

Rappels de statistique mathématique
Réponses question par question des travaux dirigés n° 8

Guillaume Lacôte
Bureau E03

✉ Guillaume.Lacote@ensae.fr

☞ <http://ensae.no-ip.com/SE222/>

Exercice corrigé 1

Soit un échantillon X_1, \dots, X_n de variables aléatoires i.i.d telles que :

- X_i a une probabilité α de valoir a , et
- une probabilité $(1 - \alpha)$ de suivre une loi normale $\mathcal{N}(0, 1)$,

ce qui s'écrit encore :

$$X_1 = a \cdot \mathbb{1}_{Z_1=1} + Y_1 \cdot \mathbb{1}_{Z_1=0}$$

où $Z_1 \sim \mathcal{B}(1, \alpha)$ est indépendante de $Y_1 \sim \mathcal{N}(0, 1)$.

☞ Q1 Montrer que $(\delta_a + \lambda)^{\otimes n}$ est une mesure dominante pour le modèle considéré.

Posons $\mu = \delta_a + \lambda$ et $\mu_n = \mu^{\otimes n}$.

Remarque : nous restreindrons les démonstrations au cas où une seule observation X_1 est disponible, dans la mesure où car la généralisation à n observations i.i.d. se réduit à un passage au produit immédiat :

$$dP_{X_1, \dots, X_n} = \prod_{i=1}^n dP_{X_i}$$

$$\frac{dP_{X_1, \dots, X_n}}{d\mu_n} = \prod_{i=1}^n \frac{dP_{X_i}}{d\mu}$$

Remarque préliminaire

Pour \mathcal{A} ensemble mesurable, on a

$$\begin{aligned} \mathbb{P}(X_1 \in \mathcal{A}) &= \mathbb{P}(X_1 \in \mathcal{A} | X_1 = a) \mathbb{P}(X_1 = a) + \mathbb{P}(X_1 \in \mathcal{A} | X_1 \neq a) \mathbb{P}(X_1 \neq a) \\ &= \alpha \mathbb{P}(X_1 \in \mathcal{A} | X_1 = a) + (1 - \alpha) \int_{\mathcal{A}} \varphi(x) d\lambda(x) \end{aligned}$$

où φ désigne la densité de la loi normale centrée réduite.

Or $\mathbb{P}(X_1 \in \mathcal{A} | X_1 = a)$ vaut 1 si $a \in \mathcal{A}$ et 0 sinon, de sorte que

$$\mathbb{P}(X_1 \in \mathcal{A}) = \alpha \mathbb{1}_{a \in \mathcal{A}} + (1 - \alpha) \int_{\mathcal{A}} \varphi(x) d\lambda(x)$$

Or μ est une mesure dominante d'une variable X_1 si :

$$\forall \mathcal{A} \text{ mesurable, } \mu(\mathcal{A}) = 0: P_{X_1}(\mathcal{A}) = 0$$

En l'occurrence pour tout \mathcal{A} mesurable :

- $\mu(\mathcal{A}) = 0: \delta_a(\mathcal{A}) = 0$
 - $\delta_a(\mathcal{A}) = 0: a \notin \mathcal{A}$, donc d'après 1 $\lambda(\mathcal{A}) = 0: P_{X_1}(\mathcal{A}) = 0$
- Donc $(\delta_a + \lambda)^{\otimes n}$ est une mesure dominante.

Montrer que :

$$\frac{dP_{X_1, \dots, X_n}(x_1, \dots, x_n)}{d(\delta_a + \lambda)^{\otimes n}} = \prod_{i=1}^n (\alpha \mathbb{1}_a(x_i) + (1 - \alpha)(1 - \mathbb{1}_a(x_i))\phi(x_i))$$

où ϕ est la densité d'une loi normale $\mathcal{N}(0, 1)$.

On a $\mathbb{1}_{a \in \mathcal{A}} = \int_{\mathcal{A}} \delta_a(x)$,
donc en reportant dans l'équation 1 il vient

$$P(X_1 \in \mathcal{A}) = \int_{\mathcal{A}} (\alpha \delta_a(x) + (1 - \alpha)\varphi(x)) d\lambda(x)$$

d'où on tire facilement

$$P(X_1 \in \mathcal{A}) = \int_{\mathcal{A}} (\alpha \mathbb{1}_a(x) + (1 - \alpha)\varphi(x)(1 - \mathbb{1}_a(x))) d\mu(x)$$

de sorte qu'en définitive

$$\frac{dP_{X_1}(x_1)}{d(\delta_a + \lambda)} = \alpha \mathbb{1}_a(x_1) + (1 - \alpha)(1 - \mathbb{1}_a(x_1))\phi(x_1)$$

d'où le résultat par indépendance de (X_1, \dots, X_n) .

Exercice corrigé 2

Cet exercice s'inspire librement des travaux de Dov Samet, 2003.¹

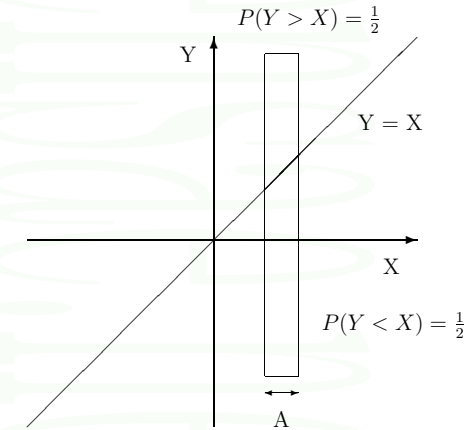
On s'intéresse pour tout couple (X, Y) de variables aléatoires réelles à la propriété \mathcal{P} suivante :

$$(\mathcal{P}) \quad \forall \mathcal{A} \text{ mesurable, } \begin{cases} \mathbb{P}(X < Y \text{ et } X \in \mathcal{A}) = \mathbb{P}(X > Y \text{ et } X \in \mathcal{A}) = \frac{1}{2}\mathbb{P}(X \in \mathcal{A}) \\ \mathbb{P}(Y < X \text{ et } Y \in \mathcal{A}) = \mathbb{P}(Y > X \text{ et } Y \in \mathcal{A}) = \frac{1}{2}\mathbb{P}(Y \in \mathcal{A}) \end{cases}$$

Q1 Comment s'interprète la propriété \mathcal{P} ?

Si un couple de variables aléatoires réelles (X, Y) vérifie \mathcal{P} , alors au voisinage de tout X , la probabilité est la même pour Y d'être au-dessus ou au-dessous de X , comme illustré sur la figure suivante :

¹Résultats présentés à la 14^{ième} conférence internationale de Théorie des jeux, Stony Brook 2003. Voir <http://www.sunysb.edu/gametheory/Conf03/twopuzzles.pdf> pour un exposé plus complet.



■ Partie 1 “Devinez quel est le plus grand !”

On s'intéresse au problème suivant, présenté originellement par Blackwell (1951) :

Deux nombres réels sont tirés aléatoirement, et chacun est placé dans une enveloppe. L'une d'elle est (indépendamment) tirée au hasard et vous est présentée. Vous devez deviner, au vu du nombre qu'elle contient, s'il s'agit du plus grand ou le plus petit des deux. Sauriez-vous deviner juste plus d'une fois sur deux en moyenne ?

Pour simplifier la présentation, supposons que vous gagnez +1 si vous avez deviné juste et sinon ; ce problème revient à savoir si vous pouvez vous garantir un gain espéré strictement positif.

Q1 Soient X et Y les deux nombres tirés aléatoirement, et supposons que (X, Y) vérifie \mathcal{P} . Quel est votre meilleur gain espéré ? Interpréter.

Soit $\sigma : \mathbb{R} \rightarrow \{-1, +1\}$ votre stratégie, et soient (x, y) une réalisation du tirage.

- Si x vous est présenté :
Si $x > y$, votre gain est $\sigma(x)$; si $x < y$, votre gain est $-\sigma(x)$. Votre gain espéré, sachant que x vous est présenté, est donc nul.
- Si y vous est présenté, il en va de même.

Le choix de l'enveloppe étant indépendant du tirage de X et de Y , votre gain espéré est toujours nul, quelle que soit votre stratégie.

Comme l'hypothèse \mathcal{P} semble “raisonnable”, l'intuition suggère qu'à ce jeu vous ne pouvez jamais obtenir un gain espéré strictement positif. Cette intuition est cependant erronée.

Q2 Cherchons néanmoins à construire une stratégie gagnante.

Pour ce faire, donnons-nous $s \in \mathbb{R}$ et soit $\sigma_s : \begin{pmatrix} \mathbb{R} & \rightarrow & \{-1, 1\} \\ x & \mapsto & \begin{cases} 1 & \text{si } x > s \\ -1 & \text{sinon} \end{cases} \end{pmatrix}$ la stratégie de seuil associée à s .

