AnSeDa - Résumé Olivier D'Ancona

Espérance

Soit X une variable aléatoire à valeurs dans $\mathbb{R}$, on définit l'espérance de X par :

Cas discret : $\mathrm{E}[X] = \sum_{i=1}^{\infty} x_i \cdot p_i$

Cas continu : $\mathrm{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$ où $f(x)$ est la densité de probabilité de X.

Propriétés

- $\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$
- $\mathrm{E}[aX] = a\mathrm{E}[X]$
- $[X^2] = [X]^2 + \mathrm{Var}[X]$
- $\mathrm{E}[XY] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$ if X and Y are independent.
- $\mathrm{E}[c] = c$ if c is a constant.

Variance

Soit X une variable aléatoire à valeurs dans $\mathbb{R}$. Cas discret // continu :

$$\mathrm{Var}[X] = \sum_{i=1}^{\infty} (x_i - \mathrm{E}[X])^2 \cdot p_i$$
$$= \mathrm{E}[(X_i - \mu)^2] = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 \cdot f(x)\, dx$$

$\mathrm{Var}[X] = \sigma^2$

Propriétés

- $\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$
- $\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$
- $\mathrm{Var}[XY] = \mathrm{Var}[X] \cdot \mathrm{Var}[Y] + \mathrm{Var}[X] \cdot \mathrm{E}[Y]^2 + \mathrm{Var}[Y] \cdot \mathrm{E}[X]^2$ if X and Y are independent.
- $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y]$
- $\mathrm{Var}[c] = 0$ if c is a constant.

Covariance

Soit X et Y deux variables aléatoires, la covariance entre X et Y est définie par :

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$
$$= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

Propriétés

- $\mathrm{Cov}(X,c) = 0$ si c est une constante
- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(aX+b, cY+d) = ac \cdot \mathrm{Cov}(X,Y)$
- $\mathrm{Cov}(X+Y, Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$
- $\mathrm{Cov}(X,Y) = 0$ Si X et Y sont indépendants

Corrélation

Soit X et Y deux variables aléatoires, le coefficient de corrélation entre X et Y est défini par :

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Propriétés

- $\rho_{XY} = \rho_{YX}$
- $-1 \leq \rho_{XY} \leq 1$
- $\rho_{XY} = 1$ ou -1 indique une corrélation linéaire parfaite
- $\rho_{XY} = 0$ si X et Y sont indépendants linéairement
- Invariance des changements linéaires : $\rho_{aX+b, cY+d} = \rho_{XY}$ pour tout a, b, c, d réels avec $ac \neq 0$

Stationnarité

La stationnarité d'un processus aléatoire décrit comment ses propriétés statistiques restent constantes au fil du temps. Pour qu'une série temporelle soit stationnaire, elle doit présenter quatre propriétés constantes dans le temps :

1. Moyenne constante
2. Variance constante
3. Structure d'autocorrélation constante
4. Aucun composant périodique (saisonnalité)

L'autocorrélation signifie que la mesure actuelle de la série temporelle est corrélée à une mesure passée.

Stationnarité au sens strict

Un processus $\{X_i\}$ est strictement stationnaire si :

- La distribution de X_i est la même que celle de X_j pour $i \neq j$.
- Les implications sont que $\mathrm{E}[X_i]$ et $\mathrm{Var}(X_i)$ sont constants et ne dépendent pas de i.
- La distribution conjointe de (X_i, X_j) est la même que celle de (X_{i+k}, X_{j+k}).
- L'autocovariance $\mathrm{Cov}(X_i, X_{i+j})$ ne dépend que de j.

Stationnarité au sens large

Un processus $\{X_i\}$ est largement stationnaire si :

- $\mathrm{E}[X_i]$ et $\mathrm{Var}(X_i)$ sont constants et ne dépendent pas de i.
- L'autocovariance $\mathrm{Cov}(X_i, X_{i+j}) = \mathrm{E}[X_i X_{i+j}] - \mathrm{E}[X_i]\mathrm{E}[X_{i+j}]$ ne dépend que de j.

Loi de probabilité

Loi	Param	$\mathrm{E}[X]$	$\mathrm{Var}(X)$	Support
Bernoulli	p	p	$p(1-p)$	$\{0,1\}$
Binomiale	n,p	np	$np(1-p)$	$\{0,\ldots,n\}$
Uniforme	a,b	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$[a,b]$
Normale	μ,σ	μ	σ^2	$\mathbb{R}$

Likelihood

Soit X une variable aléatoire à valeurs dans $\mathbb{R}$, on définit la likelihood de θ par :

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

où $f(x)$ est la densité de probabilité de X. Souvent, on utilise le logarithme de la likelihood :

$$\log L(\theta) = \sum_{i=1}^{n} \log f(x_i|\theta)$$

MLE

$\hat{\theta}_{MLE}$ est de θ est le paramètre qui maximise la likelihood de θ.
Soit θ un paramètre d'une loi de probabilité, on définit le MLE de θ par :

$$\hat{\theta} = \arg\max_{\theta} L(\theta)$$

où $L(\theta)$ est la likelihood de θ.

Procédure

1. Définir la likelihood de θ $L(\theta)$.
2. On passe au logarithme de la likelihood. $l(\theta) = \log L(\theta)$
3. On dérive la log-likelihood par rapport à θ
4. On cherche $\hat{\theta}$ tel que $\frac{\partial}{\partial\theta} l(\hat{\theta}) = 0$.
5. On vérifie que $\frac{\partial^2}{\partial\theta^2} l(\hat{\theta}) < 0$

Régression Linéaire

Soit un tableau de données :
x = Soap(g) , y = Height(cm) , $x \cdot y$, x^2

$$X = [1, Soap]$$

$$X^T X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} = \begin{bmatrix} 7 & 38.5 \\ 38.5 & 218.95 \end{bmatrix}$$

$$X^T y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} 348 \\ 1975 \end{bmatrix}$$

$$\hat{\theta} = (X^T X)^{-1} X^T y = \begin{bmatrix} -2.67 \\ 9.51 \end{bmatrix}$$

Inverse d'une matrice 2x2 :

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Normale

1. $P(X \leq x) = \Phi(x) = \int_{-\infty}^{x} N(0,1) dx$
2. $\Phi(x) = 1 - \Phi(-x)$
3. $F_x(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$ est une loi normale centrée réduite

TOPIC