Probabilités

Moh. Sedki

Université Paris-Sud

19 septembre 2017

Notions d'événement en probabilité

Evénement certain Ω toujours réalisé, impossible $\emptyset$ jamais réalisé

Composition d'événements

- Négation d'un événement $\bar{A} = \Omega \backslash A$ réalisé ssi A n'est pas réalisé
- Union d'événements $A \cup B$: A est réalisé ou B est réalisé
- Intersection d'événements $A \cap B$: A est réalisé et B est réalisé

Relation entre événements

- Inclusion $A \subset B$: si A est réalisé, B l'est aussi
- Egalité $A = B$
- Incompatibilité $A \cap B = \emptyset$ (on dit aussi événements disjoints)

A et $\bar{A}$ sont incompatibles car $A \cap \bar{A} = \emptyset$; de plus, $A \cup \bar{A} = \Omega$

Algèbre des événements

Idempotence	$A \cup A = A, \quad A \cap A = A$
Commutativité	$A \cup B = B \cup A, \quad A \cap B = B \cap A$
Associativité	$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$ $(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$
Distributivité	$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
Identités	$A \cup \emptyset = A, \quad A \cup \Omega = \Omega$ $A \cap \emptyset = \emptyset, \quad A \cap \Omega = A$
Complémentarité	$\overline{(\bar{A})} = A, \quad \bar{\emptyset} = \Omega, \quad \overline{\Omega} = \emptyset$
Lois de De Morgan	$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B}$

Système d'événements

Soient n événements $A_1, A_2, \ldots, A_n$; $\left(\bigcap_{i=1}^{n} A_i\right) \subset \left(\bigcup_{i=1}^{n} A_i\right)$

- Intersection $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n$
  tous les événements sont réalisés en même temps
- Union $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n$
  au moins l'un des événements est réalisé

L'ensemble $\{A_i\}_{i=1,2,\ldots,n}$ d'événements est une partition (dit aussi système complet d'événements ou événements exhaustifs) <u>ssi</u>

- $\forall i,\ A_i \neq \emptyset$ aucun événement impossible
- $\forall i \neq j,\ A_i \cap A_j = \emptyset$ événements incompatibles 2 à 2
- $\bigcup_{i=1}^{n} A_i = \Omega$

Calculs de probabilités

Probabilité d'un événement A : $\mathbb{P}(A) = \dfrac{\#A}{\#\Omega}, \quad 0 \leq \mathbb{P}(A) \leq 1$

$\mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\emptyset) = 0$

Formule de Poincaré : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Si A et B sont 2 événements incompatibles ($A \cap B = \emptyset$), alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$; en particulier, $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$

De façon générale, $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$, l'égalité n'étant vérifiée que pour des événements incompatibles

Formule des probabilités totales : Si $B_1, \ldots, B_n$ forment une partition de Ω, alors $\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \ldots + \mathbb{P}(A \cap B_n)$
En particulier, $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \bar{B})$

Probabilité conditionnelle

Définition : $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ avec $\mathbb{P}(B) \neq 0$
probabilité de réaliser A sachant que B a été réalisé

Propriétés

- $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B)$ pour $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$
- $\mathbb{P}(\bar{A}|B) = 1 - \mathbb{P}(A|B)$

Théorème des probabilités totales : Si $\{B_1, \dots, B_n\}$ est une partition de Ω, alors $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$

Théorème de Bayes : $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A})}$

Indépendance et indépendance conditionnelle

$A \perp B$: A et B sont indépendants
<u>ssi</u> $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
<u>ssi</u> $\mathbb{P}(A|B) = \mathbb{P}(A)$ pour $\mathbb{P}(B) > 0$
<u>ssi</u> $\mathbb{P}(B|A) = \mathbb{P}(B)$ pour $\mathbb{P}(A) > 0$
(la connaissance de A n'altère pas celle de B, et réciproquement)

Ne pas confondre avec $\mathbb{P}(A \cap B) = 0 \Leftrightarrow A$ et B disjoints

$(A \perp B)|C$: A et B sont indépendants conditionnellement à C
<u>ssi</u> $\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C)\mathbb{P}(B|C)$
<u>ssi</u> $\mathbb{P}(A|B, C) = \mathbb{P}(A|B \cap C) = \mathbb{P}(A|C)$ pour $\mathbb{P}(B) > 0$
<u>ssi</u> $\mathbb{P}(B|A, C) = \mathbb{P}(B|A \cap C) = \mathbb{P}(B|C)$ pour $\mathbb{P}(A) > 0$
(lorsque C est connu, la connaissance de A n'altère pas celle de B)

Variable aléatoire

Une variable aléatoire (v.a.) X est définie sur Ω, en associant à chaque résultat de l'expérience aléatoire un nombre réel x (la réalisation de X)
L'événement $X = x$ prend en compte tous les résultats pour lesquels la variable X prend la valeur x

Une v.a. X est dite discrète si l'ensemble des réalisations possibles $x_1, x_2, \ldots, x_n$ pour cette variable est fini ou dénombrable

Une v.a. X est dite continue si son domaine de variation est l'ensemble $\mathbb{R}$ ou un intervalle de $\mathbb{R}$

