MATH2715: Statistical Methods – Worked Examples on Estimators

---

Worked Example: For a random sample of size n from an exponential(λ) distribution, λ can be estimated using $1/\bar{X}$. What is the mean, the bias, the variance and MSE of this estimator?

Answer: If $X_i \sim$ exponential(λ), then $S = \sum_{i=1}^{n} X_i \sim$ gamma(n, λ) with probability density function $f_S(s) = \dfrac{\lambda^n s^{n-1} e^{-\lambda s}}{\Gamma(n)}$.

Then $\bar{X} = S/n = Y$ has pdf $f_Y(y) = f_S(s)\left|\dfrac{\mathrm{d}s}{\mathrm{d}y}\right| = nf_S(s) = n\dfrac{\lambda^n s^{n-1}e^{-\lambda s}}{\Gamma(n)} = \dfrac{(n\lambda)^n y^{n-1} e^{-n\lambda y}}{\Gamma(n)}$ where $s = ny$. Hence we can recognise that $Y \sim$ gamma($n, n\lambda$).

<u>Mean of $1/\bar{X}$:</u>

$$\begin{aligned}
\mathrm{E}[1/\bar{X}] = \mathrm{E}[1/Y] &= \int_0^\infty \frac{1}{y} f_Y(y)\,\mathrm{d}y = \int_0^\infty \frac{1}{y} \times \frac{(n\lambda)^n y^{n-1} e^{-n\lambda y}}{\Gamma(n)}\,\mathrm{d}y \\
&= \frac{n\lambda\Gamma(n-1)}{\Gamma(n)} \int_0^\infty \frac{(n\lambda)^{n-1} y^{n-2} e^{-n\lambda y}}{\Gamma(n-1)}\,\mathrm{d}y = \frac{n\lambda\Gamma(n-1)}{\Gamma(n)} = \frac{n\lambda}{n-1}
\end{aligned}$$

as area under a gamma($n-1, n\lambda$) probability density function is one.

<u>Bias of $1/\bar{X}$:</u> Bias $= \mathrm{E}[1/\bar{X}] - \lambda = \dfrac{n\lambda}{n-1} - \lambda = \dfrac{\lambda}{n-1}$.

<u>Variance of $1/\bar{X}$:</u>

$$\begin{aligned}
\mathrm{E}[1/\bar{X}^2] = \mathrm{E}[1/Y^2] &= \int_0^\infty \frac{1}{y^2} f_Y(y)\,\mathrm{d}y = \int_0^\infty \frac{(n\lambda)^n y^{n-1} e^{-n\lambda y}}{y^2\Gamma(n)}\,\mathrm{d}y \\
&= \frac{n^2\lambda^2\Gamma(n-2)}{\Gamma(n)} \int_0^\infty \frac{(n\lambda)^{n-2} y^{n-3} e^{-n\lambda y}}{\Gamma(n-2)}\,\mathrm{d}y = \frac{n^2\lambda^2\Gamma(n-2)}{\Gamma(n)} = \frac{n^2\lambda^2}{(n-1)(n-2)}
\end{aligned}$$

as area under a gamma($n-2, n\lambda$) probability density function is one. Thus

$$\mathrm{Var}[1/\bar{X}] = \mathrm{E}[1/Y^2] - \{\mathrm{E}[1/Y]\}^2 = \frac{n^2\lambda^2}{(n-1)(n-2)} - \frac{n^2\lambda^2}{(n-1)^2} = \frac{n^2\lambda^2}{(n-1)^2(n-2)}.$$

<u>MSE of $1/\bar{X}$:</u> MSE $= \mathrm{Var}[1/\bar{X}] + (\text{bias})^2 = \dfrac{n^2\lambda^2}{(n-1)^2(n-2)} + \dfrac{\lambda^2}{(n-1)^2} = \dfrac{(n^2+n-2)\lambda^2}{(n-1)^2(n-2)}$.

Worked Example: If $X \sim \mathrm{Bin}(n, \theta)$, the method of moments estimator of θ is X/n. Consider estimating the odds ratio $\theta/(1-\theta)$ using the simplistic estimator

$$U = \frac{X/n}{1 - X/n} = \frac{X}{n-X}.$$

What is the mean of U?

Answer:

$$\mathrm{E}[U] = \mathrm{E}\left[\frac{X}{n-X}\right] = \sum_{x=0}^{n} \frac{x}{n-x}\,\mathrm{pr}\{X = x\} = \infty$$

since $X = n$ (and $U = \infty$) with finite probability $\mathrm{pr}\{X = n\} = \theta^n$. The "obvious" estimator U of $\dfrac{\theta}{1-\theta}$ is clearly no good![1]

Worked Example: Suppose a random sample $x_1, x_2, \ldots, x_n$ of n values is taken from a distribution with probability density function $f_X(x;\theta) = \theta(1-x)^{\theta-1}$ with $0 < \theta < 1$. Obtain the maximum likelihood estimate for θ.