(a) Soit $(x \neq y)$ une réalisation du tirage; le nombre qui vous est présenté est soit x , soit y . Quel est votre gain espéré en jouant selon σ_s sachant que $x < s$ et $y < s$?

On a dans tous les cas $\sigma(x) = \sigma(y) = -1$ car $x < s$ et $y < s$.

Or si $x < y$, le gain espéré est

$$\mathbb{P}(\text{le nombre présenté est } x) \cdot 1 + \mathbb{P}(\text{le nombre présenté est } y) \cdot (-1) = 0$$

puisque ces deux alternatives sont équiprobables et que le choix de l'enveloppe est indépendant du tirage de X et de Y .

Il en va de même lorsque $x > y$, de sorte que le gain espéré, sachant que $x < s$ et $y < s$, est nul.

(b) Quel est votre gain espéré sachant que $x < s < y$?

Si le nombre présenté est x , on a $\sigma(x) = -1$ car $x < s$; or $x < y$, donc la réponse est juste et le gain est $+1$.

Si le nombre présenté est y , on a $\sigma(y) = +1$ car $y > s$; or $y > x$, donc la réponse est encore juste et le gain est toujours $+1$.

Ainsi dans tous les cas le gain espéré, sachant que $x < s < y$, est $+1$. Il en va de même bien-sûr sachant que $y < s < x$.

(c) En déduire que si votre gain espéré (sur tous les tirages possibles) est strictement positif.

D'après les résultats précédents le gain espéré en jouant σ_s est donc

$$g(s) = \mathbb{P}(X < s < Y) + \mathbb{P}(Y < s < X)$$

Or ces deux probabilités ne peuvent être simultanément nulles pour tout s : en effet

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{P}(X < s < Y) + \mathbb{P}(Y < s < X) ds &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^2} (\mathbb{1}_{x < s < y} + \mathbb{1}_{y < s < x}) f_{X,Y}(x,y) dx dy \right) ds \\ &= \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}} \mathbb{1}_{x < s < y} ds + \int_{\mathbb{R}} \mathbb{1}_{y < s < x} ds \right) f_{X,Y}(x,y) dx dy \\ &= \int_{\mathbb{R}^2} (\mathbb{1}_{x < y} (y - x) + \mathbb{1}_{y < x} (x - y)) f_{X,Y}(x,y) dx dy \\ &= \int_{\mathbb{R}^2} |x - y| \mathbb{1}_{x \neq y} f_{X,Y}(x,y) dx dy \\ &= \mathbb{E}(|X - Y|) \end{aligned}$$

Si $\mathbb{P}(X \neq Y) = 0$ alors $\mathbb{P}(X = Y) = 1$ et le jeu devient trivial : la stratégie qui présente le nombre tiré est le plus grand est presque sûrement toujours gagnante; en particulier lorsque $s = 0$ $g(0) > 0$.

Et si $\mathbb{P}(X \neq Y) \neq 0$ alors $\mathbb{E}(|X - Y|) > 0$, et donc $\int_{\mathbb{R}} g(s) ds > 0$.

Donc dans tous les cas

$$\exists s \in \mathbb{R} / g(s) > 0$$

En fait, si (X, Y) est distribué de telle sorte que toute partie de mesure non-nulle a une probabilité non-nulle, alors n'importe quel seuil s assure un gain espéré strictement positif.

Q3 En déduire qu'aucune paire de variables aléatoires réelles (X, Y) ne peut vérifier \mathcal{P} .

Si une telle distribution sur \mathbb{R}^2 existait, alors d'après la question (1) dans le jeu de Blackwell associé nul ne saurait gagner plus que zéro en moyenne, ce qui est pourtant possible d'après la question (2).

■ Partie 2 Le jeu des enveloppes

On s'intéresse désormais au célèbre jeu dit *des enveloppes*, introduit par Kraitchik (1953) sous la forme suivante :

On considère deux enveloppes contenant une certaine somme d'argent, mais l'une contenant le double de l'autre; chacune a autant de chances que l'autre de contenir la plus grosse somme. L'une d'elles vous est donnée au hasard, et vous devez choisir entre la prendre, ou prendre l'autre.

A supposer qu'elle contienne x , l'autre a une chance sur deux de contenir $2x$ et une chance sur deux de contenir $\frac{x}{2}$; donc le contenu espéré de l'autre enveloppe est $\frac{5}{4}x > x$, de sorte que vous avez intérêt à changer d'enveloppe. Mais le même raisonnement s'applique aussi à la nouvelle enveloppe, de sorte que vous avez à nouveau intérêt à changer d'enveloppe, et ainsi de suite ...

Que faire dans cette situation?

Résoudre ce paradoxe a été le prétexte à une littérature foisonnante; voici une solution.

Notons X la somme (variable aléatoire réelle) contenue dans l'enveloppe qui vous est présentée et Y celle contenue dans l'autre.

Q1 Quel sens donner à l'expression \mathcal{H}_x : "l'autre a une chance sur deux de contenir $2x$ et une chance sur deux de contenir $\frac{x}{2}$ "?

La formulation du paradoxe suggère que l'on se donne une distribution de probabilité a priori sur les deux sommes X et Y ; cette distribution est nécessairement telle que $\mathbb{P}(X = 2Y \text{ ou } Y = 2X) = 1$.

Etant donné $x \in \mathbb{R}$ et à supposer que $X = x$, l'hypothèse \mathcal{H}_x s'exprime en termes mathématiques sous la forme

$$\mathbb{P}(Y = 2x | X = x) = \mathbb{P}\left(Y = \frac{x}{2} | X = x\right) = \frac{1}{2}$$

Q2 En déduire que si cette hypothèse est vraie, alors la distribution de (X, Y) vérifie \mathcal{P} .
L'hypothèse \mathcal{H}_y analogue à la précédente en $Y = y \in \mathbb{R}$ s'écrit

$$\mathbb{P}(X = 2y | Y = y) = \mathbb{P}\left(X = \frac{y}{2} | Y = y\right) = \frac{1}{2}$$

Dans l'énoncé du paradoxe, l'hypothèse selon laquelle "chacune a autant de chances que l'autre de contenir la plus grosse somme" se traduit alors en \mathcal{H}_x quelle que soit la somme x de la première enveloppe, et \mathcal{H}_y quelle que soit la somme y de la deuxième enveloppe.

En d'autres termes

$$\forall x \in \mathbb{R}, \mathbb{P}(Y = 2x | X = x) = \mathbb{P}\left(Y = \frac{x}{2} | X = x\right) = \frac{1}{2}$$

$$\forall y \in \mathbb{R}, \mathbb{P}(X = 2y | Y = y) = \mathbb{P}\left(X = \frac{y}{2} | Y = y\right) = \frac{1}{2}$$

Or pour tout $x \in \mathbb{R}$ l'événement $Y = 2x$ est égal à l'événement $Y > x$, et l'événement $Y = \frac{x}{2}$ à $Y < x$. Par suite

$$\forall x \in \mathbb{R}, \mathbb{P}(Y > x | X = x) = \mathbb{P}(Y < x | X = x) = \frac{1}{2}$$

$$\forall y \in \mathbb{R}, \mathbb{P}(X > y | Y = y) = \mathbb{P}(X < y | Y = y) = \frac{1}{2}$$

Par conséquent pour toute partie mesurable \mathcal{A} de \mathbb{R} telle que $\mathbb{P}(X \in \mathcal{A}) \neq 0$ et $\mathbb{P}(Y \in \mathcal{A}) \neq 0$ il vient

$$\mathbb{P}(Y > X | X \in \mathcal{A}) = \mathbb{P}(Y < X | X \in \mathcal{A}) = \frac{1}{2}$$

$$\mathbb{P}(X > Y | Y \in \mathcal{A}) = \mathbb{P}(X < Y | Y \in \mathcal{A}) = \frac{1}{2}$$

de sorte que (X, Y) vérifie la propriété \mathcal{P} .

Q3 Conclure.

Ainsi, à supposer que l'on se donne une distribution a priori sur les sommes d'argent de chacune des deux enveloppes, il n'est pas possible de traduire l'hypothèse "chacune a autant de chances que l'autre de contenir la plus grosse somme" comme il est suggéré, car la distribution vérifierait alors la propriété \mathcal{P} ce qui n'est pas possible.

La démarche proposée, bien qu'intuitive, n'est donc pas raisonnable : en particulier il est faux qu'après un nombre quelconque de changements d'enveloppe on ait "à nouveau intérêt à changer d'enveloppe".

Exercice corrigé 3

Soient X_1, \dots, X_n n variables aléatoires i.i.d dans \mathbb{R}^k , de loi paramétrique \mathcal{P}_θ de densité

$$\frac{d\mathcal{P}_\theta(x)}{d\lambda} = f(x - \theta)$$

$\theta \in \mathbb{R}^k$ est appelé *paramètre de position*.