Loi de probabilité

Cas discret : A chacune des réalisations x_i de la v.a. X est associée une probabilité $\mathbb{P}(X = x_i) = p_i$ avec $0 \leq p_i \leq 1, \forall i$

L'ensemble des couples (x_i, p_i) forme une loi de probabilité si la somme de toutes ses probabilités est égale à 1, $\sum_{i=1}^{n} p_i = 1$

Cas continu : La loi de probabilité est décrite par la fonction de densité $f(x) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbb{P}[x \leq X < x + \Delta]$

- $f(x) \geq 0 \quad \forall x$
- Surface totale sous la courbe

$$\int_{-\infty}^{+\infty} f(x) \mathrm{d}x = 1$$

Fonction de répartition

Définition générale : $F(x) = F_X(x) = \mathbb{P}(X \leq x)$

–**Cas discret** : $F(x) = \sum_{x_i \leq x} \mathbb{P}(X = x_i)$

–**Cas continu** : $F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$

Propriétés générales :

- Fonction monotone croissante entre 0 et 1, $0 \leq F(x) \leq 1$
- Si $a \leq b$, alors $F(a) \leq F(b)$
- $\mathbb{P}(a < X \leq b) = F(b) - F(a)$
- $\lim_{x \to -\infty} F(x) = 0$ et $\lim_{x \to +\infty} F(x) = 1$

Espérance

–**Cas discret** : $\mathbb{E}(X) = \sum\limits_{i=1}^{n} x_i \mathbb{P}(X = x_i) = \sum\limits_{i=1}^{n} x_i p_i$

–**Cas continu** : $\mathbb{E}(X) = \int\limits_{-\infty}^{+\infty} x\, f(x) \mathrm{d}x$

Propriétés générales : soient X et Y deux v.a.

- $\mathbb{E}(a) = a \quad \forall a \in \mathbb{R}$
- $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b \quad \forall (a, b) \in \mathbb{R}^2$
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ si X et Y indépendants
- $\mathbb{E}[h(X)] = \sum\limits_{i} h(x_i)\mathbb{P}(X = x_i)$ si X discret,

  $\mathbb{E}[h(X)] = \int\limits_{-\infty}^{+\infty} h(x) f(x) \mathrm{d}x$ si X continu

Variance

$$\mathrm{Var}(X) = \sigma^2(X) = \mathbb{E}\left\{[X - \mathbb{E}(X)]^2\right\} = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

–**Cas discret** : $\mathrm{Var}(X) = \sum\limits_i [x_i - \mathbb{E}(X)]^2 \mathbb{P}(X = x_i)$

–**Cas continu** : $\mathrm{Var}(X) = \int\limits_{-\infty}^{+\infty} [x - \mathbb{E}(X)]^2 f(x) \mathrm{d}x$

Propriétés générales : soient X une v.a. et $a \in \mathbb{R}$

- $\mathrm{Var}(X) \geq 0$
- $\mathrm{Var}(a) = 0$, $\mathrm{Var}(X + a) = \mathrm{Var}(X)$
- $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$
- si X et Y indépendants, $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
  Attention, la réciproque est fausse !
- $\mathrm{Var}[h(X)] = \mathbb{E}[h^2(X)] - \mathbb{E}^2[h(X)]$

Ecart-type : $\sigma_X(X) = \sigma(X) = \sqrt{\mathrm{Var}(X)}$

Inégalités intéressantes

Inégalité de Markov : si $\forall x < 0, f(x) = 0$ ou $p(x) = 0$ (v.a. ne prenant que des valeurs positives), alors

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a} \quad \forall a > 0$$

Preuve :

$$\mathbb{E}(X) = \int\limits_0^{+\infty} xf(x)\mathrm{d}x \geq \int\limits_a^{+\infty} xf(x)\mathrm{d}x \geq \int\limits_a^{+\infty} af(x)\mathrm{d}x = a\mathbb{P}(X \geq a)$$

Inégalité de Chebyshev : pour une v.a. X quelconque

$$\mathbb{P}(|X - \mu| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2} \quad \forall a > 0$$

Moments d'une variable aléatoire

Moments d'ordre k : $\mathbb{E}(X^k) = \sum_{i=1}^{n} x_i^k p(x_i)$ (cas discret) ou $\int_{-\infty}^{+\infty} x^k f(x) \mathrm{d}x$ (cas continu)

Moments centrés d'ordre k : $\mathbb{E}[(X-\mu)^k] = \sum_{i=1}^{n} (x_i - \mu)^k p(x_i)$ (cas discret) ou $\int_{-\infty}^{+\infty} (x-\mu)^k f(x) \mathrm{d}x$ (cas continu)

Coefficient d'asymétrie (*skewness*) $\gamma = \mathbb{E}[(X-\mu)^3]/\sigma^3$

Coefficient d'aplatissement (*kurtosis*) $\kappa = \mathbb{E}[(X-\mu)^4]/\sigma^4$

