

Biais de la moyenne empirique et de la variance

Cette question est vraiment trop facile... Par linéarité de l'espérance, on a :

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

Les X_i sont *i.i.d.* donc $\mathbb{E}[X_i] = m$ pour tout i et

$$\mathbb{E}[\bar{X}] = \frac{1}{n} \times n \times m = m \quad (2)$$

On en déduit que $\bar{X}$ est toujours un estimateur sans biais de la moyenne (quand elle existe).

De la même façon,

$$\mathbb{E}[nS^2] = \mathbb{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \quad (3)$$

$$= \mathbb{E}\left[\sum_{i=1}^n (X_i - m + m - \bar{X})^2\right] \quad (4)$$

$$= \mathbb{E}\left[\sum_{i=1}^n \left((X_i - m)^2 + 2(X_i - m)(m - \bar{X}) + (m - \bar{X})^2\right)\right] \quad (5)$$

$$= n\sigma^2 + \mathbb{E}\left[2n(m - \bar{X})(\bar{X} - m) + n(m - \bar{X})^2\right] \quad (6)$$

$$= n\sigma^2 - n\mathbb{V}(\bar{X}) \quad (7)$$

$$= n\sigma^2 - \sigma^2 \quad (8)$$

$$= (n-1)\sigma^2 \quad (9)$$

Ainsi,

$$\mathbb{E}[S^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2 \quad (10)$$

La variance empirique est donc biaisée. Par contre,

$$\mathbb{E}[\tilde{S}^2] = \sigma^2 \quad (11)$$

si l'on pose

$$\tilde{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (12)$$

Estimateur du maximum de vraisemblance pour un échantillon gaussien

Partant d'un n -échantillon gaussien $X = (X_1, \dots, X_n)$ de loi $\mathcal{N}(m, \sigma^2)$, on souhaite estimer $\theta = (m, \sigma^2)$. Dans toute la suite, on posera $v = \sigma^2$. La log-vraisemblance de cet échantillon est

$$l(x, \theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) \quad (13)$$

$$= -\frac{1}{2v} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2} \ln v - \frac{n}{2} \ln(2\pi) \quad (14)$$

Le gradient de cette fonction de deux variables en $\theta = (m, v)$ est le vecteur colonne

$$\nabla_l(\theta) = \left(\frac{1}{v} \sum_{i=1}^n (x_i - m); \frac{1}{2v^2} \sum_{i=1}^n (x_i - m)^2 - \frac{n}{2v} \right)^T \quad (15)$$

Le gradient s'annule en un unique point critique $\hat{\theta}$ dont les coordonnées sont données par

$$m = \bar{x} \text{ et } v = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2 \quad (16)$$

La matrice hessienne, en θ , de la fonction l est donnée par

$$H_l(\theta) = \begin{pmatrix} \frac{\partial^2}{\partial m^2} l(x, \theta) & \frac{\partial^2}{\partial m \partial v} l(x, \theta) \\ \frac{\partial^2}{\partial v \partial m} l(x, \theta) & \frac{\partial^2}{\partial v^2} l(x, \theta) \end{pmatrix} \quad (17)$$

avec

$$\frac{\partial^2}{\partial m^2} l(x, \theta) = -\frac{n}{v} \quad (18)$$

$$\frac{\partial^2}{\partial v^2} l(x, \theta) = -\frac{1}{v^3} \sum_{i=1}^n (x_i - m)^2 + \frac{n}{2v^2} \quad (19)$$

$$\frac{\partial^2}{\partial v \partial m} l(x, \theta) = \frac{\partial^2}{\partial m \partial v} l(x, \theta) \quad (20)$$

$$= -\frac{1}{v^2} \sum_{i=1}^n (x_i - m) \quad (21)$$

Ces expressions s'entendent à x fixé, x étant une observation du n -échantillon.

