Contenu

1 Variable aléatoire

Considérons une boîte contenant trois boules rouges et deux boules blanches. On tire, sans remise, trois boules l'une après l'autre. Une question qu'on se pose, a priori, est : combien aura-t-on tiré de boules rouges à la fin de cette expérience ? Si on désigne par X ce nombre, il est clair que X est aléatoire (on ne sait pas avec certitude quelle valeur il va prendre). Cependant, on sait quelles sont les valeurs qu'il est susceptible de prendre : $1, 2$ ou 3. À chacun de ces nombres, on associe une probabilité puisque chacun de ces nombres correspond à un évènement lié à notre expérience :

$$"X = 1" \Leftrightarrow "\textit{On tire une boule rouge}"$$

etc. Pour préciser notre pensée, introduisons un ensemble fondamental lié à cette expérience ($R =$"Boule rouge" et $B =$"boule blanche") :

$$\Omega = \{RBB, BRB, BBR, RRB, RBR, BRR, RRR\}$$

Nons avons donc les correspondances

X	1	2	3	Total
Évènement	$\{RBB, BRB, BBR\}$	$\{RRB, RBR, BRR\}$	$\{RRR\}$	Ω
Probabilité	$\frac{3}{5}\frac{2}{4}\frac{1}{3}+\frac{2}{5}\frac{3}{4}\frac{1}{3}+\frac{2}{5}\frac{1}{4}\frac{3}{3}=\frac{3}{10}$	$\frac{3}{5}\frac{2}{4}\frac{2}{3}+\frac{3}{5}\frac{2}{4}\frac{2}{3}+\frac{2}{5}\frac{3}{4}\frac{2}{3}=\frac{6}{10}$	$\frac{3}{5}\frac{2}{4}\frac{1}{3}=\frac{1}{10}$	1

Nous voyons ainsi que X peut être interprétée comme une fonction réelle ayant pour domaine l'ensemble fondamental Ω :

$$\begin{aligned} X &: \Omega \to \mathbb{R} \\ \omega &\to X(\omega) = x \end{aligned}$$

ω	RBB	BRB	BBR	RRB	RBR	BRR	RRR
X	1	1	1	2	2	2	3

Définition 1 *Étant donné un espace probabilisé* $(\Omega, \mathcal{E}, P)$*, une variable aléatoire est une fonction*

$$X : \Omega \to \mathbb{R}$$

À chaque valeur x prise par X, on associe une probabilité

$$p(x) = P\{\omega \in \Omega : X(\omega) = x\} = P(X = x)$$

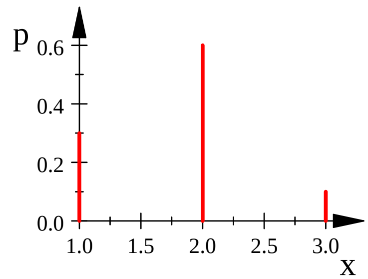
qui est celle de l'évènement "$X = x$" (nécessairement sous-ensemble de Ω). Cette probabilité $p(x)$ vérifie les deux propriétés suivantes :

1. $p(x) \geq 0$ (car c'est une probabilité).
2. $\sum_{x \in X(\Omega)} p(x) = 1$, puisqu'on exhauste tous les évènements inclus dans Ω (la fonction p crée une partition de Ω).

Noter que $X(\Omega)$ est l'ensemble des valeurs prises par la variable X (image de Ω dans $\mathbb{R}$ par la fonction X). Dans l'exemple ci-dessus, $X(\Omega) = \{1, 2, 3\}$.

Nous appellerons $p(x)$ la <u>*fonction de masse*</u> associée à la variable aléatoire X. La donnée d'une variable aléatoire X et de sa fonction de masse déterminent complètement la variable et permet de laisser l'ensemble Ω (et l'expérience aléatoire sous-jacente) en arrière plan. Dans notre exemple, nous résumons ceci dans le tableau de distribution suivant :

X	1	2	3	Total
$p(x)$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	1



<u>Code R</u>

```
valeurs <- c(1, 2, 3)
prob <- c(3/10, 6/10, 1/10)
plot(valeurs, prob, type = 'h', xlab = 'x', ylab = 'p(x)', ylim = c(0,1),
     main = 'Fonction de masse', lwd=5, col = 'red')
```

2 Fonction de répartition

La fonction de répartition associée à la variable aléatoire X est définie par

$$\begin{aligned} F &: \mathbb{R} \to [0; 1] \\ x &\to F(x) = P(X \leq x) \end{aligned}$$