Answer: Joint probability density function is

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = \prod_{i=1}^{n} f_X(x_i,\theta) = \prod_{i=1}^{n} \left\{\theta(1-x_i)^{\theta-1}\right\} = \theta^n \left\{(1-x_1)(1-x_2)\cdots(1-x_n)\right\}^{\theta-1}.$$

Likelihood function is $L(\theta;\boldsymbol{x}) = \theta^n \left\{(1-x_1)(1-x_2)\cdots(1-x_n)\right\}^{\theta-1}$. The log-likelihood function is $l(\theta;\boldsymbol{x}) = \log L(\theta;\boldsymbol{x}) = n\log\theta + (\theta-1)\log\{(1-x_1)(1-x_2)\cdots(1-x_n)\}$.

$$\frac{\mathrm{d}l(\theta;\boldsymbol{x})}{\mathrm{d}\theta} = \frac{n}{\theta} + \log\{(1-x_1)(1-x_2)\cdots(1-x_n)\}.$$

$$\frac{\mathrm{d}l(\theta;\boldsymbol{x})}{\mathrm{d}\theta} = 0 \quad \Rightarrow \quad \widehat{\theta} = \frac{-n}{\log\{(1-x_1)(1-x_2)\cdots(1-x_n)\}}.$$[2]

Worked Example: Suppose a single value X is available from a geometric(θ) distribution with $0 < \theta < 1$. Obtain the maximum likelihood estimate for θ and obtain its mean.

Answer: $\mathrm{pr}\{X = x\} = \theta(1-\theta)^x$, $x = 0, 1, 2, \ldots$, so that the likelihood function is $L(\theta;x) = \theta(1-\theta)^x$, $0 < \theta < 1$, and the log-likelihood function is $l(\theta;x) = \log\theta + x\log(1-\theta)$.

---

[1] A better estimator is $U^\star = \dfrac{X}{n-X+1}$. In this case

$$\mathrm{E}\left[\frac{X}{n-X+1}\right] = \sum_{x=0}^{n} \frac{x}{n-x+1}\mathrm{pr}\{X=x\} = \sum_{x=1}^{n} \frac{x}{n-x+1}\cdot\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$\boxed{\text{since} \quad \frac{x}{n-x+1}\binom{n}{x} = \frac{x}{n-x+1}\cdot\frac{n!}{x!(n-x)!} = \frac{n!}{(x-1)!(n-x+1)!} = \binom{n}{x-1}}$$

$$= \frac{\theta}{1-\theta}\sum_{x=1}^{n}\binom{n}{x-1}\theta^{x-1}(1-\theta)^{n-x+1} = \frac{\theta}{1-\theta}\sum_{y=0}^{n-1}\binom{n}{y}\theta^y(1-\theta)^{n-y}$$

$$\boxed{\text{putting } y = x-1\text{, and also noting that } Y \sim \mathrm{Bin}(n,\theta) \text{ satisfies} \quad \sum_{y=0}^{n}\mathrm{pr}\{Y=y\} = \sum_{y=0}^{n}\binom{n}{y}\theta^y(1-\theta)^{n-y} = 1}$$

$$= \frac{\theta}{1-\theta}\left\{\sum_{y=0}^{n}\binom{n}{y}\theta^y(1-\theta)^{n-y} - \binom{n}{n}\theta^n(1-\theta)^0\right\} = \frac{\theta}{1-\theta}\left\{1-\theta^n\right\}.$$

For large n, since $0 < \theta < 1$, it can be seen that $\mathrm{E}[U^\star] \longrightarrow \dfrac{\theta}{1-\theta}$.

[2] Notice we can write $\widehat{\theta} = \dfrac{-n}{\sum_{i=1}^{n}\log(1-x_i)}$. Also notice that $\dfrac{\mathrm{d}^2 l(\theta;\boldsymbol{x})}{\mathrm{d}\theta^2} = -\dfrac{n}{\theta^2}$, so $\left.\dfrac{\mathrm{d}^2 l(\theta;\boldsymbol{x})}{\mathrm{d}\theta^2}\right|_{\theta=\widehat{\theta}} < 0$ and $\widehat{\theta}$ does give the maximum.

