2.3 Valeurs caractéristiques

Mesure de tendance centrale

Définition: L'**espérance** d'une variable aléatoire X est

$$\mathbb{E}(X) = \begin{cases} \sum_i x_i \, f_X(x_i), & X \text{ discrète}, \\ \int_{-\infty}^{\infty} x f_X(x) \, \mathrm{d}x, & X \text{ continue}, \end{cases}$$

si la somme/intégrale converge

Propriétés :

- si $X_1, \ldots, X_n$ sont des variables aléatoires et $a, b_1, \ldots, b_n$ des constantes, alors
$$\mathbb{E}\left(a + \sum_{i=1}^{n} b_i X_i\right) = a + \sum_{i=1}^{n} b_i \mathbb{E}(X_i)$$
- pour g fonction 'raisonnable', $\mathbb{E}\{g(X)\} = \begin{cases} \sum_i g(x_i) f_X(x_i), & X \text{ discrète} \\ \\ \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x, & X \text{ continue} \end{cases}$
- Si X, Y sont des variables aléatoires on définit $\mathbb{E}[g(X, Y)]$ de la même manière à l'aide de $f_{X,Y}$
- si X, Y sont indépendantes et g, h des fonctions 'raisonnables', alors
$$\mathbb{E}\{g(X)h(Y)\} = \mathbb{E}\{g(X)\}\mathbb{E}\{h(Y)\}$$

Exemple binomial

Exemple Pour $X \sim \mathcal{B}(m, p)$, trouver $\mathbb{E}(X)$.

Exemple Poisson

Exemple Pour $X \sim \mathrm{Poiss}(\lambda)$, trouver $\mathbb{E}(X)$ et $\mathbb{E}\{X(X-1)\}$.

Exemples

Exemple Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, trouver $\mathbb{E}(X)$.

Mesure de dispersion

Définition: La **variance** d'une variable aléatoire X est définie comme

$$\text{var}(X) = \mathbb{E}[\{X - \mathbb{E}(X)\}^2] = \cdots = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Propriétés :

- Interprétation physique : variance $\equiv$ moment d'inertie relatif au centre de masse
- $\text{var}(X) \geq 0$, et $\text{var}(X) = 0$ implique que X est constante
- la **déviation standard** de X est définie comme $\mathrm{sd}(X) = \sqrt{\text{var}(X)} \geq 0$
- si a, b sont des constantes, alors $\text{var}(a + bX) = b^2\text{var}(X)$
- si $X_1, \ldots, X_n$ sont indépendantes et $a, b_1, \ldots, b_n$ des constantes, alors

$$\text{var}\left(a + \sum_{i=1}^{n} b_i X_i\right) = \sum_{i=1}^{n} b_i^2 \text{var}(X_i)$$

Exemple Si $X \sim \mathrm{Poiss}(\lambda)$, montrer que $\text{var}(X) = \lambda$.

Exemple Si $X \sim \mathcal{B}(m, p)$, montrer que $\text{var}(X) = m\,p(1 - p)$.

Exemple Si $X \sim \mathcal{N}(\mu, \sigma^2)$, montrer que $\text{var}(X) = \sigma^2$.

Exemples : variance

Exemple Si $X \sim \mathrm{Poiss}(\lambda)$, montrer que $\mathrm{var}(X) = \lambda$.

Exemple Si $X \sim \mathcal{B}(m, p)$, montrer que $\mathrm{var}(X) = m\,p(1-p)$.

Exemple variance loi normale

Exemple Si $X \sim \mathcal{N}(\mu, \sigma^2)$, montrer que $\text{var}(X) = \sigma^2$.

Covariance

Définition: La **covariance** des variables aléatoires X, Y est

$$\mathrm{cov}(X, Y) = \mathbb{E}\left[\{X - \mathbb{E}(X)\}\{Y - \mathbb{E}(Y)\}\right] = \cdots = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Interprétation : C'est une mesure de dépendance **linéaire** entre X et Y

Propriétés :

- la covariance dépend des unités dont on mesure X, Y
- $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$
- $\mathrm{cov}(X, X) = \mathrm{var}(X)$
- $\mathrm{cov}(X+Y, Z+W) = \mathrm{cov}(X, Z)+\mathrm{cov}(Y, Z)+\mathrm{cov}(X, W)+\mathrm{cov}(Y, W)$
- si a, b, c, d sont des constantes, alors $\mathrm{cov}(a\,X + b,\, c\,Y + d) = ac\,\mathrm{cov}(X, Y)$
- $\mathrm{var}(X \pm Y) = \mathrm{var}(X) + \mathrm{var}(Y) \pm 2\mathrm{cov}(X, Y)$
- si X et Y sont indépendantes, alors $\mathrm{cov}(X, Y) = 0$. Mais attention, l'inverse **n'est pas vraie** en général !

