PRÉALABLES

Espérance, variance et covariance

Espérance mathématique, variance et covariance

L'espérance et la variance d'une variable discrète *Y* dont les valeurs sont avec probabilités correspondantes $p(y_1)$, $p(y_2)$, … , $p(y_n)$ sont définies par, respectivement,

- E(*Y*) = $\sum_i^n y_i P(Y = y_i) = \sum_i^n y_i\, p(y_i)$ et
- V(*Y*) = $\sum_i^n [y_i - E(Y)]^2 P(Y = y_i) = \sum_i^n [y_i - E(Y)]^2\, p(y_i)$

Covariance

Soit (*X* ; *Y*) un couple aléatoire discret de fonction de probabilité conjointe *p*(*x*;*y*). Leur *covariance* est définie par

- Cov(X ; Y) = E{[(*X*-E(*X*))][(*Y*-E(*Y*))]}= E(*XY*) – E(*X*)E(*Y*) .
  $= \sum_y \sum_x [x - \mathrm{E}(X)][y - \mathrm{E}(Y))]P(X = x; Y = y) = \sum_y \sum_x [x - \mathrm{E}(X)][y - \mathrm{E}(Y))]p(x;y)$.

Espérance et variance et covariance conditionnelles

L'espérance et la variance *conditionnelles* de *Y* étant donné *X* = *x* sont données par

- E(*Y*|*X*=*x*) = $\sum_i^n y_i P(Y = y_i \mid X = x)$
- V(*Y*|*X*=*x*) = $\sum_i^n [y_i - E(Y)]^2 P(Y = y_i \mid X = x)$

La covariance *conditionnelle* de (*X* ; *Y*) étant donnée *Z* = *z* est donnée par

- Cov(*X* ; *Y* | *Z*=z) = $\sum_y \sum_x [x - \mathrm{E}(X)][y - \mathrm{E}(Y))]P(X = x; Y = y \mid Z = z)$

Calcul d'une espérance, d'une variance et d'une covariance par conditionnement

- E(*Y*) = E[E(*Y*|*X*)]
- V(*Y*) = E[V(*Y*|*X*)] + V[E(*Y*|*X*)]
- Cov(*X* ; *Y*) = Cov[E(*X*|*Y*);E(*Y*|*Z*)]+E[Cov(*X*;*Y*|*Z*)]

Notions d'estimation

Un estimateur d'un paramètre θ est une fonction $\hat{\theta}$ des observations d'un échantillon.

Propriétés d'un estimateur

Biais d'un estimateur

Le *biais* B($\hat{\theta}$) d'un estimateur $\hat{\theta}$ est défini par

$$B(\hat{\theta}) = E(\hat{\theta})-\theta$$

Estimateur sans biais

Un estimateur $\hat{\theta}$ est dit *sans biais* si $B(\hat{\theta}) = 0$, c'est-à-dire, si $E(\hat{\theta}) = \theta$.

Variance d'un estimateur

La variance $V(\hat{\theta})$ d'un estimateur sans biais est une mesure de la qualité de l'estimateur : plus $V(\hat{\theta})$ est petit, mieux c'est. Dans le cas d'un estimateur biaisé, la variance ne peut mesurer la précision la précision d'un estimateur.

Erreur quadratique moyenne

Une mesure de la précision d'un estimateur est donnée par l'erreur quadratique moyenne. L'erreur quadratique moyenne d'un estimateur $\hat{\theta}$ est définie par $EQM(\hat{\theta}) = E[(\hat{\theta}-\theta)^2]$. On montre aisément que

$$EQM(\hat{\theta}) = V(\hat{\theta}) + [B(\hat{\theta})]^2$$

L'erreur quadratique moyenne coïncide avec la variance lorsque $\hat{\theta}$ est sans biais.

Un estimateur est dit *asymptotiquement sans biais* si $B(\theta) \to 0$ lorsque $n \to \infty$.

Analyse combinatoire

Tiragez sans remise

On tire un échantillon ω (un ensemble aléatoire) de n unités d'une population de taille N, sans remise.

Soit Ω l'ensemble de tous les échantillons (ou toutes les valeurs de ω) possibles.

- La cardinalité de Ω est $\binom{N}{n} = \dfrac{N!}{n!(N-n)!}$.
    - Pour tout ensemble s de n unités, la probabilté $P(\omega = s)$ est $p(s) = \dfrac{1}{\binom{N}{n}}$.
- Le nombre d'échantillons qui contiennent une unité donnée i est $\binom{N-1}{n-1}$.
    - Donc la probabilité qu'une unité donnée i se trouve dans l'échantillon est $P(i \in \omega) = \dfrac{\binom{N-1}{n-1}}{\binom{N}{n}} = \dfrac{n}{N}$.
- Le nombre d'échantillons qui contiennent deux unités données i et j, $(i \neq j)$ est $\binom{N-2}{n-2}$.
    - Donc la probabilité que deux unités données i et j, $(i \neq j)$, se trouvent dans l'échantillon est $\dfrac{\binom{N-2}{n-2}}{\binom{N}{n}} = \dfrac{n(n-1)}{N(N-1)}$.

Tirages avec remise

On effectue n tirages *avec remise* d'une population de N unités. Soit Ω^* l'ensemble des échantillons possibles ω^*. Si Ω^* énumère les échantillons *ordonnés*, alors sa cardinalité est N^n et ses éléments sont équiprobables. Soit $\boldsymbol{Z} = [Z_1; Z_2; \ldots ; Z_n]$ un vecteur aléatoire où Z_i est le nombre de fois où l'unité i de la population est tirée. Alors

Z_i est de loi binomiale de paramètres n et $p = 1/N$: $\mathrm{P}(Z_i = z_i) = \binom{n}{z_i}\left(\frac{1}{N}\right)^{z_i}\left(1-\frac{1}{N}\right)^{n-z_i}$

$\boldsymbol{Z}$ est de loi multinomiale de paramètres n et $p_1 = p_2 = \ldots = p_N = 1/N$:

$$P(Z_1 = z_1;\, Z_2 = z_2;\cdots;Z_n = z_n) = \frac{n!}{N^n \prod_{i=1}^{N} z_i!}.$$

$$P(Z_i = z_i;\, Z_j = z_j) = \frac{n!(n - z_i - z_j)}{N^n z_i!z_j!(n - z_i - z_j)!}$$

$$\mathrm{Cov}(Z_i ; Z_j) = -np_ip_j = -n/N^2$$