On cherche à étudier les estimateurs T du paramètre θ . On se restreint pour cela à la classe d'estimateurs *équivariants*, c'est-à-dire des estimateurs T vérifiant :

$$\forall c \in \mathbb{R}^k, T(X_1 + c, \dots, X_n + c) = T(X_1, \dots, X_n) + c$$

Cette restriction est naturelle : en effet, pour tout $c \in \mathbb{R}^k$, $X_i + c \sim \mathcal{P}_{\theta+c}$, et on ne peut admettre qu'un simple changement d'échelle puisse mener à une estimation différente.

Q1 Montrer que :
T équivariant $\Leftrightarrow \exists T_1 : (\mathbb{R}^{k-1} \rightarrow \mathbb{R}^k) / \forall (x_1, \dots, x_n) \in \mathbb{R}^k$
 $T(x_1, \dots, x_n) = T_1(x_2 - x_1, \dots, x_n - x_1) + x_1$

Il suffit de poser $c = -x_1$ dans la définition de l'équivariance.

Q2 Soit $W(x, y) = w(x - y)$ une fonction de perte. Montrer que, si un estimateur T est équivariant, le risque associé, pour la perte W , ne dépend pas de θ .
Un estimateur équivariant T qui minimise le risque $R_w(T, 0)$ (c'est-à-dire au point $\theta = 0$) est alors un estimateur optimal (parmi les équivariants) pour la fonction de perte W .

Le risque associé vérifie :

$$\begin{aligned} R_w(T, \theta) &= E_\theta[w(T - \theta)] \\ &= \int_{(\mathbb{R}^k)^n} w(T(x_1, \dots, x_n) - \theta) \prod_{i=1}^n f(x_i - \theta) dx_i \\ &= \int_{(\mathbb{R}^k)^n} w(T(x_1 - \theta, \dots, x_n - \theta)) \prod_{i=1}^n f(x_i - \theta) dx_i \\ &= \int_{(\mathbb{R}^k)^n} w(T(u_1, \dots, u_n)) \prod_{i=1}^n f(u_i) du_i \\ &= R_w(T, 0) \end{aligned}$$

en effectuant un changement de variable $u_i = x_i - \theta$.

Q3 Soit la fonction

$$\psi(x) = \int_{\mathbb{R}^k} w(x - u) \prod_{i=1}^n f(X_i - u) du$$

et l'estimateur T^* défini par

$$\psi(T^*) = \inf_{x \in \mathbb{R}^k} \psi(x)$$

T^* est appelé *l'estimateur de Pitman* (on suppose ici que w est telle que l'équation $\psi(x^*) = \inf_{x \in \mathbb{R}^k} \psi(x)$ admette une solution et une seule).

(a) Montrer que T^* est équivariant.

Soit $c \in \mathbb{R}^k$; on a

$$\begin{aligned} \psi(x+c) &= \int w(x+c-u) \prod_{i=1}^n f(X_i-u) du \\ &= \int w(x-v) \prod_{i=1}^n f(X_i-c-v) dv \end{aligned}$$

en effectuant le changement de variable $v = u - c$.

En particulier $\psi(x + X_1) = \int w(x - v) f(-v) \prod_{i=2}^n f((X_i - X_1) - v) dv$

Donc $\psi(x + X_1)$ ne dépend que de x et des variables aléatoires $X_2 - X_1, \dots, X_n - X_1$:

soit donc T_1 tel que $\arg \min_{x \in \mathbb{R}^k} \psi(x + X_1) = T_1(X_2 - X_1, \dots, X_n - X_1)$.

Or $\forall \phi, \forall c \in \mathbb{R}^k, \arg \inf (x \mapsto \phi(x+c)) = (\arg \inf (x \mapsto \phi(x))) - c$.

Donc en définitive

$$T_1(X_2 - X_1, \dots, X_n - X_1) = (\arg \inf \psi) - X_1 = T^* - X_1$$

Autrement dit T^* est équivariant d'après la question précédente.

Montrer que, pour toute fonction φ telle que $\mathbb{E}(\|\varphi(X_1, \dots, X_n)\|) < +\infty$, on a :

$$E_\theta[\varphi(X_1, \dots, X_n) | X_2 - X_1, \dots, X_n - X_1] = \int_{\mathbb{R}^k} \varphi(X_1 + \theta - u, \dots, X_n + \theta - u) \frac{\prod_{i=1}^n f(X_i - u)}{\int_{\mathbb{R}^k} \prod_{i=1}^n f(X_i - v) dv} du$$

$$\begin{aligned} &E_\theta[\varphi(X_1, \dots, X_n) | X_2 - X_1 = y_2, \dots, X_n - X_1 = y_n] \\ &= \int \varphi(x_1, x_1 + y_2, \dots, x_1 + y_n) d\mathbb{P}(X_1 = x_1 | X_2 - X_1 = y_2, \dots, X_n - X_1 = y_n) \\ &= \int \varphi(x_1, x_1 + y_2, \dots, x_1 + y_n) \frac{f(x_1 - \theta) \prod_{i=2}^n f(x_1 + y_i - \theta)}{\int f(x_1 - \theta) \prod_{i=2}^n f(x_1 + y_i - \theta) dx_1} dx_1 \end{aligned}$$

Effectuons le changement de variables $u = \theta - x_1 + c$ (c constante quelconque) dans les deux intégrales :

$$\begin{aligned} &E_\theta[\varphi(X_1, \dots, X_n) | X_2 - X_1 = y_2, \dots, X_n - X_1 = y_n] \\ &= \int \varphi(c + \theta - u, c + y_2 + \theta - u, \dots) \frac{f(c - u) \prod_{i=2}^n f(c + y_i - u)}{\int f(c - u) \prod_{i=2}^n f(c + y_i - u) du} du \end{aligned}$$

La relation suivante est alors vraie pour tout $c \in \mathbb{R}^k$:

$$\begin{aligned} &E_\theta[\varphi(X_1, \dots, X_n) | X_2 - X_1, \dots, X_n - X_1] \\ &= \int \varphi(c + \theta - u, c + X_2 - X_1 + \theta - u, \dots) \frac{f(c - u) \prod_{i=2}^n f(c + X_i - X_1 - u)}{\int f(c - u) \prod_{i=2}^n f(c + X_i - X_1 - u) du} \end{aligned}$$

elle est donc vraie aussi pour $c = X_1$:

$$\begin{aligned} &E_\theta[\varphi(X_1, \dots, X_n) | X_2 - X_1, \dots, X_n - X_1] \\ &= \int \varphi(X_1 + \theta - u, X_2 + \theta - u, \dots) \frac{\prod_{i=1}^n f(X_i - u)}{\int f(X_1 - u) \prod_{i=2}^n f(X_i - u) du} du \end{aligned}$$

(c) Montrer finalement que T^* est optimal pour w .

Soit T un estimateur équivariant quelconque. On va montrer que T^* domine T . Pour ce il suffit de prouver que $R_w(T, 0) \geq R_w(T^*, 0)$, selon le résultat de la question (2).

$$\begin{aligned} R_w(T, 0) &= \mathbb{E}_0[w(T)] \\ &= \mathbb{E}_0[\mathbb{E}_0[w(T) | X_2 - X_1, \dots, X_n - X_1]] \\ &= \mathbb{E}_0 \left[\int_{\mathbb{R}} w(T - u) \frac{\prod_{i=1}^n f(X_i - u)}{\int_{\mathbb{R}} \prod_{i=1}^n f(X_i - u) du} du \right] \end{aligned}$$

Or par définition de T^* , pour tous $x \in \mathbb{R}^n$ et $t \in \mathbb{R}$

$$\int_{\mathbb{R}} w(T^*(x) - u) \prod_{i=1}^n f(x_i - u) du \leq \int_{\mathbb{R}} w(t - u) \prod_{i=1}^n f(x_i - u) du$$

donc

$$\begin{aligned} R_w(T, 0) &\geq \mathbb{E}_0 \left[\int_{\mathbb{R}} w(T^*(X) - u) \frac{\prod_{i=1}^n f(X_i - u)}{\int_{\mathbb{R}} \prod_{i=1}^n f(X_i - u) du} du \right] \\ &= R_w(T^*, 0) \end{aligned}$$

Donc T^* domine T .

Q4 Donner l'expression de T^* lorsque $w(x - y) = \|x - y\|^2$.