Fonction génératrice de moments

$$G(t) = \mathbb{E}\left[e^{tX}\right] = \begin{cases} \sum_{i=1}^{n} e^{tx_i} p(x_i) & \text{(cas discret)} \\ \int_{-\infty}^{+\infty} e^{tx} f(x) \mathrm{d}x & \text{(cas continu)} \end{cases}$$

$$\frac{d^k G(t)}{dt^k} = \begin{cases} \sum_{i=1}^{n} x_i^k e^{tx_i} p(x_i) & \text{(cas discret)} \\ \int_{-\infty}^{+\infty} x^k e^{tx} f(x) \mathrm{d}x & \text{(cas continu)} \end{cases}$$

$$\frac{d^k G(t)}{dt^k} \,|_{t=0} = \mathbb{E}(X^k)$$

Exemples : $G(t) = \dfrac{e^{tb} - e^{ta}}{t(b-a)}$ pour $X \sim \mathcal{U}(a,b)$, $\quad G(t) = \dfrac{\theta}{\theta - t}$ pour $X \sim \mathcal{E}(\theta)$ et $t < \theta$

Fonction caractéristique

Transformée de Fourier de la fonction de densité $f(x)$

$$\Psi(t) = \mathbb{E}\left[e^{itX}\right] = \begin{cases} \sum\limits_{i=1}^{n} e^{itx_i} p(x_i) & \text{cas discret} \\ \int\limits_{-\infty}^{+\infty} e^{itx} f(x)\mathrm{d}x & \text{cas continu} \end{cases}$$

où $i = \sqrt{-1}$ est le nombre imaginaire

Dans le cas continu $f(x) = \dfrac{1}{2\pi} \int\limits_{-\infty}^{+\infty} e^{-itx} \Psi(t)\mathrm{d}t$

Propriétés

- $\Psi(t) = G(it)$
- si $Y = aX + b$, $\Psi_Y(t) = e^{ibt}\Psi_X(at)$
- si $Y = h(X)$, $\Psi_Y(t) = \int\limits_{-\infty}^{+\infty} e^{ith(x)} f_X(x)\mathrm{d}x = \int\limits_{-\infty}^{+\infty} e^{ity} f_Y(y)\mathrm{d}y$

Processus temporels et fonction de survie

Les valeurs de la v.a. correspondent à des temps

Fonction de survie (ou fonction de fiabilité)
$S(x) = \mathbb{P}(X > x) = 1 - F(x)$ avec $x > 0$

Fonction de répartition conditionnelle
$H(x,t) = \mathbb{P}(X \leq x + t | X > x) = \dfrac{F(x+t) - F(x)}{1 - F(x)}$ avec $t > 0$
probabilité d'observer un individu présenter l'événement dans l'intervalle de temps $]x, x+t]$ après l'instant x sachant qu'il était toujours en vie à l'instant x

$$h(x) = \lim_{\Delta x \to 0} \frac{H(x, \Delta x)}{\Delta x} = \frac{f(x)}{1 - F(x)}$$

Processus sans mémoire $\mathbb{P}(X \leq x + t | X > x) = \mathbb{P}(X \leq t)$ avec $t > 0$
$\Rightarrow h(x)$ constant
Seules 2 lois vérifient cette propriété : géométrique (cas discret) et exponentielle (cas continu)

Loi conjointe

Loi du couple (X, Y) pour X et Y deux v.a.

–**Cas discret** :

$$\begin{aligned}\mathbb{P}(X = x \cap Y = y) &= \mathbb{P}(Y = y)\mathbb{P}(X = x|Y = y) \\ &= \mathbb{P}(X = x)\mathbb{P}(Y = y|X = x)\end{aligned}$$

Si X et Y indépendants, $\mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$

–**Cas continu** :

$$\begin{aligned}\mathbb{P}(x_1 < X \leq x_2 \cap y_1 < Y \leq y_2) &= \int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x, y)\mathrm{d}x\mathrm{d}y \\ &= \int_{y_1}^{y_2}\int_{x_1}^{x_2} f(x, y)\mathrm{d}y\mathrm{d}x\end{aligned}$$

Si X et Y indépendants, $f(x, y) = f_X(x)f_Y(y)$

Lois marginales

–**Cas discret** $\begin{cases} \mathbb{P}(X = x) = \sum\limits_{y} \mathbb{P}(X = x \cap Y = y) \\ \mathbb{P}(Y = y) = \sum\limits_{x} \mathbb{P}(X = x \cap Y = y) \end{cases}$

–**Cas continu** $\begin{cases} f_X(x) = \int\limits_{-\infty}^{+\infty} f(x, y)\mathrm{d}y \\ f_Y(y) = \int\limits_{-\infty}^{+\infty} f(x, y)\mathrm{d}x \end{cases}$

$$(X, Y) \sim \mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho) \quad \Rightarrow \quad \begin{cases} X \sim \mathcal{N}\left(\mu_X, \sigma_X^2\right) \\ Y \sim \mathcal{N}\left(\mu_Y, \sigma_Y^2\right) \end{cases}$$

la récriproque n'est pas vraie

Loi conditionnelle

Loi conditionnelle de $Y|X$

–**Cas discret** : $\mathbb{P}(Y = y|X = x) = \dfrac{\mathbb{P}(X = x \cap Y = y)}{\mathbb{P}(X = x)}$ si $\mathbb{P}(X = x) \neq 0$

