UNIVERSITÉ JEAN MONNET
LICENCE 3 INFORMATIQUE

KERDONCUFF Tanguy
SEBAN Marc

Correction du TD6 de probabilités

Table des matières

1 TD6 : Estimation - Biais - Erreur d'estimation

1.1 Exercice I

Supposons que $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta, Var(\hat{\theta}_1) = \sigma_1^2 et Var(\hat{\theta}_2) = \sigma_2^2$. Soit a une constante.

1. Montrer que $\hat{\theta}_3 = a\hat{\theta}_1 + (1-a)\hat{\theta}_2$ est un estimateur non biaisé de θ.

   Solution: On doit montrer que $E(\hat{\theta}_3) = \theta$

$$\begin{aligned} E(\hat{\theta}_3) =& aE(\hat{\theta}_1) + (1-a)E(\hat{\theta}_2) \\ =& a\theta + (1-a)\theta \\ =& (a+1-a)\theta \\ =& \theta \end{aligned}$$

   On a donc un estimateur non biaisé.

2. Quelle valeur de a choisir pour minimiser la variance de $\hat{\theta}_3$? (On supposera $\hat{\theta}_1$ et $\hat{\theta}_2$ indépendant. indépendants).

   Solution: On rappelle que

$$V(aX + bY) = V(aX) + V(bY) - 2ab \times cov(X, Y)$$

$$\begin{aligned} V(\hat{\theta}_3) =& V(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) \\ =& a^2V(\hat{\theta}_1) + (1-a)^2V(\hat{\theta}_2) + 2a(1-a)cov(\hat{\theta}_1, \hat{\theta}_2) \\ =& a^2V(\hat{\theta}_1) + (1-a)^2V(\hat{\theta}_2) \text{ car on a indépendance} \\ =& a^2(\sigma_1^2 + \sigma_2^2) - 2a\sigma_2^2 + a^2\sigma_2^2 \\ =& a^2(\sigma_1^2 + \sigma_2^2) - 2a\sigma_2^2 + \sigma_2^2 \end{aligned}$$

   On cherche le minimum de cette fonction en part rapport à a.

$$\frac{dV(\hat{\theta}_3)}{da} = 0 \Rightarrow 2(\sigma_1^2 + \sigma_2^2)a - 2\sigma_2^2 = 0 \Rightarrow a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

   On a trouver un bon estimateur, il est sans biais et a une variance assez faible.

1.2 Exercice II

Même question qu'à l'exerice I, mais cette fois-ci, $cov(\hat{\theta}_1, \hat{\theta}_2) = c \neq 0$

1. Quelle valeur de a choisir pour minimiser la variance de $\hat{\theta}_3$?

Solution:

$$\begin{aligned}
V(\hat{\theta}_3) =& V(a\hat{\theta}_1 + (1-a)\hat{\theta}_2)\\
=& a^2V(\hat{\theta}_1) + (1-a)^2V(\hat{\theta}_2) + 2a(1-a)cov(\hat{\theta}_1, \hat{\theta}_2)\\
=& a^2(\sigma_1^2 + \sigma_2^2) - 2a\sigma_2^2 + \sigma_2^2 + 2a(1-a)c\\
=& a^2(\sigma_1^2 + \sigma_2^2) - 2a\sigma_2^2 + \sigma_2^2 - 2a^2c + 2ac\\
=& a^2(\sigma_1^2 + \sigma_2^2 - 2c) + (2c - 2\sigma_2^2)a + \sigma_2^2
\end{aligned}$$

On cherche le minimum de cette fonction en part rapport à a.

$$\frac{dV(\hat{\theta}_3)}{da} = 0 \Rightarrow 2(\sigma_1^2 + \sigma_2^2 - 2c)a - 2\sigma_2^2 + 2c = 0 \Rightarrow a = \frac{\sigma_2^2 - c}{\sigma_1^2 + \sigma_2^2 - 2c}$$

On a trouver un bon estimateur, il est sans biais et a une variance assez faible.

1.3 Exercice III

Soit Y_1, Y_2, Y_3 , un échantillon aléatoire d'une distribution exponentielle de densité

$$\text{Pour tout } y > 0 \quad f(y) = \frac{1}{\theta}e^{-\frac{y}{\theta}}$$

On considère deux estimateur de θ : $\hat{\theta}_1 = Y_1$ et $\hat{\theta}_2 = min(Y_1, Y_2, Y_3)$.