Si $w(x - y) = \|x - y\|^2$, ψ vaut :

$$\psi(x) = \int_{\mathbb{R}^k} \|x - u\|^2 \prod_{i=1}^n f(X_i - u) du$$

et le gradient $\psi'(x)$ vaut : $\psi'(x) = \int_{\mathbb{R}^k} 2(x - u) \prod_{i=1}^n f(X_i - u) du$. T^* est défini par l'équation $\psi'(T^*) = 0$, qui se réduit ici à :

$$T^* = \frac{\int_{\mathbb{R}^k} u \prod_{i=1}^n f(X_i - u) du}{\int_{\mathbb{R}^k} \prod_{i=1}^n f(X_i - u) du}$$

☞ Q5

Dans ce cas, que vaut T^* lorsque $X_i \sim \mathcal{N}(\theta, 1)$?
 Et lorsque $X_i \sim \mathcal{U}_{[-\frac{1}{2}+\theta, \frac{1}{2}+\theta]}$?

Lorsque $X_i \sim \mathcal{N}(0, 1)$, on a

$$T^* = \frac{1}{n} \sum_{i=1}^n X_i$$

Si $X_i \sim \mathcal{U}(0, 1)$, il vient

$$T^* = \frac{\min(X_i) + \max(X_i)}{2}$$

Exercice corrigé 4

On considère une loi multinomiale à K modalités ($K \geq 3$) de probabilités p_1, p_2, \dots, p_K . On dispose de n observations indépendantes issues de cette loi. On notera N_k le nombre d'observations de la modalité $k \in \llbracket 1, K \rrbracket$.

On veut tester l'hypothèse : $H_0 : "p_1 + p_2 = \frac{1}{2}"$

☞ Q1

Donner les estimateurs du maximum de vraisemblance contraints $(\hat{p}_k^0)_k$ et non contraints $(\hat{p}_k)_k$ de p_1, \dots, p_K .

Les observations étant les effectifs N_k de chaque classe $k \in \llbracket 1, K \rrbracket$, la vraisemblance s'écrit

$$L_{N_1, \dots, N_K}(n_1, \dots, n_K; p_1, \dots, p_K) = \frac{n!}{n_1! \dots n_K!} \prod_{k=1}^K p_k^{n_k} \quad \text{en notant } n = n_1 + \dots + n_K$$

La maximisation sans hypothèse conduit aux conditions nécessaires du premier ordre, en n'omettant la contrainte $p_1 + \dots + p_K = 1$

$$\begin{aligned} & \max_{p_1, \dots, p_K} \left(\sum_{k=1}^K n_k \ln p_k \right) \\ & \text{s.c. } \{ p_1 + \dots + p_K = 1 \end{aligned}$$

Le lagrangien associé s'écrit

$$\mathcal{L}(p_1, \dots, p_K, \lambda) = \sum_{k=1}^K n_k \ln p_k + \lambda(1 - \sum_{k=1}^K p_k)$$

et il vient

$$\forall k \in \llbracket 1, K \rrbracket, \quad \frac{n_k}{\hat{p}_k} - \lambda = 0$$

Or en sommant de 1 à K il vient

$$n = \sum_{k=1}^K n_k = \lambda \sum_{k=1}^K p_k = \lambda$$

de sorte que finalement

$$\forall k \in \llbracket 1, K-1 \rrbracket, \quad \hat{p}_k = \frac{n_k}{n}$$

(et on vérifie que cette condition est également suffisante).

La maximisation sous l'hypothèse H_0 correspond au programme de maximisation sous contrainte

$$\begin{aligned} & \max_{p_1, \dots, p_K} \left(\sum_{k=1}^K n_k \ln p_k \right) \\ & \text{s.c. } \begin{cases} p_1 + \dots + p_K = 1 \\ p_1 + p_2 = \frac{1}{2} \end{cases} \end{aligned}$$

Le lagrangien associé s'écrit

$$\mathcal{L}(p_1, \dots, p_K, \lambda, \mu) = \sum_{k=1}^K n_k \ln p_k + \lambda(1 - \sum_{k=1}^K p_k) + \mu(\frac{1}{2} - p_1 - p_2)$$

et il vient

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial p_1} = \frac{n_1}{p_1} - \lambda - \mu \\ \frac{\partial \mathcal{L}}{\partial p_2} = \frac{n_2}{p_2} - \lambda - \mu \\ \frac{\partial \mathcal{L}}{\partial p_k} = \frac{n_k}{p_k} - \lambda, \quad \forall k \in \llbracket 3, K \rrbracket \\ \frac{\partial \mathcal{L}}{\partial \lambda} = 1 - \sum_{k=1}^K p_k \\ \frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{2} - p_1 - p_2 \end{cases}$$

d'où finalement

$$\begin{cases} \hat{p}_1^0 = \frac{n_1}{2(n_1+n_2)} \\ \hat{p}_2^0 = \frac{n_2}{2(n_1+n_2)} \\ \hat{p}_k^0 = \frac{n_k}{2(n_3+\dots+n_K)}, \quad \forall k \in \llbracket 3, K \rrbracket \end{cases}$$

☞ Q2

Calculer la statistique ξ^W du test de Wald de l'hypothèse H_0 .

Pour déterminer la statistique de Wald, appliquons le Théorème Central Limite à $\begin{pmatrix} \hat{p}_1 \\ \vdots \\ \hat{p}_K \end{pmatrix}$

pour ce faire, calculons tout d'abord sa variance.

On a en effet pour $k \in \llbracket 1, K \rrbracket$, $\hat{p}_k = \frac{n_k}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \in C_k}$ où C_k désigne la k -ième classe ; or

$$\mathbb{V}(\mathbb{1}_{X_i \in C_k}) = p_k \cdot 1 - (p_k)^2 = p_k(1 - p_k)$$

et

$$\text{Cov}(\mathbf{1}_{X_i \in C_k}, \mathbf{1}_{X_j \in C_l}) = \mathbb{E}(\mathbf{1}_{X_i \in C_k} \cdot \mathbf{1}_{X_j \in C_l}) - \mathbb{E}(\mathbf{1}_{X_i \in C_k})\mathbb{E}(\mathbf{1}_{X_j \in C_l}) = -p_k p_l \mathbb{1}_{i=j}$$

de sorte que

$$\sqrt{n} \begin{pmatrix} \hat{p}_1 - p_1 \\ \vdots \\ \hat{p}_K - p_K \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \cdots & -p_1 p_K \\ -p_2 p_1 & p_2(1-p_2) & \cdots & -p_2 p_K \\ \vdots & \vdots & \ddots & \vdots \\ -p_K p_1 & -p_K p_2 & \cdots & p_K(1-p_K) \end{pmatrix} \right)$$

Posons finalement $g : \begin{pmatrix} \mathbb{R}^K & \rightarrow & \mathbb{R} \\ x_1, \dots, x_K & \mapsto & x_1 + x_2 \end{pmatrix}$; alors d'après le théorème de Slutsky, si H_0 est vraie on a

$$\begin{aligned} \sqrt{n} \left(\hat{p}_1 + \hat{p}_2 - \frac{1}{2} \right) &\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \left(\frac{\partial g}{\partial p_1, \dots, p_K} \right)' \mathbb{V} \left(\begin{pmatrix} \hat{p}_1 \\ \vdots \\ \hat{p}_K \end{pmatrix} \right) \left(\frac{\partial g}{\partial p_1, \dots, p_K} \right) \right) \\ &= \mathcal{N} \left(0, (p_1 + p_2)(1 - p_1 - p_2) \right) \end{aligned}$$

d'où en centrant et en réduisant, puis en élevant au carré

$$\xi_n^W = n \frac{(\hat{p}_1 + \hat{p}_2 - \frac{1}{2})^2}{(\hat{p}_1 + \hat{p}_2)(1 - \hat{p}_1 - \hat{p}_2)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi_1^2$$

☞ Q3

En constatant que \hat{p}_3^0 est asymptotiquement efficace sous H_0 , montrer que

$$\text{Cov}_{as}(\hat{p}_3, \hat{p}_3 - \hat{p}_3^0) = 0$$

(a)

et en déduire que

$$\mathbb{V}_{as}(\hat{p}_3 - \hat{p}_3^0) = \mathbb{V}_{as}(\hat{p}_3) - \mathbb{V}_{as}(\hat{p}_3^0)$$

On sait que $\hat{p}_3^0 \xrightarrow[n \rightarrow +\infty]{p.s.} p_3$ et que $\hat{p}_3 \xrightarrow[n \rightarrow +\infty]{p.s.} p_3$.

Soit donc pour $\lambda \in \mathbb{R}$ $p_3(\lambda) = \lambda \hat{p}_3^0 + (1 - \lambda) \hat{p}_3 \xrightarrow[n \rightarrow +\infty]{p.s.} p_3$.