Maximising the likelihood:

$$\frac{\mathrm{d}L(\theta;x)}{\mathrm{d}\theta} = (1-\theta)^x - x\theta(1-\theta)^{x-1} \quad \text{so} \quad \frac{\mathrm{d}L(\theta;x)}{\mathrm{d}\theta} = 0 \;\Rightarrow\; (1-\widehat{\theta}) = x\widehat{\theta} \quad \text{giving} \quad \widehat{\theta} = \frac{1}{x+1}.$$

Maximising the log-likelihood:

$$\frac{\mathrm{d}l(\theta;x)}{\mathrm{d}\theta} = \frac{1}{\theta} - \frac{x}{1-\theta} \quad \text{and} \quad \frac{\mathrm{d}l(\theta;x)}{\mathrm{d}\theta} = 0 \;\Rightarrow\; \frac{1}{\widehat{\theta}} = \frac{x}{1-\widehat{\theta}} \quad \text{so} \quad \widehat{\theta} = \frac{1}{x+1}.^3$$

Mean:

$$\begin{aligned}
\mathrm{E}[\widehat{\theta}] &= \mathrm{E}\left[\frac{1}{X+1}\right] = \sum_{x=0}^{\infty} \frac{1}{x+1} \cdot \theta(1-\theta)^x \\
&= \theta + \tfrac{1}{2}\theta(1-\theta) + \tfrac{1}{3}\theta(1-\theta)^2 + \cdots \\
&= \frac{\theta}{1-\theta}\left((1-\theta) + \tfrac{1}{2}(1-\theta)^2 + \tfrac{1}{3}(1-\theta)^3 + \cdots\right) = \frac{-\theta\log\theta}{1-\theta}
\end{aligned}$$

since, for $|\delta| < 1$, $\log(1-\delta) = -\delta - \frac{1}{2}\delta^2 - \frac{1}{3}\delta^3 - \cdots$.

Worked Example: A simple random sample $X_1, X_2, \ldots, X_n$ of size n is drawn from a population with probability density function[4]

$$f_X(x;\theta) = \frac{\theta e^{-\theta x}}{1-e^{-\theta}}, \quad 0 < x < 1, \quad \text{with} \quad \theta > 0.$$

Show that the likelihood function is a function of the sample mean $\bar{x}$ of the observations.

Answer: The joint probability density function is

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = \prod_{i=1}^{n} f_X(x_i;\theta) = \prod_{i=1}^{n} \frac{\theta e^{-\theta x_i}}{1-e^{-\theta}} = \left(\frac{\theta}{1-e^{-\theta}}\right)^n \exp\left(-\sum_{i=1}^{n}\theta x_i\right).$$

The likelihood function is

$$L(\theta;\boldsymbol{x}) = \left(\frac{\theta}{1-e^{-\theta}}\right)^n \exp\left(-\sum_{i=1}^{n}\theta x_i\right) = \left(\frac{\theta}{1-e^{-\theta}}\right)^n e^{-n\theta\bar{x}}.$$

This is a function of $\bar{x}$ only. Thus $\widehat{\theta}$ will be a function, albeit complicated, of $\bar{x}$ only. This answers the question![5]

---

[3] Notice $\frac{\mathrm{d}^2 l(\theta;x)}{\mathrm{d}\theta^2} = -\frac{1}{\theta^2} - \frac{x}{(1-\theta)^2}$. Putting $\theta = \widehat{\theta}$ gives $\frac{\mathrm{d}^2 l(\theta;x)}{\mathrm{d}\theta^2} = \frac{-(x+1)^3}{x} < 0$ so $\widehat{\theta}$ does give the maximum.

[4] The probability density function of X is a truncated exponential density. The $1-e^{-\theta}$ term in the denominator ensures the probability density function of X integrates to one. Thus if $Z \sim$ exponential(θ), then $\mathrm{pr}\{Z \le 1\} = 1-e^{-\theta}$ and the probability density function of Z given $\{Z \le 1\}$ is

$$f_Z(z|Z \le 1) = \frac{f_Z(z)}{\mathrm{pr}\{Z \le 1\}} = \frac{\theta e^{-\theta z}}{1-e^{-\theta}}.$$

The following R commands generates values from the distribution for X.

```
z=rexp(100)                # Generate 100 values z exponential(1).
x=z[z<1]                   # Put values of z<1 into x.
```

[5] How could we obtain the maximum likelihood estimate in this case?
The log-likelihood function is $l(\theta;\boldsymbol{x}) = n\log\left(\frac{\theta}{1-e^{-\theta}}\right) - n\theta\bar{x}$.