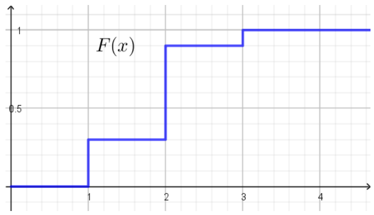
2.1 Propriétés

1. Continue à droite : $\lim_{x \to a^+} F(x) = F(a)$
2. $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ et $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$.
3. Croissante (au sens large) : $a \leq b \Rightarrow F(a) \leq F(b)$, pour tous $a, b \in \mathbb{R}$.
4. $P(a < X \leq b) = F(b) - F(a)$.
5. Pour tout $b \in \mathbb{R}$,
$$P(X = b) = F(b) - F(b^-)$$
où $F(b^-) = \lim_{x \to b^-} F(x)$. La probabilité en $x = b$ n'est pas nulle s'il y a un saut en b.
6. Pour tout $a \in \mathbb{R}$,
$$P(X > a) = 1 - F(a).$$
7. Pour tout $a \in \mathbb{R}$,
$$P(X < a) = F(a^-).$$
Ainsi, si F est continue en a, alors $F(a) = F(a^+) = F(a^-)$.

8. Pour tout $a \in \mathbb{R}$,
$$P(X \geq a) = 1 - F(a^-).$$

Exemple 1 *Reprenons l'exemple ci-dessus. On a*

X	1	2	3
$p(x)$	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$
$F(x)$	$\frac{3}{10}$	$\frac{9}{10}$	$\frac{10}{10}$



Exemple 2 *On considère une boîte contenant N boules numérotées de 1 à N. On en choisit au hasard n $(1 \leq n \leq N)$, sans remise. Soit X le numéro maximum obtenu. Trouver la fonction de répartition de X.*

Réponse *Il est clair que X est concentré sur les nombres $n, n+1, \ldots, N$. Considérons l'évènement "$X \leq k$". Il est équivalent à l'évènement "Tous les numéros sortis sont $\leq k$". Pour réaliser cet évènement, il faut et il suffit de prendre un sous-ensemble de n éléments dans l'ensemble $\{1, 2, \ldots, k\}$, c'est-à-dire une combinaison de n éléments parmi k. Il y a $\binom{k}{n}$ façons de faire cette opération. Comme il y a $\binom{N}{n}$ sous ensembles de taille n parmi N éléments, la réponse est alors*

$$F(k) = P(X \leq k) = \frac{\binom{k}{n}}{\binom{N}{n}}, \; k = n, n+1, \ldots, N.$$

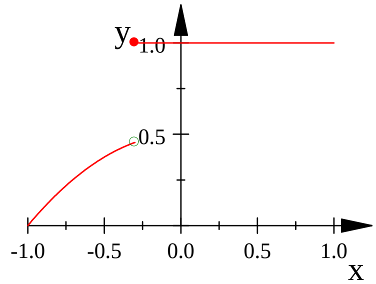
Exemple 3 *On considère la fonction de répartition de la variable aléatoire X :*

$$F(x) = \begin{cases} 0 & si \quad x < -1 \\ \dfrac{1-x^2}{2} & si \quad -1 \leq x < b \\ 1 & si \quad x \geq b. \end{cases}$$

1. *Montrer que $-1 < b \leq 0$.*
2. *Vérifier que $P\left(X = \frac{b-1}{2}\right) + P(X \geq b) = \frac{1}{2} + \frac{b^2}{2}$.*

Réponse

1. *Par hypothèse, $b > -1$. Maintenant, pour $x > -1$, la fonction $\frac{1-x^2}{2}$ (parabole) est croissante et positive tant que $x \leq 0$, puis elle décroît. La valeur maximale permissible pour b est donc 0 puisqu'une fonction de répartition est non décroissante.$\frac{1-x^2}{2}$*



2. *On a* $P\left(X=\frac{b-1}{2}\right)=0$ *car* F *est continue en ce point (milieu de* b *et de* -1*). Il reste*

$$\begin{aligned} P(X \geq b) &= 1 - P(X<b) = 1 - F\left(b^{-}\right) \\ &= 1 - \lim_{x \to b^{-}} \frac{1-x^2}{2} = 1 - \frac{1-b^2}{2} \\ &= \frac{1}{2} + \frac{b^2}{2}. \end{aligned}$$

C'est tout simplement la valeur du saut vertical quand on passe de b^{-} *à* b*.*

3 Variable aléatoire discrète

La variable aléatoire X est dite discrète si l'ensemble image $X(\Omega)=\{x_1, x_2, \ldots\}$ est dénombrable (on supposera toujours les x_n ordonnés par ordre croissant : $x_1 < x_2 < \ldots$). On dit que le support de Ω est l'ensemble $\{x_1, x_2, \ldots\}$, ou que X est concentrée sur $x_1, x_2, \ldots$.