–**Cas continu** : $f_{Y|X=x}(y) = \dfrac{f(x,y)}{f_X(x)}$ si $f_X(x) \neq 0$

Propriétés générales :

- $\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|X)], \quad \mathbb{E}(Y|W) = \mathbb{E}\left[\mathbb{E}(Y|X,W)|W\right]$
- $\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}[\mathbb{E}(Y|X)]$

Covariance

Définition générale : soient X,Y deux v.a.
$\text{Cov}(X, Y) = \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

Propriétés

- X et Y indépendants $\Rightarrow \text{Cov}(X, Y) = 0$
  $\Leftrightarrow \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$
- Invariance par changement d'origine
  $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$
- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$

Coefficient de corrélation linéaire

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Propriétés

- $-1 \leq \rho(X,Y) \leq 1$, les bornes sont atteintes <u>ssi</u> $Y = a + bX$
- $\rho(X,X) = 1$
- si X et Y sont indépendants, alors $\rho(X,Y) = 0$
  la réciproque est toujours vraie uniquement dans le cas où X et Y suivent une loi normale
- $\rho(\lambda X + a, \mu Y + b) = \frac{\lambda\mu}{|\lambda\mu|}\rho(X,Y)$

Transformation d'une variable aléatoire

$Y = h(X)$

Loi de probabilité

- Si h est bijective, c'est-à-dire $h'(x) \neq 0$ alors
  $f_Y(y) = f_X[u(y)]|u'(y)|$ où $u = h^{-1}$ fonction réciproque de h
- Si h n'est pas bijective, il faut raisonner à partir de la fonction de répartition

Δ-méthode : on note $\mathbb{E}(X) = \mu$

- Espérance $\mathbb{E}(Y) \approx h(\mu), \quad \mathbb{E}[h(X)] \approx h[\mathbb{E}(X)]$
- Variance $\mathrm{Var}(Y) \approx [h'(\mu)]^2 \mathrm{Var}(X)$,
  $\mathrm{Var}[h(X)] \approx (h'[\mathbb{E}(X)])^2 \mathrm{Var}(X)$

Matrice de variance-covariance (1)

Soit un vecteur de v.a. $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$

Définition : $\underset{(p \times p)}{\Sigma} = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}]) \cdot (\mathbf{X} - \mathbb{E}[\mathbf{X}])^\top\right] \quad \Leftrightarrow$
$\forall (i,j) \in \{1, \ldots, p\} \times \{1, \ldots, p\} \quad \Sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$ où
$\mu_i = \mathbb{E}(X_i)$ est l'espérance de la v.a. X_i (scalaire)

Note : matrice de variance comme généralisation de la variance $\sigma^2 = \mathbb{E}[(X - \mu)^2]$ pour un vecteur de dimension 1 (William Feller), ou matrice de covariance entre les composantes scalaires de $\mathbf{X}$ (variances en termes diagonaux)

Source : `http://en.wikipedia.org/wiki/Covariance_matrix`

Matrice de variance-covariance (2)

Σ^{-1} inverse de la matrice de variance-covariance est aussi appelée *concentration matrix* ou *precision matrix*

Propriétés

- $\Sigma = \mathbb{E}[\mathbf{X}\mathbf{X}^\top] - \mu\mu^\top$ où $\mu = \mathbb{E}(\mathbf{X})$ vecteur de dimension p
- Σ est positive semi-définie (*positive semi-definite*)
- $\text{Var}(\mathbf{A}\mathbf{X} + \mathbf{a}) = \mathbf{A}\text{Var}(\mathbf{X})\mathbf{A}^\top$ pour $\mathbf{a}$ vecteur $q \times 1$ et $\mathbf{A}$ matrice $q \times p$

Covariance entre 2 v.a. de dimension multiple (1)

Soit $X \in \mathbb{R}^p, Y \in \mathbb{R}^q$

$$\underset{(p\times q)}{\mathrm{Cov}(X,Y)} = \mathbb{E}\left[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])^\top\right]$$
$$= \underset{(p\times 1)(1\times q)}{\mathbb{E}(XY^\top)} - \underset{(p\times 1)}{\mathbb{E}(X)}\underset{(1\times q)}{\mathbb{E}(Y)^\top}$$

Note : on parle aussi de *cross-covariance* pour distinguer de la matrice de variance-covariance pour un vecteur de v.a.

Covariance entre 2 v.a. de dimension multiple (2)