Exemple

Exemple (voir diapositive 123) Soient X et Y de densité conjointe

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{si} \quad 0<x<1,\ 0<y<1, \\ 0 & \text{sinon.} \end{cases}$$

Trouver $\mathrm{Var}(X)$, $\mathrm{Var}(Y)$, et $\mathrm{Cov}(X,Y)$.

Corrélation

Définition: La **corrélation** de X et Y est

$$\rho_{X,Y} = \rho(X,Y) = \text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

(zéro si une des variances est zéro).

Propriétés :

- $\rho_{X,Y}$ mesure la dépendance **linéaire** (et seulement linéaire !) entre X et Y
- $\rho(a+bX, c+dY) = \mathrm{sign}(bd)\rho(X,Y)$
- $\text{corr}(X,Y) = \text{corr}(Y,X)$
- $\text{corr}(X,X) = 1$ (si X n'est pas constante)
- $\text{corr}(X,-X) = -1$ (si X n'est pas constante)
- $-1 \leq \text{corr}(X,Y) \leq 1$ (inegalité de Cauchy–Schwarz)
- si X et Y sont indépendantes, alors $\text{corr}(X,Y) = 0$, mais la réciproque est faux !
- corrélation $\neq$ causalité !

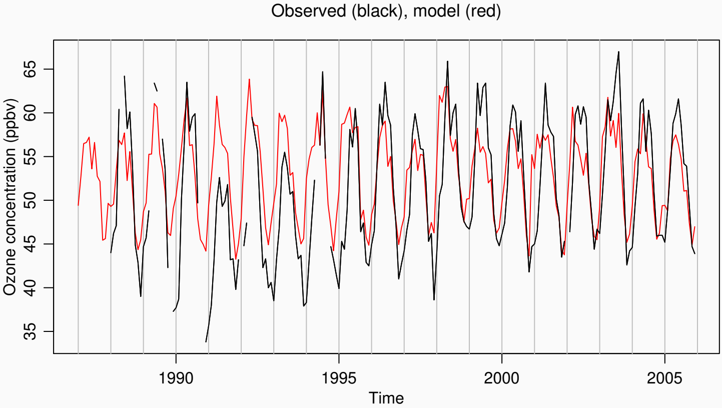
Corrélation empirique

Version empirique (si $\Pr((X = x_i, Y = y_i) = 1/n$ pour $i = 1, \ldots, n$)

$$\frac{n^{-1}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\left\{n^{-1}\sum_{j=1}^{n}(x_j - \bar{x})^2 \times n^{-1}\sum_{j=1}^{n}(y_j - \bar{y})^2\right\}^{1/2}},$$

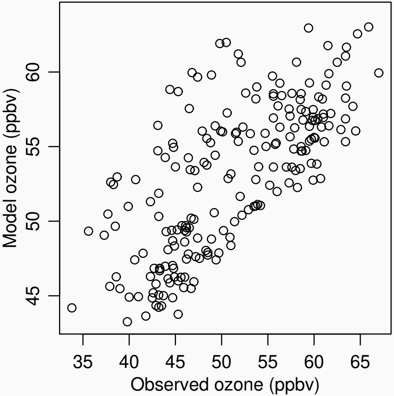
Exemple : ozone atmosphérique

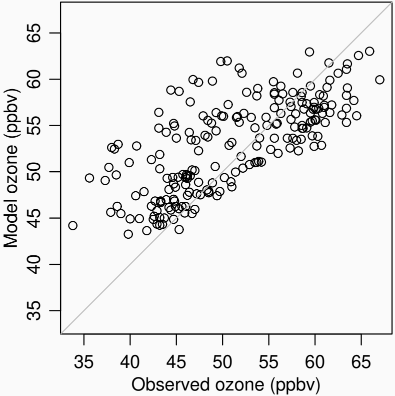
Prof. Isabelle Bey (SIE) : observations de la concentration d'ozone au Jungfraujoch de janvier 1987 à décembre 2005 (quelques valeurs manquantes), et résultats d'une modélisation.



La modélisation vous paraît-elle bonne ?