La fonction de répartition d'une variable discrète est une fonction en escalier.

3.1 Fonction de masse

C'est la fonction associée à la variable aléatoire discrète X et définie par

$$\begin{aligned} & p : \mathbb{R} \rightarrow [0;1] \\ & x \rightarrow p(x) = \begin{cases} P(X=x) & \text{si } x \in X(\Omega) \\ 0 \text{ sinon} \end{cases} \end{aligned}$$

3.1.1 Propriétés

1. $p(x) \geq 0$
2. $\sum_n p(x_n) = 1$
3. $p(x_n) = F(x_n) - F(x_{n-1})$
4. $F(x_n) = \sum_{i \leq n} p(x_i)$
5. Probabilités totales. Pour tout évènement E :

$$P(E) = \sum_n P(X \in E) = \sum_n P(E \,|X = x_n)\, p(x_n)$$

Exemple 4 *On lance une pièce de monnaie régulière une fois. Soit* X *le nombre de piles obtenus. Trouver la fonction de masse de* X*.*

Réponse *Il est évident que* X *ne peut prendre que deux valeurs (c'est l'indicatrice de l'évènement "pile") :* 0 *(on obtient face) et* 1 *(on obtient pile). La fonction de masse est immédiatement donnée sous forme de tableau :*

X	0	1	*Total*
$p(x)$	$\frac{1}{2}$	$\frac{1}{2}$	1

Exemple 5 *Reprenons l'exemple du maximum ci-dessus. Nous avons vu que la fonction de répartition est donnée par*

$$F(k) = P(X \leq k) = \frac{\binom{k}{n}}{\binom{N}{n}}, \; k = n, n+1, \ldots, N.$$

On en déduit la fonction de masse

$$p(k) = F(k) - F(k-1) = \frac{\binom{k}{n} - \binom{k-1}{n}}{\binom{N}{n}} = \frac{\binom{k-1}{n-1}}{\binom{N}{n}},\ k = n, n+1, \dots, N.$$

On a utilisé la propriété du triangle de Pascal

$$\binom{k}{n} = \binom{k-1}{n} + \binom{k-1}{n-1}$$

Maintenant, on peut trouver la fonction de masse directement sans passer par la fonction de répartition. En effet, l'évènement "$X = k$" signifie que le plus grand numéro tiré vaut k. Ce numéro étant fixé, les $n-1$ autres numéros sont tous inférieurs à k, donc entre 1 et $k-1$. Il y a $\binom{k-1}{n-1}$ façons de faire ces choix. La réponse finale est donc

$$p(k) = P(X = k) = \frac{\binom{k-1}{n-1}}{\binom{N}{n}},\ k = n, n+1, \dots, N.$$

3.1.2 Exercice

Exercice 1 *Reprendre l'exemple ci-dessus où, cette fois-ci, X est le plus petit des n nombres.*

3.2 Lois discrètes usuelles

3.2.1 Bernoulli $X \sim Bernoulli(p)$

Considérons une expérience aléatoire où on s'intéresse à la réalisation (ou non réalisation) d'un évènement donné A (de probabilité $p = P(A)$). Une telle expérience est dite de Bernoulli. Soit X l'indicatrice de A, c'est-à-dire la variable aléatoire qui prend la valeur 1 si A se réalise et 0 sinon. La distribution (loi) de X est donnée dans la table suivante :

X	0	1	Total
$p(x)$	p	$q = 1-p$	1

On peut résumer ceci comme suit :

$$p(x) = p^x (1-p)^{1-x}, \quad x = 0, 1$$

3.2.2 Loi binomiale $X \sim Bin(n, p)$

On répète une expérience de Bernoulli n fois (n est un nombre fixé à l'avance) de manière indépendante et dans des conditions identiques (jet d'une pièce de monnaie, d'un dé, etc.). Soit X le nombre de succès obtenus au bout des n épreuves. Il est clair que X est concentrée sur les entier $0, 1, \dots, n$. On a donc

$$X = \sum_{i=1}^{n} X_i$$

où les X_i iid $\sim Bernoulli(p)$. La distribution de X est

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \dots, n$$

Exemple 6 *Tirage avec remise. Une urne contient r boules rouges et b boules blanches. On en tire, avec remise, n. Quelle est la probabilité d'obtenir k boules rouges ?*

Réponse *Manifestement, nous avons n épreuves de Bernoulli où "succès" = "tirer une boule rouge" et $p = \frac{r}{r+b}$. La réponse est donc*

$$P(X = k) = \binom{n}{k} \left(\frac{r}{r+b}\right)^k \left(\frac{b}{r+b}\right)^{n-k}, \; k = 0, 1, \cdots, n.$$