Alors

$$\begin{aligned} \mathbb{V}_{as}(p_3(\lambda)) &= \mathbb{V}_{as}(\hat{p}_3^0 + \lambda(\hat{p}_3 - \hat{p}_3^0)) \\ &= \mathbb{V}_{as}(\hat{p}_3^0) + 2\lambda \text{Cov}_{as}(\hat{p}_3^0, \hat{p}_3 - \hat{p}_3^0) + \lambda^2 \mathbb{V}_{as}(\hat{p}_3 - \hat{p}_3^0) \\ &\geq \mathbb{V}_{as}(\hat{p}_3^0) \end{aligned}$$

car \hat{p}_3^0 est asymptotiquement efficace, car sa variance est égale à la borne FDCR $I(p_3)^{-1}$.

Par conséquent le polynôme du second degré $P(X) = \mathbb{V}_{as}(\hat{p}_3^0 - \hat{p}_3)X^2 + 2\text{Cov}_{as}(\hat{p}_3^0, \hat{p}_3 - \hat{p}_3^0)X + \mathbb{V}_{as}(\hat{p}_3^0) - \mathbb{V}_{as}(\hat{p}_3)$ est de signe constant. En particulier,

$$\text{Cov}_{as}(\hat{p}_3^0, \hat{p}_3 - \hat{p}_3^0) = 0$$

et donc

$$\begin{aligned} \mathbb{V}_{as}(\hat{p}_3) &= \mathbb{V}_{as}(\hat{p}_3^0 - (\hat{p}_3^0 - \hat{p}_3)) \\ &= \mathbb{V}_{as}(\hat{p}_3^0) + \mathbb{V}_{as}(\hat{p}_3 - \hat{p}_3^0) \end{aligned}$$

ce qui s'écrit encore

$$\mathbb{V}_{as}(\hat{p}_3 - \hat{p}_3^0) = \mathbb{V}_{as}(\hat{p}_3) - \mathbb{V}_{as}(\hat{p}_3^0)$$

(b) Calculer la loi limite de $(\hat{p}_3^0 - p_3)$, et en déduire $\mathbb{V}_{as}(\hat{p}_3^0)$.

La loi asymptotique de $(\hat{p}_3 - \hat{p}_3^0)$ se déduit de celle de $\begin{pmatrix} \hat{p}_1 \\ \vdots \\ \hat{p}_K \end{pmatrix}$ déterminée précédemment.

Soit en effet $g : \begin{pmatrix} \mathbb{R}^K & \rightarrow & \mathbb{R} \\ x_1, \dots, x_K & \mapsto & \frac{x_3}{2(1-x_1-x_2)} \end{pmatrix}$; ainsi $g(\hat{p}_3) = \hat{p}_3^0$. Alors

$$\begin{aligned} \frac{\partial g}{\partial x_1}(p_1, \dots, p_K) &= \frac{p_3}{2(1-p_1-p_2)^2} \\ &= 2p_3 \end{aligned}$$

$$\begin{aligned} \frac{\partial g}{\partial x_2}(p_1, \dots, p_K) &= \frac{p_3}{2(1-p_1-p_2)^2} \\ &= 2p_3 \end{aligned}$$

$$\frac{\partial g}{\partial x_3}(p_1, \dots, p_K) = 1$$

$$\frac{\partial g}{\partial x_k}(p_1, \dots, p_K) = 0, \quad \forall k \in \llbracket 3, K \rrbracket$$

Donc d'après le théorème de Slutsky

$$\begin{aligned}
\sqrt{n} \left(\widehat{p}_3^0 - p_3 \right) &\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left(0, \begin{pmatrix} 2p_3 \\ 2p_3 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}' \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_K \\ -p_2p_1 & p_2(1-p_2) & \cdots & -p_2p_K \\ \vdots & \vdots & \ddots & \vdots \\ -p_Kp_1 & -p_Kp_2 & \cdots & p_K(1-p_K) \end{pmatrix} \begin{pmatrix} 2p_3 \\ 2p_3 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \\
&= \mathcal{N} \left(0, \begin{pmatrix} 2p_3 \\ 2p_3 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}' \times \begin{pmatrix} 2p_3p_1(1-p_1) - 2p_1p_2p_3 - p_1p_3 \\ -2p_1p_2p_3 + 2p_3p_2(1-p_2) - p_2p_3 \\ -2p_3^2p_1 - 2p_3^2p_2 + p_3(1-p_3) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \\
&= \mathcal{N} \left(0, \begin{cases} 4p_3^2p_1(1-p_1) - 4p_1p_2p_3^2 - 2p_1p_3^2 \\ -4p_1p_2p_3^2 + 4p_3^2p_2(1-p_2) - 2p_2p_3^2 \\ -2p_3^2p_1 - 2p_3^2p_2 + p_3(1-p_3) \end{cases} \right) \\
&= \mathcal{N} \left(0, \begin{cases} [4p_1(1-p_1) - 4p_1p_2 - 2p_1 - 4p_1p_2 + \\ 4p_2(1-p_2) - 2p_2 - 2p_1 - 2p_2] p_3^2 \\ -p_3^2 + p_3 \end{cases} \right) \\
&= \mathcal{N} \left(0, \{-4(p_1 + p_2)^2 p_3^2 - p_3^2 + p_3\} \right) \\
&= \boxed{\mathcal{N}(0, p_3(1-2p_3))} \text{ sous } H_0 : " p_1 + p_2 = \frac{1}{2} "
\end{aligned}$$

Ainsi $\boxed{\mathbb{V}_{as}(\widehat{p}_3^0) = p_3(1-2p_3)}$.

(c)

Donner un estimateur $\mathbb{V}_{as}(\widehat{p}_3 - \widehat{p}_3^0)$ convergent sous H_0 de $\mathbb{V}_{as}(\widehat{p}_3 - \widehat{p}_3^0)$.
En déduire l'expression de la statistique du test d'Hausman de H_0

$$\xi^H = n \frac{(\widehat{p}_3 - \widehat{p}_3^0)^2}{\mathbb{V}_{as}(\widehat{p}_3 - \widehat{p}_3^0)}$$

On a

$$\begin{aligned}
\mathbb{V}_{as}(\widehat{p}_3 - \widehat{p}_3^0) &= \mathbb{V}_{as}(\widehat{p}_3) - \mathbb{V}_{as}(\widehat{p}_3^0) \\
&= p_3(1-p_3) - p_3(1-2p_3) \\
&= \boxed{p_3^2}
\end{aligned}$$

Posons donc $\mathbb{V}_{as}(\widehat{p}_3 - \widehat{p}_3^0) = \widehat{p}_3^2 \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{V}_{as}(\widehat{p}_3 - \widehat{p}_3^0)$. On a par ailleurs

$$\begin{aligned}
\widehat{p}_3 - \widehat{p}_3^0 &= \frac{n_3}{n} - \frac{n_3}{2(n-n_1-n_2)} \\
&= \frac{n_3}{n} \left(1 - \frac{1}{2(1-\frac{n_1}{n}-\frac{n_2}{n})} \right) \\
&= \widehat{p}_3 \left(1 - \frac{1}{2(1-\widehat{p}_1-\widehat{p}_2)} \right) \\
&= \widehat{p}_3 \frac{1-2\widehat{p}_1-2\widehat{p}_2}{2-2\widehat{p}_1-2\widehat{p}_2}
\end{aligned}$$

de sorte que

$$\begin{aligned}
\xi_n^H &= n \frac{\widehat{p}_3^2 \left(\frac{1-2\widehat{p}_1-2\widehat{p}_2}{2-2\widehat{p}_1-2\widehat{p}_2} \right)^2}{\widehat{p}_3^2} \\
&= \boxed{n \left(\frac{1-2\widehat{p}_1-2\widehat{p}_2}{2-2\widehat{p}_1-2\widehat{p}_2} \right)^2} \\
&\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \lambda_1^2
\end{aligned}$$

(d) $\boxed{\text{Vérier que le test d'Hausman est convergent.}}$

Si H_0 est fautive, alors $p_1 + p_2 \neq \frac{1}{2}$ et donc

$$\begin{aligned}
\frac{1}{n} \xi_n^H &= \left(\frac{1-2\widehat{p}_1-2\widehat{p}_2}{2-2\widehat{p}_1-2\widehat{p}_2} \right)^2 \\
&\xrightarrow[n \rightarrow +\infty]{p.s.} \left(\frac{p_1 + p_2 - \frac{1}{2}}{1 - p_1 - p_2} \right) \\
&> 0
\end{aligned}$$

de sorte que

$$\xi_n^H \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} (+\infty)$$

Ainsi, si H_0 est fautive alors $\mathbb{P}(\xi_n^H < q_{1-\alpha}^{\chi_1^2}) \xrightarrow[n \rightarrow +\infty]{} \mathbb{P}((+\infty) < q_{1-\alpha}^{\chi_1^2}) = 0$:
le test d'Hausman est donc convergent.