Propriétés

- $\underset{(p\times q)}{\mathrm{Cov}(\mathbf{X},\mathbf{Y})} \neq \underset{(q\times p)}{\mathrm{Cov}(\mathbf{Y},\mathbf{X})}$ mais $\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = \mathrm{Cov}(\mathbf{Y},\mathbf{X})^\top$
- $\mathrm{Cov}(\mathbf{X}_1+\mathbf{X}_2,\mathbf{Y}) = \mathrm{Cov}(\mathbf{X}_1,\mathbf{Y}) + \mathrm{Cov}(\mathbf{X}_2,\mathbf{Y})$
- Si $p = q$, alors
  $\mathrm{Var}(\mathbf{X}+\mathbf{Y}) = \mathrm{Var}(\mathbf{X}) + \mathrm{Cov}(\mathbf{X},\mathbf{Y}) + \mathrm{Cov}(\mathbf{Y},\mathbf{X}) + \mathrm{Var}(\mathbf{Y})$
- $\mathrm{Cov}(\mathbf{A}\mathbf{X},\mathbf{B}^\top\mathbf{Y}) = \mathbf{A}\,\mathrm{Cov}(\mathbf{X},\mathbf{Y})\,\mathbf{B}$
- Si $\mathbf{X}$ et $\mathbf{Y}$ sont indépendants, alors $\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = 0$

pour $\mathbf{X}$, $\mathbf{X}_1$ et $\mathbf{X}_2$ vecteurs aléatoires $p\times 1$, $\mathbf{Y}$ vecteur aléatoire $q\times 1$, $\mathbf{a}$ vecteur (non aléatoire) $q\times 1$, $\mathbf{A}$ et $\mathbf{B}$ matrices $q\times p$

Loi de Bernoulli

Expérience aléatoire avec 2 résultats possibles : E et $\bar{E}$, avec $\mathbb{P}(E) = p$ et $\mathbb{P}(\bar{E}) = 1 - p = q$

$$\text{v.a. } X = \begin{cases} 1 \text{ si } E \\ 0 \text{ si } \bar{E} \end{cases} \quad \text{notée } X \sim \mathcal{B}(p), \mathbb{P}(X = 1) = p$$

$$\mathbb{E}(X) = p, \quad \mathrm{Var}(X) = pq$$

Exemple : Lancer de pièce
Ensemble des événements $= \{\text{Pile, Face}\}$
Pour une pièce non truquée, $p = 0.5$
donc $\mathbb{E}(X) = 0.5$ et $\mathrm{Var}(X) = 0.25$

Loi binomiale

Somme de n v.a. de Bernoulli indépendantes (tirage avec remise), de même paramètre p

v.a. $Y = \sum_{i=1}^{n} X_i$ notée $Y \sim \mathcal{B}(n, p)$

Y a $(n+1)$ valeurs possibles : pour $k = 0, 1, \ldots, n$

$$\mathbb{P}(Y = k) = \binom{n}{k} p^k q^{n-k} \quad \text{où} \quad \binom{n}{k} \equiv C_n^k = \frac{n!}{k!(n-k)!}$$

$$\mathbb{E}(Y) = np, \quad \mathrm{Var}(Y) = npq$$

Note : quand n devient grand, le calcul des factorielles devient laborieux $\rightarrow$ voir approximations plus loin

Loi hypergéométrique

Répétition d'une épreuve de Bernoulli sans remise $\rightarrow$ Somme de n v.a. de Bernoulli non indépendantes parmi N

Y a pour valeurs possibles : $max(0, n+k-N), \ldots, min(n,k)$

$$\mathbb{P}(Y=x) = \frac{C_k^x C_{N-k}^{n-x}}{C_N^n}$$

Loi géométrique

Nombre de fois qu'il faut répéter une épreuve de Bernoulli pour obtenir une première réussite

Y a pour valeurs possibles $\{1, 2, \ldots\}$

$$\mathbb{P}(Y = k) = p(1-p)^{k-1}$$

$$F(y) = 1 - (1-p)^y$$

Loi de Pascal ou loi binomiale négative

Nombre de répétitions d'épreuves de Bernoulli indépendantes nécessaire pour obtenir k réussites
Somme de k variables géométriques indépendantes

$$\mathbb{P}(Y = x) = C_{x-1}^{k-1} p^k (1-p)^{x-k}, \quad x = k, k+1, \ldots$$

Loi de Poisson

La v.a. X suit une loi de Poisson de paramètre λ, $\lambda > 0$ (réel), si X est un comptage dévénements avec

$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad \forall k \geq 0$$

$$X \sim \mathcal{P}(\lambda)$$

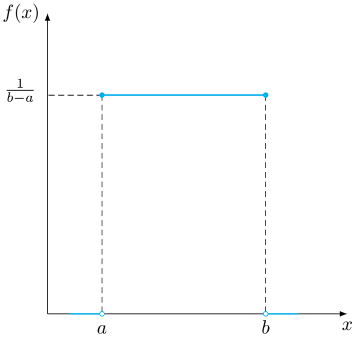
$$\mathbb{E}(X) = \lambda, \quad \mathrm{Var}(X) = \lambda$$

Note : $\mathbb{E}(X) = \mathrm{Var}(X)$ est une condition nécessaire mais pas suffisante de la loi de Poisson, elle constitue une contrainte forte pour la modélisation

Approximation de la loi binomiale par la loi de Poisson
Conditions : $n \to \infty$ et $p \to 0$, alors $\mathcal{B}(n, p) \to \mathcal{P}(\lambda)$ avec $np = \lambda$

Loi uniforme

Une variable aléatoire continue X est dite de loi (ou distribution) uniforme sur $[a, b]$ si $f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$



On note $X \sim \mathcal{U}(a, b)$; on a $\mathbb{E}(X) = \dfrac{a+b}{2}$, $\text{Var}(X) = \dfrac{(b-a)^2}{12}$