1. Trouver si ces 2 estimateurs sont sans biais.

Solution:

$$E(\hat{\theta}_1) = \int_0^\infty yf(y)dy = \int_0^\infty y\frac{1}{\theta}e^{-\frac{y}{\theta}}dy$$

On va intégrer par parti avec pour tout $y \quad u(y) = y$ et $v'(y) = \frac{1}{\theta}e^{-\frac{y}{\theta}}$ donc on a $u'(y) = 1$ et $v(y) = -e^{-\frac{y}{\theta}}$

$$\begin{aligned}
E(\hat{\theta}_1) =& \int_0^\infty u(y)v'(y)dy\\
=& [u(y)v(y)]_0^\infty - \int_0^\infty u'(y)v(y)dy\\
=& [y \times (-e^{-\frac{y}{\theta}})]_0^\infty - \int_0^\infty 1 \times (-e^{-\frac{y}{\theta}})dy\\
=& 0 - 0 + \int_0^\infty e^{-\frac{y}{\theta}}dy\\
=& [-\theta e^{-\frac{y}{\theta}}]_0^\infty\\
=& -\theta(0 - e^0)\\
=& \theta
\end{aligned}$$

Donc non biaisé.
On pose F_{min} la fonction de répartition associé à la variable aléatoire $\hat{\theta}_2$.

$$F_{min}(y) = P(Y_{min} \leq y) = 1 - P(Y_{min} > y) = 1 - P(y < Y_1, y < Y_2, y < Y_3) = 1 - (1 - F(y))^3$$

Or on a

$$F(y) = \int_0^y \frac{1}{\theta} e^{\frac{-t}{\theta}} dt = [-e^{-\frac{t}{\theta}}]_0^y = 1 - e^{\frac{-y}{\theta}}$$

Donc on a

$$F_{min}(y) = 1 - (1 - F(y))^3 = 1 - e^{-\frac{3y}{\theta}}$$

Et en dérivant on a immédiatement

$$f_{min} = \frac{3}{\theta} e^{-\frac{3y}{\theta}}$$

$$\begin{aligned}
E(\hat{\theta}_2) = & \int_0^\infty \frac{3}{\theta} y e^{-\frac{3y}{\theta}} dy \\
=&[y \times (-e^{-\frac{3y}{\theta}})]_0^\infty - \int_0^\infty 1 \times (-e^{-\frac{3y}{\theta}}) dy \\
=&0 - 0 + \int_0^\infty e^{-\frac{3y}{\theta}} dy \\
=&[-\frac{\theta}{3} e^{-\frac{3y}{\theta}}]_0^\infty \\
=&-\frac{\theta}{3}(0 - e^0) \\
=&\frac{\theta}{3}
\end{aligned}$$

Donc on a un estimateur biaisé.

1.4 Exercice IV

On étudie l'influence de l'ordre de naissance chez des bacheliers d'environnement socio-économiques équivalent. Sur un échantillon de 180 bacheliers, 126 sont enfants uniques ou aînés. Sur un échantillon de 100 non-bacheliers, 54 sont enfants uniques ou aînés.

1. Estimer la différence de proportion entre les deux populations, et trouver une borne sur l'erreur d'estimation de cette différence avec une probabilité de 0.95.

   Solution: Notons p_1 et p_2 la proportion des deux populations et $\theta = p_1 - p_2$. On cherche un estimateur de θ donc on pose

$$\hat{\theta} = \hat{p_1} - \hat{p_2} = \frac{Y_1}{n_1} - \frac{Y_2}{n_2} = \frac{126}{180} - \frac{54}{100} = 0.7 - 0.54 = 0.16$$

On cherche maintenant à contrôler l'erreur d'une telle estimation $\epsilon = |\hat{\theta} - \theta|$. On cherche $P(\epsilon < x) = 0.95$.
On va essayer de transformer $\hat{\theta}$ en une variable suivant une loi normale centré réduite.

$$V(\hat{\theta}) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = \frac{0.7 \times 0.3}{180} + \frac{0.54 \times 0.46}{100} = 0.00365$$

Donc $\sigma_{\hat{\theta}} = 0.0604$. $Z = \frac{\hat{\theta}-\theta}{\sigma}$ suit bien une loi normale : $\hat{\theta}$ est une soustraction de deux loi normale indépendante. Et chacun des deux estimateurs p suit bien une loi normale comme somme de Bernoulli indépendant et de même paramètres (+ loi normal car > 30 expérience). Cette variable est ensuite centré et réduite (θ est sa moyenne).

$$\begin{aligned}
P(|\hat{\theta} - \theta| < x) = 0.95 &\Rightarrow P(-\frac{x}{\sigma} < \frac{\hat{\theta} - \theta}{\sigma} < \frac{x}{\sigma}) = 0.95 \\
&\Rightarrow P(-\frac{x}{\sigma} < Z < \frac{x}{\sigma}) = 0.95 \\
&\Rightarrow 1 - 2P(Z > \frac{x}{\sigma}) = 0.95 \\
&\Rightarrow P(Z > \frac{x}{\sigma}) = 0.025 \\
&\Rightarrow P(Z < \frac{x}{\sigma}) = 0.975 \\
&\Rightarrow a = 1.96 \times \sigma \\
&\Rightarrow a = 0.12
\end{aligned}$$

Donc $P(|\hat{\theta} - \theta| < 0.12)$