Examples of sufficient statistics and estimation

Marianna Bolla, Prof, DSc.
Institute of Mathematics, BME

December 19, 2020

Poisson distribution

Let $X_1, \dots, X_n$ be i.i.d. sample from Poisson distribution with parameter λ.

$$L_\lambda(\mathbf{x}) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \left(\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}\right) \cdot \left(\prod_{i=1}^n \frac{1}{x_i!}\right) = g_\lambda(\sum_{i=1}^n x_i) \cdot h(\mathbf{x}),$$

so $\sum_{i=1}^n X_i$ is sufficient statistic for λ, akin to its one-to-one function $\bar{X}$.

To find the MLE,

$$\ln L_\lambda(\mathbf{x}) = \ln \left[\prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right] = \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i! - \lambda n.$$

Differentiating with respect to λ, the likelihood equation is

$$\frac{\partial \ln L_\lambda(\mathbf{x})}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0.$$

The solution is $\hat{\lambda} = \bar{x}$, which indeed gives a local and global maximum. So $T(\mathbf{X}) = \bar{X}$ is the MLE of λ, provided it is not 0, i.e., not all the sample entries are zero at the same time (it can happen with positive, albeit 'small' probability).

This distribution belongs to the exponential family, as

$$p_\lambda = \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} e^{x \ln \lambda} \frac{1}{x!}, \quad x = 0, 1, 2, \dots,$$

where $\ln \lambda$ is the canonical parameter and $\sum_{i=1}^n X_i$ is the canonical sufficient statistic. It is also complete, so making it unbiased, $\bar{X}$ will be the efficient, unbiased estimator for λ, based on the Rao–Blackwell theorem. It also implies, that starting with any other unbiased estimator, e.g., with $\sum_{i=1}^n a_i X_i$, where $\sum_{i=1}^n a_i = 1$, and 'blackwellizing' it with $\sum_{i=1}^n X_i$, we always get $\bar{X}$.

$I_1(\lambda) = \frac{1}{\lambda}$, $I_n(\lambda) = \frac{n}{\lambda}$, so the information bound for λ is $\frac{1}{I_n(\lambda)} = \frac{\lambda}{n} = \text{Var}\,(\bar{X})$, and it is attained by $\bar{X}$. Therefore, $\bar{X}$ is the efficient estimator of λ, for this reason too.

Exponential distribution

Let $X_1, \ldots, X_n$ be i.i.d. sample from exponential distribution with parameter λ. Then

$$L_\lambda(\mathbf{x}) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i},$$

that is $g_\lambda(T(\mathbf{x}))$, and $h(\mathbf{x}) = 1 \cdot I_{(0,\infty)}$. Therefore, $\sum_{i=1}^n X_i$ is sufficient akin to $\bar{X}$ or $\frac{1}{\bar{X}}$. It is also complete, so making it unbiased, $\frac{n-1}{n}\frac{1}{\bar{X}}$ will be the efficient, unbiased estimator for λ, based on the Rao–Blackwell theorem. However, it can be shown that it does not reach the information bound. Consequently, no unbiased estimator reaches the bound for λ. What is this bound?

$$\mathbb{E}_\lambda\left(\frac{\partial}{\partial\lambda}(\ln f_\lambda(X)\right) = \mathbb{E}_\lambda\left(\frac{1}{\lambda} - X\right) = 0,$$

therefore

$$I_1(\lambda) = \text{Var}_\lambda\left(\frac{1}{\lambda} - X\right) = \text{Var}_\lambda(X) = \frac{1}{\lambda^2}.$$

So

$$I_n(\lambda) = \frac{n}{\lambda^2}.$$

If $\psi(\lambda) = \lambda$, then the information bound is $\frac{1}{I_n(\lambda)} = \frac{\lambda^2}{n}$, whereas

$$\text{Var}_\lambda\left(\frac{n-1}{n}\frac{1}{\bar{X}}\right) = \frac{\lambda^2}{n-2}$$

which is larger than the bound (though, asymptotically approaches it as $n \to \infty$).

However, if we consider the parameter function $\psi(\lambda) = \frac{1}{\lambda}$, this is the expectation of the underlying distribution. For it, $\bar{X}$ is an unbiased estimator and reaches the bound, which is

$$\frac{\psi'(\lambda)^2}{I_n(\lambda)} = \frac{[-\frac{1}{\lambda^2}]^2}{\frac{n}{\lambda^2}} = \frac{1}{n\lambda^2}.$$

Indeed, $\text{Var}(\bar{X}) = \frac{1/\lambda^2}{n}$ that is the same as the above bound. So $\bar{X}$ is efficient estimator of $\frac{1}{\lambda}$.

As for the MLE of λ,

$$\ln L_\lambda(\mathbf{x}) = \ln \left[\prod_{i=1}^{n} \lambda e^{-\lambda x_i} \right] = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i,$$

from which, after differentiating, we get that $\hat{\lambda} = 1/\bar{x}$, that gives a local and global maximum. Consequently, $T(\mathbf{X}) = 1/\bar{X}$ is the MLE of λ with probability 1 ($\bar{X}$ can be 0 only with probability 0).