Loi exponentielle

Une v.a. X est dite de loi exponentielle de paramètre $\theta > 0$ si

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

$$X \sim \mathcal{E}(\theta)$$

$$\mathbb{E}(X) = \frac{1}{\theta}, \quad \mathrm{Var}(X) = \frac{1}{\theta^2}$$

Exemples : temps d'attente, analyse de survie

Loi normale (Laplace-Gauss)

Si $Y \sim \mathcal{N}(\mu, \sigma^2)$, sa densité est $f(y) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{y-\mu}{\sigma}\right)^2\right]$

$Z = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ loi normale centrée réduite $\rightarrow$ tables

Théorème central limite : La loi de probabilité de la somme de n variables indépendantes de même loi (continue quelconque) tend vers une loi normale quand n est grand

Loi Beta

Une v.a. X est dite de loi beta de paramètres $a > 0$ et $b > 0$ si

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} & \text{si } x \in [0,1] \\ 0 & \text{sinon} \end{cases}$$

$$X \sim \mathrm{Beta}(a, b)$$

$$\mathbb{E}(X) = \frac{a}{a+b}, \quad \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Loi du Chi-deux

Somme de ν v.a. de loi normale centrée réduite $\mathcal{N}(0,1)$ indépendantes
v.a. $Y = \sum_{i=1}^{\nu} X_i^2$ notée $Y \sim \chi^2_\nu$ avec $\nu \in \mathbb{N}$

$$f(y) = \frac{e^{-y/2} y^{\nu/2-1}}{2^{\nu/2}\Gamma(\nu/2)} \quad \text{pour } y \geq 0$$

$$\mathbb{E}(Y) = \nu, \quad \mathrm{Var}(Y) = 2\nu$$

En pratique : table

Loi de Student

Si $Z \sim \mathcal{N}(0,1)$ et $U \sim \chi^2_\nu$ sont indépendantes alors la v.a. $Y = Z/\sqrt{\frac{U}{\nu}}$ notée $Y \sim t_\nu$ suit une loi de student à ν degrés de liberté.

$$f(y) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\dfrac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{y^2}{\nu}\right)^{-\dfrac{\nu+1}{2}}$$

$$\mathbb{E}(Y) = 0 \quad \text{si} \quad \nu > 1, \quad \mathrm{Var}(Y) = \frac{\nu}{\nu-2} \quad \text{si} \quad \nu > 2$$

Loi de Fisher-Snedecor

Générateur aléatoire uniforme

Formule de Lehmer : $x_{i+1} = ax_i \mod b \quad i = 0, 1, \ldots, n-1$ où $b \in \mathbb{N}$ et $a \in \{1, 2, \ldots, b-1\}$
Permet de générer une séquence pseudo-aléatoire de nombres entiers allant de 0 à $b-1$ à partir du germe (*seed* = valeur de départ) $x_0 \in \{1, \ldots, b-1\}$

Avec un bon choix des paramètres a et b, les valeurs $x_1, x_2, \ldots$ peuvent raisonnablement être considérées comme des réalisations de v.a. $X_1, X_2, \ldots$ indépendantes et identiquement distribuées, de fonction de probabilité $p_X(x) = \dfrac{1}{b} \quad x = 0, 1, \ldots, b-1$

$$a = 2^7 - 1, \quad b = 2^{31} - 1$$

Pour obtenir une séquence de nombres pseudo-aléatoires uniformément distribués entre 0 et 1, prendre $y_i = \dfrac{x_i}{b} \quad i = 1, \ldots, n$

Génération d'une variable aléatoire par inversion

$$\begin{cases} X \sim \mathcal{U}(0,1) \\ Y = h^{-1}(X) \\ h^{-1} \nearrow \end{cases} \quad \Rightarrow \quad F_Y(y) = \begin{cases} 0 & y < h^{-1}(0) \\ h(y) & y \in \left[h^{-1}(0), h^{-1}(1)\right] \\ 1 & y > h^{-1}(1) \end{cases}$$

Exemple– Pour générer $y_1, y_2, \ldots$, réalisations de $Y \sim \mathcal{E}(\lambda)$ à partir de $x_1, x_2, \ldots$, réalisations de $X \sim \mathcal{U}(0,1)$

$$F_Y(y) = h(y) = 1 - e^{-\lambda y} \quad \Rightarrow \quad F_Y^{-1}(x) = h^{-1}(x) = -\frac{1}{\lambda} \ln(1-x)$$

d'où la transformation adéquate $y_i = -\dfrac{1}{\lambda} \ln(1 - x_i)$

Deuxième partie II

Statistique inférentielle

Introduction à la statistique

Soit un échantillon de n individus indépendants pour lesquels on observe les réalisations $x_1, x_2, \ldots, x_n$ de la v.a. X
Définition : on appelle statistique toute fonction exprimée en terme des variables composant l'échantillon $T = h(X_1, \ldots, X_n)$
Cette statistique est fonction de l'échantillon, sa valeur réalisée sur l'échantillon $t = h(x_1, \ldots, x_n)$
On note $x_{(1)}, \ldots, x_{(n)}$ les valeurs ordonnées par ordre croissant des réalisations de X sur l'échantillon

Paramètres/statistiques de position : "où est la mesure ?"