Commandes R

– Fonction de masse : dbinom(x, n, p)
– Fonction de répartition : pbinom(x, n, p)
– Graphe : plot(). Exemple : $n = 20, p = .3$.

```
n <- 20 ; p <- 0.3 ; x <- 0 :n
plot(x,dbinom(x, n, p), main = 'Loi binomiale Bin(n=20, p=0.3)', type = 'h', ylab = 'p(x)')
abline (0,0)
```

3.2.3 Loi géométrique $X \sim Geo(p)$

On s'intéresse à la réalisation d'un certain évènement S (succès). On répète une expérience de Bernoulli autant de fois qu'il faut, et ce jusqu'à ce que S se produise. Si, à chaque épreuve, on a $p = P(S)$, et si X désigne le nombre d'épreuves, on dit que X suit une loi géométrique. On écrira $X \sim Geo(p)$.

Il est clair que X est concentrée sur les entiers $1, 2, \ldots$. L'évènement "$X = k$" $(k = 1, 2, \ldots)$ est réalisé si et seulement si on a une séquence de $k-1$ échecs suivis d'un succès. La distribution de X est

$$p(k) = P(X = k) = (1-p)^{k-1} p, \qquad k = 1, 2, \cdots$$

Il est intéressant de noter la propriété suivante :

$$P(X > k) = (1-p)^k, \qquad k = 1, 2, \cdots$$

On peut l'utiliser pour montrer que la loi géométrique n'a pas de mémoire :

$$P(X > n+m \,|\, X > m) = P(X > n), \qquad n, m = 1, 2, \cdots$$

Exemple 7 *On lance une pièce de monnaie régulière jusqu'à obtention de pile. Trouver*

1. *La probabilité que le nombre de lancers dure indéfiniment.*
2. *Qu'il faille un nombre pair de lancers.*
3. *Qu'il faille un nombre premier de lancers.*

Réponse *Il s'agit évidemment d'une variable géométrique $X \rightsquigarrow$ Géo $\left(p = \frac{1}{2}\right)$.*

1. *0 de toute évidence.*
2. *On veut*

$$\begin{aligned}\sum_{k=1}^{\infty} P(X = 2k) &= \sum_{k=1}^{\infty} pq^{2k-1} = pq^{-1} \sum_{k=1}^{\infty} \left(q^2\right)^k \\ &= pq^{-1}\left(\frac{1}{1-q^2} - 1\right) = q^{-1}\left(\frac{1}{1+q} - p\right) \\ &= q^{-1}\left(\frac{1}{2-p} - p\right) = \frac{1-p}{2-p}.\end{aligned}$$

Pour $p = \frac{1}{2}$, on a donc

$$\sum_{k=1}^{\infty} P(X = 2k) = \frac{1-\frac{1}{2}}{2-\frac{1}{2}} = \frac{1}{3},$$

comme on pouvait s'y attendre.

3. *On ne peut pas calculer la valeur exacte de cette probabilité (bien qu'elle soit bien définie). Quelle est la difficulté ?*

3.2.4 Loi binomiale négative $X \sim NB(r,p)$

Nombre d'épreuves jusqu'au $r^{ème}$ "succès" dans une suite d'épreuves de Bernoulli.

$$p(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \qquad x = r, r+1, \ldots$$

Commandes R : dnbinom(x,r,p) et pnbinom(x,r,p)

3.2.5 Loi hypergéométrique (tirage sans remise)

Une boîte contient r boules rouges et b boules blanches ($N = r + n =$ nombre total de boules). On en tire n au hasard sans remise (par exemple d'un seul coup) et soit X le nombre de rouges parmi les n tirées. On dit que X est une variable hypergéométrique. Sa fonction de masse est

$$p(k) = P(X = k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{N}{n}}, \qquad k = 0, 1, \cdots, \min(r,n)$$

Exemple 8 *Capture-recapture (estimation d'un paramètre par* maximum de vraisemblance*). Voir exercices corrigés (et le chapitre 1).*

3.2.6 Loi de Poisson $X \sim Poi(\lambda)$

Modélise le nombre d'arrivées d'évènements "rares".

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \cdots$$

La paramètre $\lambda > 0$ est le taux d'arrivées par unité de temps.