Si nous évaluons maintenant cette matrice hessienne au point critique $\hat{\theta} = (\bar{x}, s^2)$ où $\bar{x}$ et s^2 sont respectivement la moyenne et la variance empirique de l'observation, alors

$$\frac{\partial^2}{\partial m^2} l(x, \hat{\theta}) = -\frac{n}{s^2} \quad (22)$$

$$\frac{\partial^2}{\partial v^2} l(x, \hat{\theta}) = -\frac{1}{s^6} \sum_{i=1}^n (x_i - \bar{x})^2 + \frac{n}{2s^4} \quad (23)$$

$$= -\frac{ns^2}{s^6} + \frac{n}{2s^4} = -\frac{n}{2s^4} \quad (24)$$

$$\frac{\partial^2}{\partial v \partial m} l(x, \theta) = -\frac{1}{v^2} \sum_{i=1}^n (x_i - m) \quad (25)$$

$$= -\frac{1}{s^2} \sum_{i=1}^n (x_i - \bar{x}) = 0 \quad (26)$$

Finalement,

$$H_l(\hat{\theta}) = \begin{pmatrix} -\frac{n}{s^2} & 0 \\ 0 & -\frac{n}{2s^4} \end{pmatrix} \quad (27)$$

Cette matrice est clairement définie négative et par conséquent $\hat{\theta}$ est l'unique maximum de vraisemblance. L'estimateur du maximum de vraisemblance est donc

$$\hat{\theta}_{MV} = (\bar{X}, S^2) \quad (28)$$

Estimateur du maximum de vraisemblance pour la loi uniforme sur $[\theta, \theta + 1]$

La vraisemblance d'un échantillon de cette loi s'écrit

$$L(x, \theta) = \mathbb{1}_{[\theta \leq x_{(1)} \leq x_{(n)} \leq \theta + 1]}(x) \quad (29)$$

Cette fonction (de θ) vaut 1 ou 0. Pour une observation fixée x de l'échantillon, elle est nulle si $\theta > x_{(1)}$ ou bien si $\theta < x_{(n)} - 1$. Elle est constante et vaut 1 dans l'intervalle $[x_{(n)} - 1, x_{(1)}]$. Ainsi, toute valeur de θ dans cet intervalle réalise le maximum de la vraisemblance:

Tout $\hat{\theta}$ choisit dans l'intervalle (aléatoire) $[X_{(n)} - 1, X_{(1)}]$ est un estimateur du maximum de vraisemblance.

Estimateur gaussien de la moyenne *a posteriori*

À la fin du chapitre 1, nous avons trouvé l'expression de la loi *a posteriori* d'un échantillon gaussien dont la moyenne inconnue suit une loi *a priori* gaussienne, d'hyperparamètres a (pour sa moyenne) et b^2 (pour sa variance). La loi *a posteriori* est gaussienne, de moyenne m et de variance τ^2 avec :

$$\begin{cases} m &= \bar{x} \frac{b^2}{b^2 + \sigma^2/n} + a \frac{\sigma^2/n}{b^2 + \sigma^2/n} \\ \tau^2 &= \frac{b^2 \sigma^2/n}{b^2 + \sigma^2/n} \end{cases} \quad (30)$$

m est une combinaison linéaire convexe de la moyenne de l'échantillon et de la moyenne *a priori*. Les coefficients de cette combinaison linéaire sont proportionnels aux variances de l'échantillon et variance *a priori*.

L'estimateur de la moyenne *a posteriori* est donc $\hat{\theta}$ défini par :

$$\hat{\theta} = \bar{X} \frac{b^2}{b^2 + \frac{\sigma^2}{n}} + a \left(1 - \frac{b^2}{b^2 + \frac{\sigma^2}{n}} \right) \quad (31)$$

Estimateur du coefficient de corrélation linéaire empirique

$(X_1, Y_1), \dots, (X_n, Y_n)$ est un n -échantillon d'un vecteur gaussien (X, Y) dont le coefficient de corrélation linéaire est noté ρ . Le coefficient de corrélation empirique est par définition égal à

$$\hat{\rho}_n = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \times \sum_{i=1}^n (Y_i - \bar{Y})^2 \right]^{1/2}} \quad (32)$$

Montrons que $\sqrt{n}(\hat{\rho}_n - \rho)$ est asymptotiquement normale et que

$$\sqrt{n}(\hat{\rho}_n - \rho) \rightsquigarrow \mathcal{N}(0, (1 - \rho^2)^2) \quad (33)$$

La méthode de stabilisation de la variance amène à chercher une fonction différentiable ψ dont la dérivée vaut $(1 - \rho^2)^{-1}$. On pose donc naturellement $\psi(\rho) = \operatorname{arctanh} \rho$ et une nouvelle application de la méthode delta montre que

$$\sqrt{n}(\psi(\hat{\rho}_n) - \psi(\rho)) \rightsquigarrow \mathcal{N}(0, 1) \quad (34)$$