Maximum de vraisemblance

Estimation des paramètres d'une loi normale

Soit la loi normale univariée d'epérance μ et de variance σ^2

$$f(x;\mu,\sigma^2) = \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \tag{1}$$

étant donné un échantillon d'observations *i.i.d.* $(x_1,\ldots,x_n)$ où $X_i \sim \mathcal{N}(x_i;\mu,\sigma^2)$

1. calculer l'estimateur du maximum de vraisemblance (MV) pour μ
2. vérifier que l'estimateur est sans biais et à variance minimale ; en déduire sur son efficacité
3. calculer l'estimateur du MV pour σ^2
4. Est-il sans biais ? sinon, proposez une correction du biais pour fournir un estimateur sans biais

Solution

The likelehood to be maximized is given as the joint probability density function for sample $(x_1,\ldots,x_n)$ of n independent identically distributed normal random variables

$$f(x_1,\ldots,x_n;\mu,\sigma^2) = \prod_{i=1}^{n} f(x_i;\mu,\sigma^2) = \prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}. \tag{2}$$

Maximizing this likelihood is equivalent to maximzing the following log-likehood function

$$\begin{aligned}
\mathcal{L}(\mu,\sigma^2) &= \log f(x_1,\ldots,x_n;\mu,\sigma^2) = \log\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}\\
&= \sum_{i=1}^{n}\log\frac{1}{\sigma\sqrt{2\,\pi}} + \sum_{i=1}^{n}\log\mathrm{e}^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}\\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\,. \qquad (3)
\end{aligned}$$

Lets first start by maximzing (3) with respect to μ. We have to set its partial derivative w.r.t μ equat to zero. By taking into account the fact that in (3), only the quantity $-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$ depends on μ, we can therefore write

$$\frac{\partial\mathcal{L}(\mu,\sigma^2)}{\partial\mu} = \frac{\partial - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}{\partial\mu} = 0, \tag{4}$$

that is

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}2\,(x_i-\mu)\,\frac{\partial(x_i-\mu)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0. \tag{5}$$

Finally we have $\sum_{i=1}^{n}(x_i-\mu) = \sum_{i=1}^{n}x_i - n\mu = 0$ which results in the ML estimate $\hat{\mu}$ of μ

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i,$$

that is the sample mean !.

We can see that the ML estimator of μ given by

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n}X_i$$

is unbiased. Indeed, we have

$$\mathbb{E}[\widehat{\mu}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

where we have used the fact that $\mathbb{E}[X_i] = \mu$ which is the true mean.

To estimate σ^2, we similarly differentiate the log-likelihood (3) with respect to σ and equate to zero. Since in (3) only the quantity $-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$ depends on σ, we have

$$\frac{\partial\mathcal{L}(\mu,\sigma^2)}{\partial\sigma} = \frac{\partial\left(-\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)}{\partial\sigma} = 0, \tag{6}$$

that is

$$-\frac{n}{2\sigma^2} - \frac{-1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \tag{7}$$

whiche gives

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = n. \tag{8}$$

Finally we therefore have

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{9}$$

and by replacing μ by its ML estimate we get the ML estimate $\hat{\sigma}^2$ for σ^2

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 \tag{10}$$

which is the empirical variance!

However, in contrast to the ML estimator for the mean which is unbiased, the one for the variance, as we can see it is biased.

To calculate the expectation of the ML estimator $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\widehat{\mu})^2$, by using the fact that

$$\left(\sum_{i=1}^{n}X_i\right)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j$$

we first write

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\widehat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 + \left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)^2 - \frac{2}{n}X_i\sum_{j=1}X_j\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{h=1}^{n}X_jX_h - \frac{2}{n}\sum_{j=1}X_iX_j\right),\end{aligned} \tag{11}$$

the expectation of $\widehat{\sigma}^2$ is then given by

$$\mathbb{E}[\widehat{\sigma}^2] = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{E}[X_i^2] + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{h=1}^{n}\mathbb{E}[X_jX_h] - \frac{2}{n}\sum_{j=1}\mathbb{E}[X_iX_j]\right).$$

Since the variables are mutually independent, we have, for $i \neq j$, $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i][X_j] = \mu^2$. When $i = j$, we have $\mathbb{E}[X_i X_i] = \mathbb{E}[X_i^2] = \sigma^2 + \mu^2$ from the variance formula. The previous equation is therefore wreritten as

$$\begin{aligned}
\mathbb{E}[\widehat{\sigma}^2] &= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \mu^2 + \frac{1}{n^2}\left(n\left((n-1)\,\mu^2 + \left(\sigma^2+\mu^2\right)\right)\right) - \frac{2}{n}\left((n-1)\,\mu^2 + \left(\sigma^2+\mu^2\right)\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \mu^2 + \frac{1}{n^2}\left(n^2\mu^2 + n\sigma^2\right) - \frac{2}{n}\left(n\mu^2 + \sigma^2\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \mu^2 + \mu^2 + \frac{1}{n}\sigma^2 - 2\mu^2 - \frac{2}{n}\sigma^2\right) \\
&= \frac{1}{n}(n\sigma^2 - \sigma^2) \\
&= \frac{n-1}{n}\sigma^2. \qquad (12)
\end{aligned}$$