Commandes R : dpois(x,lambda) et ppois(x,lambda)

Approximation de la loi binomiale Soit $X \sim Bin(n,p)$. Si p (ou $1-p$) est "petit", on peut utiliser la loi de Poisson comme approximation de la loi binomiale ($\lambda = np$). Plus précisément, si on fait tendre $n \to +\infty$, $p \to 0$ tout en gardant $np = \lambda$ constant, alors il est facile de montrer que

$$p(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \cdots$$

Exemple 9 *Quelle est la probabilité que, dans un groupe de n personnes, il y ait au moins une qui soit née le 1er janvier ? On ignorera les années bisextiles et on suppose que la probabilité de naître est équiprobable sur les* 365 *jours.*

Réponse. Clairement, $p = \frac{1}{365}$ *= probabilité de naître le 1er janvier. Si* X *est le nombre de personnes nées le 1er janvier, on a (en supposant l'indépendance)*

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{n}{0} p^0 (1-p)^{n-0} = 1 - \left(1 - \frac{1}{365}\right)^n = 1 - \left(\frac{364}{365}\right)^n$$
$$\overset{Poisson}{\approxeq} 1 - \frac{(np)^0}{0!} e^{-np} = 1 - e^{-n/365}$$

Prenons $n = 50$*. Alors la loi binomiale donne*

$$P(X \geq 1) = 1 - \left(\frac{364}{365}\right)^{50} = 1 - pbinom(0, 50, 1/365) = 0.1281817$$

et l'approximation de Poisson donne

$$P(X \geq 1) \approxeq 1 - e^{-50/365} = 1 - ppois(0, 50/365) = 0.128178$$

Processus de Poisson Modèle important. Compte le nombre d'"arrivées" dans un intervalle (aire, volume) $[0;t]$:

$$p(k) = P(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \qquad k = 0, 1, 2, \ldots$$

4 Variable aléatoire continue

4.1 Fonction densité

Soit X une variable aléatoire de fonction de répartition F. On dit que X est une (absolument) continue s'il existe une fonction $f(x) \geq 0$ telle que

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

Plus généralement,

$$P(X \in E) = \int_E f(u)\, du$$

pour tout E tel que $X^{-1}(E) \in \mathcal{E}$. La fonction $f(x)$ s'appelle *fonction de densité* de X. Elle vérifie les propriétés suivantes :

1. $f(x) \geq 0$.
2. $\int_{-\infty}^{+\infty} f(u)\, du = 1$.
3. $P(a \leq X \leq b) = \int_a^b f(u)\, du = F(b) - F(a^-)$.
4. $F'(x) = f(x)$ pp (presque partout).

Les deux premières propriétés définissent complètement une densité.

Exemple 10 *Soit la densité*

$$f(x) = \begin{cases} ax^2 \text{ sur } [-1,1] \\ 0 \text{ ailleurs} \end{cases}$$

Quelle est la valeur de la constante a ?

Réponse *On doit avoir*

$$\int_{-\infty}^{+\infty} f(u)\, du = \int_{-1}^{1} ax^2 dx = \frac{2}{3}a = 1.$$

Donc $a = \frac{3}{2}$. *Noter que*

$$F(x) = \begin{cases} 0 & si & x < -1 \\ \frac{1}{2}x^3 + \frac{1}{2} & si & -1 \leq x \leq 1 \\ 1 & si & x > 1 \end{cases}$$

Exemple 11 *Soit la fonction de densité (de X) suivante*

$$f(x) = \begin{cases} 3/4 \text{ si } x \in [0;1] \\ 1/4 \text{ si } x \in [2;3] \\ 0 \text{ ailleurs} \end{cases}$$

1. *Vérifier que f est bien une fonction de densité.*
2. *Trouver la fonction de répartition correspondante.*
3. *Calculer* $P\left(\frac{1}{2} \leq X \leq \frac{5}{2}\right)$.

Réponse

1. *Clairement, $f \geq 0$. La seconde condition à vérifier est que l'aire totale sous le graphe de f est égale à 1.Ce qui est évident.*

2. Il suffit d'intégrer la fonction $f(x)$. Le calcul (exercice) donne :

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \begin{cases} 0 \ \ si \ x < 0 \\ 3x/4 \ \ si \ x \in [0;1] \\ 3/4 \ \ si \ x \in [1;2] \\ x/4 + 1/4 \ \ si \ x \in [2;3] \\ 1 \ \ si \ x > 3 \end{cases}$$

3. On a (puisque F est continue partout sur $\mathbb{R}$)

$$P\left(\frac{1}{2} \leq X \leq \frac{5}{2}\right) = F\left(\frac{5}{2}\right) - F\left(\frac{1}{2}\right) = \left(\frac{1}{4}\cdot\frac{5}{2} + \frac{1}{4}\right) - \frac{3}{4}\cdot\frac{1}{2} = \frac{1}{2}$$

4.2 Lois continues usuelles

4.2.1 Loi uniforme $X \sim U[a;b]$

Concentrée sur $[a;b]$. A pour densité

$$f(x) = \begin{cases} \dfrac{1}{b-a} \text{ si } x \in [a;b] \\ 0 \text{ ailleurs} \end{cases}$$

La fonction de répartition est donc

$$F(x) = \begin{cases} 0 & si \quad x < a \\ \dfrac{x-a}{b-a} & si \quad x \in [a;b] \\ 1 & si \quad x > b \end{cases}$$

Commandes R : dunif(x,a,b) (densité), punif(x,a, b) (FR), qunif(q,a,b) (quantile d'ordre $q \in]0;1[$).