Minimum $L = \min(X_1, \ldots, X_n) = X_{(1)}$
Maximum $U = \max(X_1, \ldots, X_n) = X_{(n)}$
Mode M_o : valeur (ou centre de la classe des valeurs) la plus fréquente, c'est-à-dire x_m telle que $\forall x, p(x_m) \geq p(x)$ ou $f(x_m) \geq f(x)$

Médiane M_e : valeur qui partage les observations ordonnées en 2 groupes d'effectifs égaux, c'est-à-dire il y a autant d'observations strictement inférieures à M_e que d'observations strictement supérieures à M_e : $\mathbb{P}(X \leq \tilde{m}) = F(\tilde{m}) = 1/2$

$$\tilde{X} = \begin{cases} X_{(\frac{n+1}{2})} & n \text{ impair} \\ \dfrac{1}{2}\left[X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\right] & n \text{ pair} \end{cases}$$

Quantile (ou fractile) : x_{100p} telle que $\mathbb{P}(X \leq x_{100p}) = p, \quad p \in [0, 1]$
Exemples : médiane x_{50} ; quartiles x_{25}, x_{50}, x_{75} ; déciles $x_{10}, x_{20}, \ldots, x_{90}$; centiles $x_1, x_2, \ldots, x_{99}$

Moyenne arithmétique $\bar{X} = \dfrac{1}{n} \sum_{i=1}^{n} X_i$

Paramètres/statistiques d'échelle (dispersion/précision de la mesure)

Etendue (*range*) $R = U - L = X_{(n)} - X_{(1)}$

Intervalle interquartiles $x_{75} - x_{25}$

Variance $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \dfrac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \dfrac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \dfrac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$

Ecart-type $S = \sqrt{S^2} \geq 0$ s'exprime dans la même unité que les valeurs observées

Coefficient de variation $\mathrm{CV} = \dfrac{s}{\bar{x}}$ sans unité, souvent exprimé en %

Estimateurs ponctuels des moments

$$\widehat{\mathbb{E}}[X] = \hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\widehat{\mathbb{E}}[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

$$\widehat{\mathbb{E}}[(X-\mu)^k] = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^k$$

$$\widehat{\sigma}^2 = S^2 = \frac{n}{n-1}\hat{\mathbb{E}}[(X-\mu)^k]$$

$$\widehat{\sigma} = S = \sqrt{S^2}$$

Estimateurs ponctuels de "probabilités"

$$\widehat{\mathbb{P}}(a < X \leq b) = \frac{\#\{a < X_i \leq b\}}{n}$$

$$\widehat{F}(x) = \frac{\#\{X_i \leq x\}}{n}$$

$$\widehat{\mathbb{P}}\big(X = x\big) = \frac{\#\{X_i = x\}}{n}$$

$$\widehat{f}(x) = \frac{\#\{x - \Delta x < X_i \leq x + \Delta x\}}{2n\Delta x} \qquad \Delta x > 0 \text{ petit}$$

Représentations graphiques pour une variable aléatoire

Variable discrète : diagramme en bâtons, *pie* (camembert)...

Variable continue : histogramme des fréquences $\hat{\mathbb{P}}(a_k < X \leq a_{k+1})$, *box plot* (boîte à moustaches)...

Statistiques d'ordre

Si les X_i sont i.i.d. de loi de probabilité continue $f(x)$ alors

$$f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} \left[F(x)\right]^{k-1} \left[1 - F(x)\right]^{n-k} f(x)$$

En particulier, pour $\min(X_1, \ldots, X_n) = X_{(1)}$ et $\max(X_1, \ldots, X_n) = X_{(n)}$

$$f_{(1)}(x) = n \left[1 - F(x)\right]^{n-1} f(x), \quad f_{(n)}(x) = n \left[F(x)\right]^{n-1} f(x)$$

Qualités d'un estimateur

θ valeur théorique du paramètre, $\widehat{\Theta}$ estimateur pour ce paramètre

Biais $b\big(\widehat{\Theta}\big) = \mathbb{E}\big[\widehat{\Theta}\big] - \theta$

- Estimateur sans biais (non biaisé) $b\big(\widehat{\Theta}\big) = 0$
- Estimateur asymtotiquement sans biais $\lim\limits_{n\to\infty} b\big(\widehat{\Theta}\big) = 0$

Qualités d'un estimateur

θ valeur théorique du paramètre, $\widehat{\Theta}$ estimateur pour ce paramètre

Biais $b(\widehat{\Theta}) = \mathbb{E}\big[\widehat{\Theta}\big] - \theta$

- Estimateur sans biais (non biaisé) $b(\widehat{\Theta}) = 0$
- Estimateur asymtotiquement sans biais $\lim\limits_{n\to\infty} b(\widehat{\Theta}) = 0$

$\bar{X}$ et S^2 sont des estimateurs sans biais de μ et σ^2, respectivement
$\widehat{\mathbb{E}}\big[(X-\mu)^k\big]$ est un estimateur biaisé mais asymptotiquement sans biais de σ^2