(e) $\boxed{\text{Montrer que les statistiques } \xi^W \text{ et } \xi^H \text{ sont asymptotiquement équivalentes sous } H_0.}}$

On a enfin

$$\begin{aligned} \frac{\xi_n^H}{\xi_n^W} &= \frac{n \left(\frac{1-2\hat{p}_1-2\hat{p}_2}{2-2\hat{p}_1-2\hat{p}_2} \right)^2}{n \frac{(\hat{p}_1+\hat{p}_2-\frac{1}{2})^2}{(\hat{p}_1+\hat{p}_2)(1-\hat{p}_1-\hat{p}_2)}} \\ &= \frac{\hat{p}_1 + \hat{p}_2}{1 - \hat{p}_1 - \hat{p}_2} \\ &\xrightarrow[n \rightarrow +\infty]{p.s.} (1) \end{aligned}$$

Ainsi, les deux statistiques de test ξ_n^H et ξ_n^W sont asymptotiquement équivalentes.

Exercice corrigé 5

La mise en œuvre de nombreuses techniques d'analyse, et notamment la quasi-totalité des techniques d'estimation et de test de modèles statistiques, nécessitent d'engendrer des réalisations x d'une variable aléatoire réelle X de loi \mathcal{L} donnée.

L'objet de cet exercice est de présenter quelques techniques permettant à un ordinateur, machine essentiellement déterministe, d'engendrer de telles variables. Il s'agit donc de construire pour toute loi \mathcal{L} une fonction récursive $f^\mathcal{L} : \mathbb{N} \rightarrow I$ qui énumère des réalisations d'un variable $X \rightsquigarrow \mathcal{L}$, i.e. telle que la distribution empirique des $(f^\mathcal{L}(n))_{n \in \mathbb{N}}$ converge vers la vraie distribution $F^\mathcal{L}$.

- ☞ Q1 Cherchons tout d'abord à engendrer une variable aléatoire **entière** uniforme. Soit $N \in \mathbb{N}^*$ et considérons $\mathcal{L} = \mathcal{U}_{[0, N-1]}$.

Définissons pour $\phi : [0, N-1] \rightarrow [0, N-1]$ et $s \in [0, N-1]$

(a)
$$f_{\phi,s} : \begin{pmatrix} [0, N-1] & \rightarrow & [0, N-1] \\ t & \mapsto & \phi^t(s) \end{pmatrix}$$

Montrer que $f_{\phi,s}$ est périodique à partir d'un certain rang ; on note $T_{\phi,s}$ sa période. En quel sens peut-on dire que $f_{\phi,s}$ "génère" une variable uniformément distribuée sur $[0, N-1]$?

Comme $[0, N-1] \subsetneq \mathbb{N}$ $f_{\phi,s}$ n'est pas injective, soient $i < j$ les deux premiers entiers tels que $f_{\phi,s}(i) = f_{\phi,s}(j)$, et posons $T = j - i \geq 1$. Alors $f_{\phi,s}(i + T) = f_{\phi,s}(i)$, donc $f_{\phi,s}(i + 1 + T) = f_{\phi,s}(i + 1)$ et par récurrence $\forall k \geq i, f_{\phi,s}(k + T) = f_{\phi,s}(k)$ et donc $f_{\phi,s}$ est périodique au-delà de i .

Remarquons que la suite $(f_{\phi,s}(k))_{k \in \mathbb{N}}$ est *déterministe* et pas aléatoire ; il s'agit de déterminer si elle peut être la suite des *réalisations* d'une variable aléatoire uniformément distribuée sur $[0, N-1]$.

Définissons $g(k) = \lim_{N \rightarrow +\infty} \frac{1}{N} |\{i \in \mathbb{N}/y(i) = k\}|$ la fréquence empirique de $k \in [0, N-1]$; alors $f_{\phi,s}$ "génère" une variable distribuée selon la loi \mathcal{L} sur $[0, N-1]$ si la fréquence empirique de tout k est égale à sa probabilité selon \mathcal{L} .

En particulier, $f_{\phi,s}$ est uniformément distribuée ssi $\forall k \in [0, N-1] f_{\phi,s}(k) = \frac{1}{N-1}$.

- (b) En quel sens peut-on dire que $f_{\phi,s}$ est d'autant meilleure que $T_{\phi,s}$ est grande ?

On a $T \leq N$: en effet pour tout $a \in \mathbb{N}, t \mapsto f_{\phi,s}(a+t)$ ne peut pas être injective de $[0, N-1]$, donc il existe $i < j \in [0, N-1]$ tel que $f_{\phi,s}(a+j) = f_{\phi,s}(a+i)$, et par suite $T_{\phi,s} \leq N$.

Or si $T_{\phi,s} < N$, alors $f_{\phi,s}([i, +\infty[) = \{f_{\phi,s}(i), f_{\phi,s}(i+1), \dots, f_{\phi,s}(i+T_{\phi,s}-1)\}$, donc $|f_{\phi,s}([i, +\infty[)| \leq T_{\phi,s} < N$, et donc $f_{\phi,s}([i, +\infty[) \subsetneq [0, N-1]$. Ainsi l'un au moins entiers de $[0, N-1]$ n'est plus jamais atteint au-delà de i , et sa fréquence empirique donc nulle : $f_{\phi,s}$ ne "génère" pas une loi uniforme sur $[0, N-1]$.

Remarquons réciproquement que si $f_{\phi,s}$ est périodique de période exactement N , alors $f_{\phi,s}$ est uniforme.

Ainsi, un "bon générateur" de nombres pseudo-aléatoires sur $[0, N-1]$ est périodique de période N .

- (c) Soient $a \in [2, N-1], c \in [0, N-1]$ et $p \in [2, N-1]$ et définissons $\psi_{a,c,p,s} : \begin{pmatrix} \mathbb{N} & \rightarrow & [0, N-1] \\ n & \mapsto & (an+c) \bmod p \end{pmatrix}$. A quelles conditions $f_{\psi_{a,c,p,s}}$ est-elle un bon générateur uniforme sur $[0, N-1]$?

Par définition de la division euclidienne, $f_{\psi_{a,c,p,s}}$ est à valeur dans $[0, p-1]$ donc pour qu'elle soit un bon générateur sur $[0, N-1]$ il est nécessaire que $p = N$.

Par ailleurs dans $\mathbb{Z}/p\mathbb{Z}$ la suite $(a^n)_{n \in \mathbb{N}}$ est périodique et sa période divise p . Donc si $c \neq 0$ et si p est premier, $f_{\psi_{a,c,p,s}}$ est de période 1 ou p , donc p (car $a \geq 2$), et est donc un bon générateur uniforme sur $[0, N-1]$ (en fait il suffit que p soit premier avec a , et il n'est pas nécessaire que $c = 0$).

Ces générateurs ont été introduits par Lehmer et sont encore souvent utilisés². On sait notamment que $2^{31} - 1 = 2147483647$ est premier (vérifié par Euler en 1750), ainsi que $2^{61} - 1$ (Pervouchine 1883) et $2^{127} - 1$ (Lucas, 1876).³

- ☞ Q2 Cherchons à engendrer une variable aléatoire réelle uniforme sur $[0, 1]$.

²cf Knuth, D.E., 1981 ; The Art of Computer Programming, Volume 2 Seminumerical Algorithms, Addison-Wesley Reading Mass., 688 pages, ISBN 0-201-03822-6

³Parmi les nombres sous la forme $2^n - 1$ seuls ceux où $n \in \{2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127\}$ sont premiers non pas seulement pour $n \in \{2, 3, 5, 7, 13, 17, 19, 31, 67, 127, 257\}$ comme prétendu incorrectement par Marin Mersenne dans la préface de Cogitata Physica-Mathematica (1644). Parmi les $2^{2^n-1} - 1$ courants en informatique, ceux pour $n \in \{2, 3, 5, 7\}$ sont premiers.

Soit $(x_n)_{n \in \mathbb{N}} \in [0, 1]^{\mathbb{N}}$, et définissons pour $f : [0, 1] \rightarrow \mathbb{R}$ continue et $n \in \mathbb{N}^*$

$$S_n^x(f) = \frac{1}{n} \sum_{k=1}^n f(x_k)$$

(a) On dit de x qu'elle vérifie le critère de Weyl si pour toute f continue

$$S_n^x(f) \xrightarrow{n \rightarrow +\infty} \int_0^1 f$$

Comment s'interprète ce critère ?

Il s'agit d'une sorte de critère ergodique : la suite x est suffisamment "équi-répartie" et "équilibrée" pour que sommer aux points chargés par x ou sommer en tout point de $[0, 1]$ soit équivalent.