Exemple 12 *Soit $X \sim U[0;2]$. Trouver : a) $P\left(X > \frac{1}{2}\right)$; b) $P\left(X > 1 \mid X > \frac{1}{2}\right)$.*
Réponse *On a*

$$F(x) = \begin{cases} 0 \ \ si \ x < 0 \\ x/2 \ \ si \ x \in [0;2] \\ 1 \ \ si \ x > 2 \end{cases}$$

On demande
a) $P(X > 1/2) = 1 - F(1/2) = 1 - 1/4 = 3/4$.
b)

$$\begin{aligned} P(X > 1 \mid X > 1/2) &= 1 - P(X \leq 1 \mid X > 1/2) \\ &= 1 - \frac{P(1/2 < X \leq 1)}{P(X > 1/2)} \end{aligned}$$

$$1 - \frac{F(1) - F((X > 1/2))}{P(X > (X > 1/2))} = \frac{2}{3}.$$

4.2.2 Loi exponentielle $X \sim Exp(\lambda = 1/\theta)$

Concentrée sur $[0;\infty[$. A pour densité

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si} \quad x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Propriétés

1. $F(x) = 1 - e^{-\lambda x}$ pour $x > 0$.
2. $P(X > x) = e^{-\lambda x}$.
3. Absence de mémoire : $P(X > s + t \mid X > s) = P(X > t)$.
4. Lien avec le processus de Poisson.

4.2.3 Loi Gamma $X \sim Gamma(\alpha, \lambda)$

Dépend de deux paramètres positifs α et λ. Sa densité est

$$f(x) = \begin{cases} \dfrac{\lambda}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} e^{-\lambda x} \text{ si } x > 0 \\ \\ 0 \text{ ailleurs} \end{cases}$$

Ici,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

est la fonction Gamma (définie pour $\alpha > 0$). Elle vérifie notamment les propriétés suivantes.

1. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.
2. $\Gamma(n + 1) = n!$ si n est un entier positif.
3. $\Gamma(1/2) = \sqrt{\pi}$.

Commande R : dgamma(x,shape=α,scale=β) ($\beta = 1/\lambda$) (densité), pgamma(x,shape=α,scale=β) ($\beta = 1/\lambda$) (FR), qgamma(q,shape=α,scale=β) ($\beta = 1/\lambda$) (quantile d'ordre q).

Cas particuliers

1. Si $\alpha = 1$, on retrouve la loi exponentielle.
2. Si $\lambda = 1/2$ et $\alpha = r/2$, où r est un entier positif, alors nous avons une loi du Khi carré à r degrés de liberté : χ^2_1.
3. Si $\alpha = r$ (entier positif), on a la loi d'Erlang (utilisée dans les *files d'attente*).

4.2.4 Loi normale $X \sim N(\mu, \sigma^2)$

Dépend de deux paramètres μ et σ ($\sigma > 0$). Sa densité est donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \qquad x \in \mathbb{R}$$

Le graphe de f est symétrique par rapport à la droite $x = \mu$ et a deux points d'inflexion en $\mu \pm \sigma$.

La règle des 3 sigmas dit :

$$P(|X - \mu| \leq \sigma) \approxeq 68\%$$
$$P(|X - \mu| \leq 2\sigma) \approxeq 95\%$$
$$P(|X - \mu| \leq 3\sigma) \approxeq 99.7\%$$

En fait, pour 2σ, on a la valeur plus précise (et très fréquente)

$$P(|X - \mu| \leq 1.96\sigma) \approxeq 95\%$$

Commande R : dnorm(x,μ, σ), pnorm(x,μ, σ), qnorm(q,μ, σ).

Variable centrée réduite

$$Z = \frac{X - \mu}{\sigma}$$

Elle suit une loi normale $N(\mu = 0; \sigma = 1)$.