The ML estimator $\mathbb{E}[\widehat{\sigma}^2]$ for σ^2 is therefore unbiased since $\mathbb{E}[\widehat{\sigma}^2] \neq \sigma^2$. However it is asymptotically unbiased : $\lim_{n\to\infty} \mathbb{E}[\widehat{\sigma}^2] = \sigma^2$.

We can correct the bias by taking $\sigma^{\star 2} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)^2$ as an unbiased estimator of the variance, rather than the ML estimator $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)^2$. It is indeed easy to check that $\mathbb{E}[\sigma^{\star 2}] = \sigma^2$.

Estimation des paramètres d'une loi de Poisson

La loi de Poisson de paramètre λ pour $x \in \mathbb{N}$ est donnée par

$$p(x;\lambda) = P(X = x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad (13)$$

Étant donné un échantillon d'observations *i.i.d.* $(x_1, \ldots, x_n)$ généré suivant une loi de Poisson de paramètre inconnu λ, $p(x;\lambda)$, calculer l'EMV du paramètre λ

Solution

La vraisemblance s'écrit :

$$\begin{aligned}
L(\lambda; x_1, \ldots, x_n) &= p(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^{n} p(x_i;\lambda) &(14)\\
&= \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} &(15)\\
&= e^{-\lambda n}\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!} &(16)
\end{aligned}$$

La log-vraisemblance est donc donnée par

$$\begin{aligned}
\ln L(\lambda; x_1, \ldots, x_n) &= \ln e^{-\lambda n}\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!} &(17)\\
&= \ln e^{-\lambda n} + \ln\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!} &(18)\\
&= -\lambda n + \sum_{i=1}^{n}\ln\frac{\lambda^{x_i}}{x_i!} &(19)\\
&= -\lambda n + \ln\lambda\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\ln(x_i!) &(20)\\
& &(21)
\end{aligned}$$

La dérivée première s'annule quand :

$$\frac{\partial \ln L(\lambda; x_1, \dots, x_n)}{\partial \lambda} = 0 \tag{22}$$
$$-n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0 \tag{23}$$

donc

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$$

La dérivée seconde s'écrit :

$$\frac{\partial^2 \ln L(\lambda; x_1, \dots, x_n)}{\partial \lambda^2} = -\frac{\sum_{i=1}^n x_i}{\lambda^2} \leq 0$$

car $x_i \in \mathbb{N}$. L'estimateur est donné par :

$$\Lambda_n = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

Remarque : on peut remarquer que l'estimateur s'exprime à l'aide de la statistique exhaustive pour λ : $\Lambda = \frac{T}{n}$

1 Espérance et Variance de l'estimateur des moindres carrés (EMC)

1. Calculer l'espérance de l'EMC de $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \cdot \tag{24}$$

du modèle linéaire

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \cdot \tag{25}$$

où le bruit $\mathbf{e}$ est centré et de matrice de covariance $\sigma^2\mathbf{I}$

2. En remarquant que

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e} \tag{26}$$

calculer la matrice de covariance de l'EMC $\hat{\boldsymbol{\beta}}$

Solution On a

$$\begin{aligned}
\mathbb{E}[\hat{\boldsymbol{\beta}}] &= \mathbb{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}[\mathbf{y}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}[\mathbf{X}\boldsymbol{\beta} + \mathbf{e}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned} \tag{27}$$

On a également

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[\mathbf{X}\boldsymbol{\beta} + \mathbf{e}] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e} \\
&= \boldsymbol{\beta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}
\end{aligned} \tag{28}$$

donc

$$\begin{aligned}
cov[\hat{\boldsymbol{\beta}}] &= \mathbb{E}\left[(\hat{\boldsymbol{\beta}} - \mathbb{E}[\hat{\boldsymbol{\beta}}])(\hat{\boldsymbol{\beta}} - \mathbb{E}[\hat{\boldsymbol{\beta}}])^T\right] \\
&= \mathbb{E}\left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\right] \\
&= \mathbb{E}\left[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e})((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e})^T\right] \\
&= \mathbb{E}\left[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}^T\mathbf{e}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\right] \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}\left[\mathbf{e}^T\mathbf{e}\right]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[\sigma^2\mathbf{I}]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \qquad (29)
\end{aligned}$$