]$, on a

— $V(X) = \sigma_1^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 = \mathbb{E}\left[X^2\right] - m_1^2$.
— $V(Y) = \sigma_2^2 = \mathbb{E}\left[Y^2\right] - \mathbb{E}\left[Y\right]^2 = \mathbb{E}\left[Y^2\right] - m_2^2$.
Ainsi grâce à la formule (1) avec $g(z) = z^2$, on a

$$\mathbb{E}\left[Z^2\right] = p\mathbb{E}\left[X^2\right] + (1-p)\mathbb{E}\left[Y^2\right] = p(\sigma_1^2 + m_1^2) + (1-p)(\sigma_2^2 + m_2^2).$$

Enfin, on peut calculer $V(Z)$,

$$V(Z) = \mathbb{E}\left[Z^2\right] - \mathbb{E}\left[Z\right]^2 = p(\sigma_1^2 + m_1^2) + (1-p)(\sigma_2^2 + m_2^2) - (pm_1 + (1-p)m_2)^2,$$

et on peut s'amuser à développer.

Exercice 4 : Gestion optimale de stock

Pour la vente d'un magazine spécialisé, un buraliste constitue un stock, au début de chaque mois. On note X la variable aléatoire réelle (VAR) qui représente le nombre de magazines vendus ou demandés à l'achat par les clients auprès de ce buraliste, dans le mois considéré. On peut admettre que X suit une loi uniforme sur $[\![11;30]\!]$, c'est-à-dire que $\forall l, \quad 11 \leq l \leq 30, \quad \mathbb{P}[X = l] = \frac{1}{20}$. Tout magazine non vendu est perdu. Le buraliste achète les magazines 2 euros et les revend 3 euros. On note Y_k la VAR qui représente le gain du commerçant pour un mois donné, en ayant fait un stock de k magazines au début du mois.

1. *Soit k fixé $\in [\![11;30]\!]$, calculer la loi de $Z = \inf(X, k)$.*
2. *En considérant d'abord $k < 11$, puis $11 \leq k \leq 30$, donner la loi de Y_k . Montrer que pour $k \in [\![11;30]\!]$, on a*
$$\mathbb{E}(Y_k) = \frac{1}{40}\left(-3k^2 + 103k - 330\right)$$
3. *Quelle valeur de k conseillez-vous au buraliste ? Justifiez votre réponse.*

1) Dans l'énoncé k représente le stock de journaux achetés par le commerçant pour le mois en cours. La variable $Z = \inf(X, k)$ est donc à valeur dans $[\![11, k]\!]$.
Soit $l \in [\![11, k]\!]$, si $l < k$, $\mathbb{P}\left[Z = l\right] = \mathbb{P}\left[X = l\right] = \frac{1}{20}$. Sinon, si $l = k$, $\mathbb{P}\left[Z = l\right] = \mathbb{P}\left[Z = k\right] = \frac{30-k+1}{20}$.

2) Si $k \leq 10$, le commerçant vend forcément tous ses exemplaires. Donc le bénéfice est $Y_k = (3-2)k = k$.
Si maintenant $k \in [\![11;30]\!]$, alors $Y_k = 3\inf(X, k) - 2k = 3Z - 2k$, et on connait la loi de Z.
Cela nous suffit pour calculer l'espérance de Y_k,

$$\mathbb{E}\left[Y_k\right] = \mathbb{E}\left[3Z - 2k\right] = 3\mathbb{E}\left[Z\right] - 2k.$$

Or par définition,

$$\mathbb{E}\left[Z\right] = \sum_{l=11}^{30} l\mathbb{P}\left[Z = l\right] = \left(\sum_{l=11}^{k-1} l\frac{1}{20}\right) + k\frac{31-k}{20} = \frac{1}{20}\left(\frac{k(k-1)}{2} - 55 + 31k - k^2\right).$$

Ici le -55 vient de $\left(\sum_{l=11}^{k-1} l\right) = \left(\sum_{l=1}^{k-1} l\right) - \left(\sum_{l=1}^{11} l\right) = \frac{k(k-1)}{2} - \frac{11\times 10}{2}$
On trouve donc enfin,

$$\mathbb{E}\left[Y_k\right] = 3\mathbb{E}\left[Z\right] - 2k = \frac{1}{40}\left(-3k^2 + 183k - 330\right) - 2k = \frac{1}{40}\left(-3k^2 + 103k - 330\right).$$

3) Le buraliste veut maximiser le gain qu'il peut espérer avoir (en moyenne). Il veut donc choisir la taille k de son stock de manière à maximiser son gain moyen $\mathbb{E}\left[Y_k\right]$. On fait la traditionnelle étude du maximum sur la fonction $x \mapsto -3x^2 + 103x - 330$. Le maximum est atteint en

$$-6x + 103 = 0 \iff x = \frac{103}{6} \simeq 17,16.$$

Le buraliste a donc intérêt à acheter entre 17 et 18 magazine par mois.