La fonction de répartition de la variable centrée réduite est notée

$$\Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}u^2\right) du$$

Théorème de Moivre-Laplace Si $X \sim Bin(n, p)$ avec n grand, alors on a l'approximation

$$P(a \leq X \leq b) \approxeq \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right)$$

Exemple 13 *Prenons $n = 20$ et $p = \frac{1}{2}$. Alors*

$$\begin{aligned} P(X = 10) &= \binom{20}{10}\left(\frac{1}{2}\right)^{20} = \frac{46\,189}{262\,144} = 0.176\,20 \\ &= dbinom(10, 20, 0.5) = 0.1761971 \end{aligned}$$

Avec l'approximation ci-dessus, avec $np = 10$ et $npq = 5$,

$$\begin{aligned} P(X = 10) &= P(9.5 \leq X \leq 10.5) \approxeq \Phi\left(\frac{10.5 - 10}{\sqrt{5}}\right) - \Phi\left(\frac{9.5 - 10}{\sqrt{5}}\right) \\ &= pnorm(10.5, 10, sqrt(5)) - pnorm(9.5, 10, sqrt(5)) = 0.1769367 \\ &= \Phi(0.223\,6) - \Phi(-0.223\,6) \\ &= 2\Phi(0.223\,6) - 1 \\ &\overset{Table}{=} 2 \times 0.5885 - 1 = 0.177 \end{aligned}$$

4.2.5 Loi béta

Dépend de deux paramètres positifs $\alpha > 0$ et $\beta > 0$. Concentrée sur $]0; 1[$.

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha - 1}(1 - x)^{\beta - 1}, \qquad x \in]0; 1[$$

Si $\alpha = \beta = 1$, on retrouve la loi uniforme. Utile dans l'approche bayésienne.

Commande R : dbeta(x,α, β), pbeta(x,α, β), qbeta(q,α, β).

4.2.6 Loi de Cauchy

Sa densité est

$$f(x) = \frac{1}{\pi}\frac{1}{1 + x^2}, \qquad x \in \mathbb{R}$$

5 Fonction d'une variable aléatoire

La variable Y est une fonction de la variable X : $Y = g(X)$, X concentrée sur un intervalle I.

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

Dans le cas continu, si g est strictement monotone sur I, la fonction de densité de Y est donnée par

$$f_Y(y) = \left|\frac{d}{dy} g^{-1}(y)\right| f_X\left(g^{-1}(y)\right)$$

Généralement, dans le cas continu, il est plus simple de passer par la fonction de répartition.

Exemple 14 *Soit* $X \sim N(0,1)$*. Trouver la loi de* $Y = X^2$*.*

Il est clair que si $y < 0$*,* $F_Y(y) = 0$ *(*Y *est concentrée sur* $[0; +\infty[$*). Soit donc* $y \geq 0$ *donné. On a*

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P\left(X^2 \leq y\right) = P\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) \\ &= \Phi\left(\sqrt{y}\right) - \Phi\left(-\sqrt{y}\right) = 2\Phi\left(\sqrt{y}\right) - 1 \end{aligned}$$

On en tire la densité :

$$\begin{aligned} f_Y(y) &= 2\left(\frac{d}{dy}\sqrt{y}\right) f_X\left(\sqrt{y}\right) = 2\left(\frac{1}{2\sqrt{y}}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-y/2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{y}}e^{-y/2} = \frac{(1/2)^{1/2}\, y^{-1/2}}{\Gamma(1/2)}e^{-y/2} \end{aligned}$$

On reconnaît une loi Gamma $(\alpha = 1/2, \lambda = 1/2) = \chi_1^2$*.*

5.1 Exercices

Exercice 2 *Montrer que si* $X \sim N(0,1)$*, alors* $Y = \sigma X + \mu \sim N\left(\mu, \sigma^2\right)$*. Généraliser : si* $X \sim N\left(\mu, \sigma^2\right)$*, alors* $aX + b \sim N\left(a\mu + b, a^2\sigma^2\right)$ *(*$a \neq 0$*).*

Exercice 3 *Montrer que si* $X \sim U[a;b]$*, alors* $Y = \frac{X-a}{b-a} \sim U[0;1]$*.*

Exercice 4 *Montrer que* $X \sim U[0;1] \Leftrightarrow (1 - X) \sim U[0;1]$

Exercice 5 *Soit* X *continue avec FR* $F(X)$*. Montrer que* $Y = F(X) \sim U[0;1]$*. Ceci est utilisé pour générer des nombres aléatoires de loi de* X*. Exemple* $X \sim Exp(\lambda)$*, i.e.* $F(x) = 1 - e^{-\lambda x}$ $(x \geq 0)$ $\Rightarrow U = \frac{\ln(1-U)}{-\lambda} \sim Exp(\lambda)$*.*

6 Quantiles

Les quantiles (ou percentiles) d'une distribution sont des caractéristiques facilement interprétables.

Définition 2 *Soit* $0 < p < 1$ *donné. Le quantile d'ordre* p *de la distribution d'une variable aléatoire* X *est une valeur* x_p *telle que* $P(X < x_p) \leq p$ *et* $P(X \leq x_p) \geq p$*.*

Remarque 1 *Si* $p = 1/2$*,* $x_{1/2} = q_2$ *est la médiane (ou le second quartile). Le premier et le 3eme quartiles sont respectivement* $x_{1/4} = q_1$ *et* $x_{3/4} = q_3$*. La différence* $q_3 - q_1$ *est le rang interquartile.*

Remarque 2 *Commande R : quantile(x,...)*

7 Quelques rappels de calcul

Il est utile de rappeler quelques résultats de calcul souvent utilisés en statistiques.