Soit $r \in \mathbb{R}$, et définissons $x = (\text{frac}(rk))_{k \in \mathbb{N}}$ où $\text{frac}(u) = u - E(u) \in [0, 1[$ désigne la partie fractionnaire de $u \in \mathbb{R}$.

(b) Montrer que x vérifie le critère de Weyl ssi r est irrationnel.

Supposons que r est rationnel.

– Supposons tout d'abord que $f_p = x \mapsto e^{2\pi i p x}$, $p \in \mathbb{Z}$. Par définition $rk = \text{frac}(rk) + E(rk)$ et donc

$$\begin{aligned} S_n^x(f_p) &= \frac{1}{n} \sum_{k=1}^n e^{2\pi i p x_k} \\ &= \frac{1}{n} \sum_{k=1}^n e^{(2p\pi i) rk} \text{ car } e^{(2p\pi i)E(rk)} = 1 \\ &= \frac{1}{n} \sum_{k=1}^n (e^{(2p\pi i)r})^k \end{aligned}$$

Comme r est irrationnel, si $p \neq 0$, alors $e^{(2p\pi i)r} \neq 1$ et donc

$$\begin{aligned} S_n^x(f_p) &= \frac{1}{n} \frac{e^{(2p\pi i)r} - (e^{(2p\pi i)r})^{n+1}}{1 - e^{(2p\pi i)r}} \\ &= \frac{1}{n} \frac{e^{(2p\pi i)r \frac{n+1}{2}} \sin(\pi n r)}{\sin(\pi r)} \end{aligned}$$

et donc

$$|S_n^x(f_p)| \leq \frac{1}{n} \frac{1}{|\sin(\pi r)|} \xrightarrow{n \rightarrow \infty} 0$$

soit

$$S_n^x(f_p) \xrightarrow{n \rightarrow \infty} \int_{[0,1]} e^{2\pi i p u} du$$

Lorsque $p = 0$, on a bien-sûr $S_n^x(f_0) = 1 = \int_{[0,1]} f_0$.

- Le résultat reste donc vrai par linéarité de l'intégrale pour tout polynôme trigonométrique.
- Soit enfin f continue sur $[0, 1]$; d'après le théorème de Dirichlet f est limite uniforme d'une suite $(P_l)_{l \in \mathbb{N}}$ de polynômes trigonométriques. Donc

$$\begin{aligned} S_n^x(f) &= \frac{1}{n} \sum_{k=1}^n f(x_k) \\ &= \frac{1}{n} \sum_{k=1}^n \lim_{l \rightarrow \infty} P_l(x_k) \\ &= \lim_{k \rightarrow \infty} S_{nb}^x(P_k) \text{ car la somme } \sum_{k=1}^n \text{ est finie} \\ &= \lim_{k \rightarrow \infty} \int_{[0,1]} P_l \text{ d'après le résultat précédent} \\ &= \int_{[0,1]} \lim_{k \rightarrow \infty} P_l \text{ car la convergence de } (P_l)_{l \in \mathbb{N}} \text{ est uniforme} \\ &= \int_{[0,1]} f \end{aligned}$$

ce qui achève la preuve.

Dans le cas en revanche où $r = \frac{a}{b}$ est rationnel, alors $(rk)_{k \in \mathbb{N}} = \{0, \frac{1}{b}, \dots, \frac{b-1}{b}\}$: la suite x ne prend que b valeurs distinctes et le résultat est faux, par exemple lorsque f est fonction affine par morceaux telle que

$$\begin{cases} f\left(\frac{k}{b}\right) = 1 \\ f\left(\frac{k+\frac{1}{2}}{b}\right) = 2 \end{cases}$$

car alors $S_n^x(f) = 1 < \frac{3}{2} = \int_{[0,1]} f$.

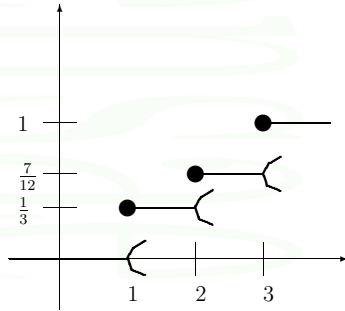
☞ Q3 On cherche désormais à engendrer des réalisations d'une variable aléatoire de loi quelconque

Soit \mathcal{L} la loi de densité $f_{\mathcal{L}} = \frac{1}{3}\delta_1 + \frac{1}{4}\delta_2 + \frac{5}{12}\delta_3$.
Tracer la fonction de répartition $F_{\mathcal{L}}$ de \mathcal{L} .
En déduire son inverse généralisée

(a)
$$F^{-1} : \begin{pmatrix} [0, 1] & \rightarrow & \mathbb{R} \\ s & \mapsto & \inf\{x \in \mathbb{R} / F_{\mathcal{L}}(x) \geq s\} \end{pmatrix}$$

Soit $U \sim \mathcal{U}_{[0,1]}$ et définissons $X = F_{\mathcal{L}}^{-1}(U)$; quelle est la loi de X ?

Le graphe de $F_{\mathcal{L}}$ est le suivant :



On a donc

$$F_{\mathcal{L}}^{-1}(t) = \begin{cases} -\infty & \text{si } 0 = t \\ 1 & \text{si } 0 < t < \frac{1}{3} \\ 2 & \text{si } \frac{1}{3} \leq t < \frac{7}{12} \\ 3 & \text{si } \frac{7}{12} \leq t \leq 1 \end{cases}$$

Par conséquent lorsque $U \sim \mathcal{U}_{[0,1]}$, $X = F_{\mathcal{L}}^{-1}(U)$ ne charge positivement que $\{1, 2, 3\}$ et on a

$$\begin{aligned} \mathbb{P}(X = 1) &= \mathbb{P}\left(0 \leq U < \frac{1}{3}\right) = \frac{1}{3} \\ \mathbb{P}(X = 2) &= \mathbb{P}\left(\frac{1}{3} \leq U < \frac{7}{12}\right) = \frac{1}{4} \\ \mathbb{P}(X = 3) &= \mathbb{P}\left(\frac{7}{12} \leq U < 1\right) = \frac{5}{12} \end{aligned}$$

de sorte que $X \sim \mathcal{L}$.

- (b) Soit alors $F_{\mathcal{L}}$ la fonction de répartition supposée continue et strictement croissante d'une loi \mathcal{L} quelconque.
On définit de même $X = F_{\mathcal{L}}^{-1}(U)$ pour $U \sim \mathcal{U}_{[0,1]}$; quelle est la loi de X ?

On a pour tout $t \in [0, 1]$

$$\begin{aligned} \mathbb{P}(X \leq t) &= \mathbb{P}(F_{\mathcal{L}}^{-1}(U) \leq t) \\ &= \mathbb{P}(U \leq F_{\mathcal{L}}(t)) \quad \text{car } F_{\mathcal{L}} \text{ est croissante} \\ &= F_{\mathcal{L}}(t) \quad \text{car } U \sim \mathcal{U}_{[0,1]} \end{aligned}$$

Par conséquent $X \sim \mathcal{L}$.

Remarquons que $F_{\mathcal{L}}$ étant strictement croissante, c'est sa continuité qui nous assure qu'elle est bijective.

- (c) Proposer un algorithme qui génère des réalisations successives d'une variable X de loi exponentielle $\mathcal{E}(\lambda)$ de densité $f_{\mathcal{E}(\lambda)}(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, +\infty[}(x)$ pour $\lambda > 0$.

$$F_{\mathcal{E}(\lambda)}(u) = \int_0^u \lambda e^{-\lambda x} dx = (1 - e^{-\lambda u})$$

Soit donc $U \sim \mathcal{U}_{[0,1]}$; alors $-\frac{1}{\lambda} \ln(1 - U) \sim \mathcal{E}(\lambda)$; or par ailleurs $(1 - U) \sim \mathcal{U}_{[0,1]}$, donc $-\frac{1}{\lambda} \ln(U) \sim \mathcal{E}(\lambda)$.

On génère donc u_n réalisation de $U \sim \mathcal{U}_{[0,1]}$ et on renvoie $x_n = -\frac{1}{\lambda} \ln(u_n)$.

- (d) Soit $(X, Y) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 I_2\right)$.
Soient $R^2 = X^2 + Y^2$ et $\theta = \arctan \frac{Y}{X}$.
Montrer que R^2 et θ sont indépendantes, et donner leurs lois.
En déduire un algorithme de génération d'une variable gaussienne.