Gaussian distribution

Let $X_1, \ldots, X_n$ be i.i.d. sample from normal (Gaussian) distribution with unknown parameter $\theta = (\mu, \sigma^2)$. Then

$$L_\theta(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right) =$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2 \right] \right).$$

It is $g_\theta(T(\mathbf{x}))$, where $T(\mathbf{X}) = (\bar{X}, S^2)$ sufficient for θ, and $h(\mathbf{x}) = 1$. Obviously, $(\bar{X}, S^{*2})$ or $(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ are also sufficient. It also belongs to the exponential family.

To find MLE,

$$\ln L_\theta(\mathbf{x}) = \ln \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \sum_{i=1}^{n} \left[-\ln(\sqrt{2\pi\sigma^2}) - \frac{(x_i - \mu)^2}{2\sigma^2} \right] =$$

$$= -\frac{n}{2}(\ln(2\pi) + \ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Taking partial derivatives,

$$\frac{\partial \ln L_\theta(\mathbf{x})}{\partial \mu} = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(x_i - \mu)(-1) = 0 \Longrightarrow \hat{\mu} = \bar{x}$$

and

$$\frac{\partial \ln L_\theta(\mathbf{x})}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 = 0.$$

Since the solution $\hat{\mu} = \bar{x}$ does not depend on the actual value of σ^2, substituting it to the second equation, we get that $\hat{\sigma^2} = S_n^2$, that is only asymptotically unbiased for σ^2. The Hessian at $(\bar{x}, s_n^2)$ is:

$$H = \begin{pmatrix} -\frac{n}{s_n^2} & 0 \\ & \\ 0 & -\frac{n}{2(s_n^2)^2} \end{pmatrix},$$

which is negative definite, so we indeed have a local and global maximum here.

If both parameters are unknown, then the information bound cannot be attained for θ. However, if the variance is known (σ_0^2), then the bound can be attained for μ. Indeed,

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-\mu)^2}{2\sigma_0^2}}.$$

Therefore,

$$I_1(\mu) = \operatorname{Var}_\mu^2\left(\frac{\partial}{\partial\mu}\ln f_\mu(X)\right) = \operatorname{Var}_\mu^2\left(\frac{2(X-\mu)}{2\sigma_0^2}\right) = \frac{1}{\sigma_0^4}\operatorname{Var}_\mu^2(X-\mu) = \frac{1}{\sigma_0^4}\sigma_0^2 = \frac{1}{\sigma_0^2} \neq 0.$$

So

$$I_n(\mu) = n\frac{1}{\sigma_0^2},$$

and it is $\frac{1}{\operatorname{Var}_\mu^2(\bar{X})}$.

The moment estimation of the parameters follows. The theoretical first two moments are:

$$m_1 = \mu, \quad m_2 = \sigma^2 + \mu^2.$$

It can easily be checked that the Jacobian of the $(\mu, \sigma^2) \to (m_1, m_2)$ map is not zero, and the inverse map is

$$\mu = m_1, \qquad \sigma^2 = m_2 - m_1^2.$$

Eventually, by

$$\hat{m}_1 = \bar{X}, \qquad \hat{m}_2 - \hat{m}_1^2 = S_n^2,$$

the moment estimates of the parameters are

$$\hat{\mu} = \bar{X}, \qquad \hat{\sigma}^2 = S_n^2,$$

that are the same as the ML estimetes.

Continuous uniform distribution

Let $X_1, \dots, X_n$ be i.i.d. sample from continuous uniform distribution on $[a, b]$. Here $\theta = (a, b)$.

$$L_\theta(\mathbf{x}) = \prod_{i=1}^{n} f_\theta(x_i) = \frac{1}{(b-a)^n}, \quad \text{if} \quad x_1, \dots, x_n \in [a, b],$$

and 0, otherwise. $L_\theta(\mathbf{x}) = (b-a)^{-n} I(x_1^* \geq a, \, x_n^* \leq b) = g_\theta(x_1^*, x_n^*)$ and $h(\mathbf{x}) = 1$. So the pair (X_1^*, X_n^*) is sufficient for (a, b). It also gives the MLE, as we maximize the likelihood on the constraint that $[a, b]$ should contain all the sample entries.

This distribution does not belong to the exponential family, as its support depends on the parameters. Therefore, the moment estimate of the parameters is not the same as the MLE, in contrast to the first three exampes that belonged to the exponential family. The moment estimation here as follows.

The theoretical first two moments are:

$$m_1 = \frac{a+b}{2}, \quad m_2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4}.$$

It can easily be checked that the Jacobian of the $(a, b) \to (m_1, m_2)$ map is not zero, and the inverse map is

$$a = m_1 - \sqrt{3(m_2 - m_1^2)}, \qquad b = m_1 + \sqrt{3(m_2 - m_1^2)}.$$

Eventually, by

$$\hat{m}_1 = \bar{X}, \qquad \hat{m}_2 - \hat{m}_1^2 = S_n^2$$

the moment estimates of the parameters are

$$\hat{a} = \bar{X} - \sqrt{3} S_n, \qquad \hat{b} = \bar{X} + \sqrt{3} S_n$$

that are not the same as the ML estimetes.