Exemple : ozone atmosphérique

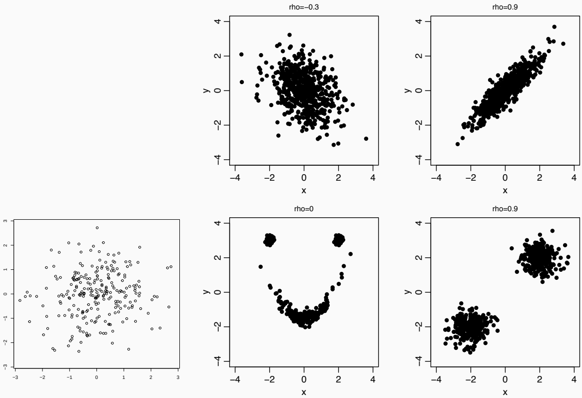




La corrélation empirique est $\rho = 0.707$.

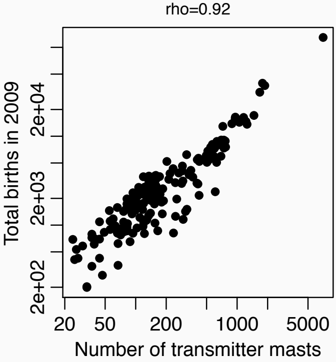
Limitations de la corrélation

- ρ mesure la dépendance linéaire (panneaux supérieurs)
- On peut avoir $\rho \approx 0$, mais dépendance forte mais non-linéaire (en bas au milieu)
- Une corrélation pourrait être forte mais **specieuse**, comme en bas à droite, ou deux sous-groupes, chacun sans corrélation, sont combinés
- Une corrélation entre deux variables n'implique pas une causalité entre elles



Corrélation $\neq$ causalité

Deux variables peuvent être très corrélées sans lien de causalité. Le graphique à gauche ici montre une corrélation forte entre le nombre de naissances et les mâts de communication dans les villes anglaises ...



Danger !

- Les espérances/variances/covariances/corrélations ne sont pas définies si les intégrales/sommes ne convergent pas
- Ceci est notamment le cas lorsque la distribution de X a des **queues lourdes** : la densité de X décroît trop lentement vers zéro, et X a une probabilité élevée de prendre des valeurs énormes.

Exemple

- Considérons la fonction de densité $f(x) = \alpha x^{-1-\alpha}$ sur $[1, \infty)$ et $f(x) = 0$ pour $x < 1$ (loi **Pareto**). Pour $r \in \mathbb{R}$ on a

$$\mathbb{E}(X^r) = \alpha \int_1^\infty x^{r-1-\alpha} dx = \begin{cases} \frac{\alpha}{\alpha - r} & r < \alpha \\ \infty & r \geq \alpha \end{cases}$$

- En particulier, $\mathbb{E}(X) < \infty$ si et seulement si $\alpha > 1$, et $\mathrm{var}(X) < \infty$ si et seulement si $\alpha > 2$
- Pour α petit la densité tend lentement vers zéro

Espérance d'une variable aléatoire mixte

Théorème de l'espérance totale Pour une partition $A_1, A_2, \ldots$

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X|A_i)\mathrm{Pr}(A_i)$$

Exemple : pluie (diapositive 112) La probabilité qu'il pleuve pendant la journée est 0.2. S'il pleut, la quantité de pluie qui tombe suit une loi exponentielle de parametre $\lambda = 0.05\text{mm}^{-1}$. Trouver l'espérance de la quantité de pluie journalière.

Quantiles

Définition: Soit $0 < p < 1$. On définit le $p$ième **quantile** d'une fonction de répartition F par

$$x_p = \inf\{x : F(x) \geq p\}$$

- Pour des variables aléatoires continues, $F(x_p) = p$, donc x_p est tel que $\Pr(X \leq x_p) = p$
- Pour la plupart des variables aléatoires continues, ceci implique que $\boxed{x_p = F^{-1}(p)}$, où F^{-1} est la fonction inverse de F
- "La plupart" : celles ayant une fonction de densité strictement positive (sur $\{x : 0 < F(x) < 1\}$)
- Pour des variables aléatoires discrètes la situation est plus complexe
- Les quantiles empiriques (diapositive 35) sont des estimations (cf les prochains cours) des quantiles à partir des données à disposition.

En particulier, on appelle le 0.5ème quantile la **médiane** de F

Exemple quantiles

Exemple Calculer les quantiles des lois (a) $U(a, b)$, (b) Pareto (diapositive 141)