Soit $G : \mathbb{R}^2 \rightarrow \mathbb{R}^n$ absolument continue; alors

$$\begin{aligned} \mathbb{E}(G(R^2, \theta)) &= \int_{\mathbb{R}^2} G\left(x^2 + y^2, \arctan \frac{y}{x}\right) \frac{1}{\sqrt{2\pi\sigma^2}^2} e^{-\frac{x^2+y^2}{2\sigma^2}} dy dx \\ &= \int_{\mathbb{R}^+ \times [0, 2\pi]} G(\rho^2, \theta) \frac{1}{2\pi\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}} \rho d\rho d\theta \end{aligned}$$

par le changement de variables en coordonnées polaires⁴. En particulier si G est un produit $G(a, b) = G_1(a)G_2(b)$ alors

$$\mathbb{E}(G_1(R^2)G_2(\theta)) = \left(\frac{1}{\sigma^2} \int_{\mathbb{R}^+} \underbrace{G_1(\rho^2) e^{-\frac{\rho^2}{2\sigma^2}} \rho}_{\text{ne dépend que de } \rho} d\rho\right) \left(\frac{1}{2\pi} \int_{[0, 2\pi]} \underbrace{G_2(\theta)}_{\text{que de } \theta} d\theta\right)$$

Par conséquent, les variables R^2 et θ sont indépendantes.

En outre on a pour toute G_1 absolument continue

$$\begin{aligned} \mathbb{E}(G_1(R^2)) &= \int_{\mathbb{R}^+} \frac{1}{\sigma^2} G_1(\rho^2) e^{-\frac{\rho^2}{2\sigma^2}} \rho d\rho \\ &= \int_{\mathbb{R}^+} \frac{1}{2\sigma^2} G_1(u) e^{-\frac{1}{2\sigma^2}u} du \end{aligned}$$

⁴Soit $f : \mathbb{R}^2 \rightarrow E$ mesurable absolument continue, et soit à calculer $\int_{\mathcal{A} \subset \mathbb{R}^2} f(x, y) dx dy$.
Posons

$$\phi : \begin{pmatrix} \mathbb{R}^* \times \mathbb{R} \\ \begin{pmatrix} x \\ y \end{pmatrix} \end{pmatrix} \mapsto \begin{pmatrix} \mathbb{R}^+ \times \mathbb{R} \\ \begin{cases} x^2 + y^2 \\ \arctan(\frac{y}{x}) \text{ si } x > 0 \\ \arctan(\frac{y}{x}) + \pi \text{ si } x < 0 \end{cases} \end{pmatrix}$$

ϕ est différentiable et de jacobienne

$$Jac(\phi)_{(x,y)} = \begin{pmatrix} 2x & 2y \\ -\frac{y}{x^2} \frac{1}{1+(\frac{y}{x})^2} & \frac{1}{x} \frac{1}{1+(\frac{y}{x})^2} \end{pmatrix}$$

donc de jacobien

$$|Jac(\phi)_{(x,y)}| = 2x \frac{1}{x} \frac{1}{1+(\frac{y}{x})^2} - 2y \left(-\frac{y}{x^2} \frac{1}{1+(\frac{y}{x})^2}\right) = 2$$

de sorte que

$$\int_{\mathcal{A} \subset \mathbb{R}^2} f(x, y) dx dy = \int_{\phi(\mathcal{A})} f(\phi^{-1}(\rho^2, \theta)) 2 \left(\frac{1}{2} \rho d\rho\right) d\theta$$

et donc $R^2 \sim \mathcal{E}(\frac{1}{2\sigma^2})$.

Par ailleurs pour toute G_2 absolument continue

$$\mathbb{E}(G_2(\theta)) = \int_{[0, 2\pi[} G_2(\theta) \frac{1}{2\pi} d\theta$$

et donc $\theta \sim \mathcal{U}_{[0, 2\pi]}$

Pour générer une réalisation d'une variable qui suivrait la loi $\mathcal{N}(m, \sigma^2)$, on génère donc $\frac{\theta_n}{2\pi}$ uniformément sur $[0, 1]$ et r_n^2 selon la loi $\mathcal{E}(\frac{1}{\sigma^2})$ et on renvoie le nombre

$$x_n = m + \sqrt{r_n^2} \cos(2\pi(\frac{\theta_n}{2\pi}))$$

(e) Même question si X suit une loi de Weibull $\mathcal{W}(\alpha, \beta)$ de densité $f_{\mathcal{W}(\alpha, \beta)}(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta} \mathbf{1}_{x>0}$.

$$F_{\mathcal{W}(\alpha, \beta)}(y) = \int_0^y f_{\mathcal{W}(\alpha, \beta)}(x) dx = 1 - e^{-\alpha y^\beta}$$

et donc on génère u_n réalisation de $U \sim \mathcal{U}_{[0, 1]}$ et on renvoie

$$x_n = (-\frac{1}{\alpha} \ln(u_n))^{\frac{1}{\beta}}$$

(f) Même question si X suit une loi de Poisson $\mathcal{P}(\lambda)$ de densité $f_{\mathcal{P}(\lambda)}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$. On pourra comparer $x_n = u_1 \times \dots \times u_n$, où chaque u_n est une réalisation de $U \sim \mathcal{U}_{[0, 1]}$, à $e^{-\lambda}$.

La densité $f_{\mathcal{P}(\lambda)}$ s'écrit sur \mathbb{R} comme une somme infinie de masses de Diracq $f_{\mathcal{P}(\lambda)}(s) = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} \delta_k(s)$ donc un algorithme consiste à générer u uniformément sur $[0, 1]$ et à le placer dans l'un des intervalles $[S_n, S_{n+1}[$ où $S_n = \sum_{k=0}^n \frac{\lambda^k}{k!} e^{-\lambda}$.

Le calcul des $(S_n)_{n \in \mathbb{N}}$ peut néanmoins être dissuasif en pratique; il est possible en fait de s'en passer.

Soient en effet $(Y_n)_{n \in \mathbb{N}} \stackrel{iid}{\sim} \mathcal{E}(\mu)$; alors $\forall n \in \mathbb{N}^*, Y_1 + \dots + Y_n \sim \Gamma(n, \mu)$ (voir TD 1, exercice 2), et donc

$$\mathbb{P}(Y_1 + \dots + Y_n > 1) = \int_{\frac{1}{\mu}}^{+\infty} \frac{t^{n-1}}{(n-1)!} e^{-t} dt$$

Définissons alors $\nu(Y) = \min\{n / Y_1 + \dots + Y_n > 1\} - 1$; alors

$$\begin{aligned} \mathbb{P}(\nu(Y) = n) &= \mathbb{P}(Y_1 + \dots + Y_{n+1} > 1) - \mathbb{P}(Y_1 + \dots + Y_n > 1) \\ &= \int_{\frac{1}{\mu}}^{+\infty} \frac{t^n}{n!} e^{-t} dt - \int_{\frac{1}{\mu}}^{+\infty} \frac{t^{n-1}}{(n-1)!} e^{-t} dt \\ &= \frac{1}{\mu^n n!} e^{-\frac{1}{\mu}} \end{aligned}$$

et donc $\nu(Y) \sim \mathcal{P}(\frac{1}{\mu})$.

Ainsi, engendrer une loi de Poisson de paramètre λ revient à engendrer une succession de lois exponentielles d'espérance $\frac{1}{\lambda}$. Or $\mathcal{E}(\frac{1}{\lambda}) = \mathcal{W}(\frac{1}{\lambda}, 1)$ donc d'après la question précédente si $U \sim \mathcal{U}_{[0, 1]}$, alors $-\frac{1}{\lambda} \ln U \sim \mathcal{E}(\frac{1}{\lambda})$. Par conséquent posons $U_i = 1 - e^{-\frac{1}{\lambda} Y_i}$; U_i uniforme, et en outre

$$\begin{aligned} Y_1 + \dots + Y_n > 1 &\Leftrightarrow -\frac{1}{\lambda} \sum_{k=1}^n \ln U_k > 1 \\ &\Leftrightarrow U_1 \dots U_n > e^{-\lambda} \\ &\Leftrightarrow X_n > e^{-\lambda} \end{aligned}$$

où $X_n = U_1 \dots U_n$

Définissons donc finalement $N(X) = \min\{n / X_n < e^{-\lambda}\} - 1$; alors $N(X) = \nu(X)$, donc $N(X) \sim \mathcal{P}(\lambda)$.

Pour générer une loi $\mathcal{P}(\lambda)$, il suffit donc de générer une suite u_1, \dots, u_n, \dots uniformes $[0, 1]$, et de renvoyer le premier entier $n - 1$ tel que $u_1 \dots u_n > e^{-\lambda}$.

Un tel algorithme s'écrit : ⁵

- 1 $n \leftarrow 0$ et $x \leftarrow 1$
- 2 Tant que $x > e^{-\lambda}$
- 3 Générer u_n uniformément sur $[0, 1]$
- 4 $x \leftarrow x \times u_n$ et $n \leftarrow n + 1$
- 5 Renvoyer n

⁵Pour un exposé plus complet, voir D. Knuth, *The Art of Computer Programming*, vol. 2, p 132.