Sommes finies de puissances d'entiers

$$\begin{aligned} \sum_{k=1}^{n} k &= \frac{1}{2}n(n+1) \\ \sum_{k=1}^{n} k^2 &= \frac{1}{6}n(n+1)(2n+1) \\ \sum_{k=1}^{n} k^3 &= \frac{1}{4}n^2(n+1)^2 \\ \sum_{k=1}^{n} k^4 &= \frac{1}{30}n(n+1)(2n+1)\left(3n^2+3n-1\right) \end{aligned}$$

Des formules générales existent (utilisent les polynômes de Bernoulli).

Sommes infinies de puissances d'entiers (séries) Posons (fonction dzéta de Riemann)

$$\zeta(p) = \sum_{k=1}^{\infty} \frac{1}{n^p}$$

Alors $\zeta(p)$ converge si et seulement si $p > 1$. On sait en particulier que

$$\begin{aligned}
\zeta(2) &= \frac{1}{6}\pi^2 \\
\zeta(3) &\approxeq 1.202\,1 \\
\zeta(4) &= \frac{1}{90}\pi^4 \\
\zeta(5) &\approxeq 1.036\,9 \\
\zeta(6) &= \frac{1}{945}\pi^6
\end{aligned}$$

Limites et séries

$$\begin{aligned}
\lim_{n\to\infty} n^{1/n} &= 1 \\
\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n &= e^x
\end{aligned}$$

Si $f(x)$ est une fonction n fois différentiable sur un intervalle $]a;b[$ telle que $f^{(n-1)}(x)$ est continue sur $[a;b]$. Soit $c \in [a;b]$. Alors, pour tout $x \in [a;b]$, il existe un nombre ξ entre c et x tel que

$$f(x) = f(c) + \sum_{k=1}^{n-1} \frac{(x-c)^k}{k!} f^{(k)}(c) + \frac{(x-c)^n}{n!} f^{(n)}(\xi)$$

Pour $n = 1$, on retrouve le théorème des accroissements finis (ou théorème de Lagrange).

$$\begin{aligned}
e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} \\
\ln(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \frac{1}{5}x^5 - \cdots \quad (x \in]-1;1])
\end{aligned}$$

et

$$(1-x)^{-a} = 1 + \frac{a}{1!}x + \frac{a(a+1)}{2!}x^2 + \frac{a(a+1)(a+2)}{3!}x^3 + \cdots = \sum_{n=0}^{\infty} \binom{a}{n} x^n$$

pour tout $|x| < 1$, $a \in \mathbb{R}$. Ici, $\binom{a}{n}$ est le coefficient du binôme généralisé :

$$\begin{aligned}
\binom{a}{n} &= \frac{a(a+1)\cdots(a+n-1)}{n!}, \quad n = 1, 2, \cdots \\
\binom{a}{0} &= 1
\end{aligned}$$

Dérivation sous le signe $\int$ Supposons que $\varphi(x,t)$, $a(t)$ et $b(t)$ sont différentiables par rapport à t pour $x \in \mathbb{R}$ et $t \in \mathbb{R}$. Alors

$$\frac{d}{dt} \int_{a(t)}^{b(t)} \varphi(x,t)\, dx = \varphi(b(t), t)\, b'(t) - \varphi(a(t), t)\, a'(t) + \int_{a(t)}^{b(t)} \frac{\partial}{\partial t} \varphi(x,t)\, dx$$

ou $b'(t) = \frac{d}{dt} b(t)$ et $a'(t) = \frac{d}{dt} a(t)$.

Supposons que $\varphi(x,t)$ est différentiable par rapport à t pour $x \in \mathbb{R}$ et $t \in \mathbb{R}$, et s'il existe une fonction $g(x,t)$ telle que

1. $\left|\dfrac{\partial}{\partial t}\varphi(x,t)\right|_{t=t_0} \leq g(x,t) \qquad \text{pour } t \in \left]t_0 - h, t_0 + h\right[\;\; (\text{voisinage de } t_0)$
2. $\displaystyle\int_{-\infty}^{\infty} g(x,t) < 0 \qquad (\text{converge})$

alors

$$\frac{d}{dt}\int_{-\infty}^{\infty} \varphi(x,t)\,dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial t}\varphi(x,t)\,dx$$

Approximation de Stirling de $n!$

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

dans le sens où

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1$$