Efficacité et consistance

Efficacité : Erreur quadratique moyenne
$EQM(\hat{\Theta}) = \mathbb{E}[(\hat{\Theta} - \theta)^2] = \mathrm{Var}(\hat{\Theta}) + b^2(\hat{\Theta})$ finie et minimale

Efficacité et consistance

Efficacité : Erreur quadratique moyenne
$EQM(\hat{\Theta}) = \mathbb{E}[(\hat{\Theta} - \theta)^2] = \mathrm{Var}(\hat{\Theta}) + b^2(\hat{\Theta})$ finie et minimale

Consistance : $\lim_{n \to \infty} \mathbb{P}[|\hat{\Theta} - \theta| > \varepsilon] = 0$ convergence en probabilité

Méthode des moments

En supposant que les moments théoriques $\mathbb{E}[X], \mathbb{E}[X^2], \ldots, \mathbb{E}[X^p]$ peuvent s'écrire comme fonctions des paramètres inconnus $\theta_1, \theta_2, \ldots, \theta_p$

$$\begin{cases} \mathbb{E}[X] & = h_1\big(\theta_1, \ldots, \theta_p\big) = h_1(\theta) \\ \mathbb{E}\big[X^2\big] & = h_2\big(\theta_1, \ldots, \theta_p\big) = h_2(\theta) \\ & \vdots \\ \mathbb{E}\big[X^p\big] & = h_p\big(\theta_1, \ldots, \theta_p\big) = h_p(\theta) \end{cases}$$

Méthode des moments

En supposant que les moments théoriques $\mathbb{E}[X], \mathbb{E}[X^2], \ldots, \mathbb{E}[X^p]$ peuvent s'écrire comme fonctions des paramètres inconnus $\theta_1, \theta_2, \ldots, \theta_p$

$$\begin{cases} \mathbb{E}[X] & = h_1\big(\theta_1, \ldots, \theta_p\big) = h_1(\theta) \\ \mathbb{E}\big[X^2\big] & = h_2\big(\theta_1, \ldots, \theta_p\big) = h_2(\theta) \\ & \vdots \\ \mathbb{E}\big[X^p\big] & = h_p\big(\theta_1, \ldots, \theta_p\big) = h_p(\theta) \end{cases}$$

Si ce système d'équations peut être résolu par rapport à $\theta_1, \ldots, \theta_p$, on obtient

$$\begin{cases} \theta_1 = g_1\Big(\mathbb{E}[X], \mathbb{E}\big[X^2\big], \ldots, \mathbb{E}\big[X^p\big]\Big) \\ \quad \vdots \\ \theta_p = g_p\Big(\mathbb{E}[X], \mathbb{E}\big[X^2\big], \ldots, \mathbb{E}\big[X^p\big]\Big) \end{cases}$$

Méthode des moments

On remplace les moments théoriques par leurs estimateurs
$\widehat{\mathbb{E}}[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad \forall k = 1, \ldots, p$

$$\begin{cases} \widehat{\Theta}_1 = g_1\left(\widehat{\mathbb{E}}[X], \widehat{\mathbb{E}}\left[X^2\right], \ldots, \widehat{\mathbb{E}}\left[X^p\right]\right) \\ \quad\vdots \\ \widehat{\Theta}_p = g_p\left(\widehat{\mathbb{E}}[X], \widehat{\mathbb{E}}\left[X^2\right], \ldots, \widehat{\mathbb{E}}\left[X^p\right]\right) \end{cases}$$

Note : on peut aussi utiliser les moments centrés $\mathbb{E}[(X-\mu)^k]$

Estimateurs pour les lois usuelles par la méthode des moments

$$X \sim \mathcal{B}(n,p) \quad \longrightarrow \quad \widehat{p} = \frac{1}{n}\bar{X}$$

$$X \sim \mathcal{G}(p) \quad \longrightarrow \quad \widehat{p} = \frac{1}{\bar{X}}$$

$$X \sim \mathcal{P}(\mu) \quad \longrightarrow \quad \widehat{\mu} = \bar{X}$$

Certains de ces estimateurs sont biaisés mais ils sont tous asymptotiquement sans biais

Estimateurs pour les lois usuelles par la méthode des moments

$$X \sim \mathcal{U}(0,b) \quad \longrightarrow \quad \widehat{b} = 2\bar{X}$$

$$X \sim \mathcal{U}(a,b) \quad \longrightarrow \quad \widehat{a} = \bar{X} - \sqrt{3\frac{n-1}{n}S^2}, \widehat{b} = \bar{X} + \sqrt{3\frac{n-1}{n}S^2}$$

$$X \sim \mathcal{E}(\lambda) \quad \longrightarrow \quad \widehat{\lambda} = \tfrac{1}{\bar{X}}$$

$$X \sim \mathcal{N}(0,\sigma^2) \quad \longrightarrow \quad \widehat{\sigma}^2 = \widehat{\mathbb{E}}[X^2]$$

$$X \sim \mathcal{N}(\mu,\sigma^2) \quad \longrightarrow \quad \widehat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{n-1}{n}S^2$$

Certains de ces estimateurs sont biaisés mais ils sont tous asymptotiquement sans biais