Worked Example: A simple random sample $X_1, X_2, \ldots, X_n$ of size n is drawn from a population with probability density function

$$f_X(x;\theta) = \frac{\theta e^{-\theta x}}{1-e^{-\theta}}, \quad 0 < x < 1, \quad \text{with} \quad \theta > 0.$$

Suppose however that the only information available about each sample value is whether or not it is greater than $\frac{1}{2}$. Thus the sample consists of independent values $Y_1, \ldots, Y_n$ satisfying

$$Y_i = \begin{cases} 0 \ \text{ if } \ x \le \frac{1}{2}, & \text{with probability } 1-\phi, \\ 1 \ \text{ if } \ x > \frac{1}{2}, & \text{with probability } \phi, \end{cases}$$

where[6] $\phi = \text{pr}\{Y_i = 1\} = \dfrac{e^{-\frac{1}{2}\theta}}{1+e^{-\frac{1}{2}\theta}}$ and satisfies $0 < \phi < 0.5$

Derive the maximum likelihood estimator of ϕ. Use the 1–1 relationship between ϕ and θ to deduce the maximum likelihood estimator for θ in this case.

Answer: The sample now consists of independent values $Y_1, \ldots, Y_n$ taking values 0 and 1 with probabilities $1-\phi$ and ϕ respectively. Thus, if $Y = \sum_{i=1}^{n} y_i$ is the number of Y_i equal to one, the joint probability function of $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$ is

$$\text{pr}\{\boldsymbol{Y} = \boldsymbol{y}; \phi\} = \text{pr}\{Y_1 = y_1\} \times \text{pr}\{Y_2 = y_2\} \times \text{pr}\{Y_3 = y_3\} \times \cdots \times \text{pr}\{Y_n = y_n\} = \phi^Y (1-\phi)^{n-Y}.$$

The likelihood function is $L(\phi; \boldsymbol{y}) = \phi^Y (1-\phi)^{n-Y}$.

The log-likelihood is $l(\phi; \boldsymbol{y}) = \log L(\phi; \boldsymbol{y}) = Y \log(\phi) + (n-Y)\log(1-\phi)$.

Hence $\dfrac{\mathrm{d}l(\phi;\boldsymbol{y})}{\mathrm{d}\phi} = \dfrac{Y}{\phi} - \dfrac{n-Y}{1-\phi}$ so $\dfrac{\mathrm{d}l(\phi;\boldsymbol{y})}{\mathrm{d}\phi} = 0 \quad \Rightarrow \quad \widehat{\phi} = Y/n.$[7] [8]

$$\frac{\mathrm{d}l}{\mathrm{d}\theta} = n\left(\frac{1-e^{-\theta}}{\theta}\right)\left(\frac{(1-e^{-\theta}) - \theta e^{-\theta}}{(1-e^{-\theta})^2}\right) - n\bar{x} = \frac{(1-e^{-\theta}) - \theta e^{-\theta}}{\theta(1-e^{-\theta})} - \bar{x}.$$

The maximum likelihood estimator $\widehat{\theta}$ thus satisfies $\dfrac{(1-e^{-\widehat{\theta}}) - \widehat{\theta} e^{-\widehat{\theta}}}{\widehat{\theta}(1-e^{-\widehat{\theta}})} - \bar{x} = 0.$

Clearly it is not possible to obtain an analytical solution for $\widehat{\theta}$. One approach is to simply plot $l(\theta; \boldsymbol{x})$ against θ to obtain information about the maximum. A Newton-Raphson method gives another possible approach to find a numerical solution for the maximum likelihood estimate.

Suppose $\theta = \theta_1$ is an approximation to the root of $g(\theta) = 0$. Let the exact root be at $\theta = \theta_1 + h$. Then, by Taylor's expansion,

$$0 = g(\theta_1 + h) = g(\theta_1) + hg'(\theta_1) + \tfrac{1}{2}h^2 g''(\theta_1) + \cdots.$$

Neglecting terms in h^2 and higher order gives $h \approx -\dfrac{g(\theta_1)}{g'(\theta_1)}$. An improved estimate for the root is thus $\theta_2 = \theta_1 + h$. This process can be iterated to give a better approximation to the root of $g(\theta)$. In this case we have

$$g(\theta) = \frac{(1-e^{-\theta}) - \theta e^{-\theta}}{\theta(1-e^{-\theta})} - \bar{x}, \qquad g'(\theta) = \frac{\theta^2 e^{-\theta} - (1-e^{-\theta})^2}{\theta^2(1-e^{-\theta})^2}.$$

[6] $\phi = \text{pr}\{Y = 1\} = \text{pr}\{X > \frac{1}{2}\} = \displaystyle\int_{\frac{1}{2}}^{1} \frac{\theta e^{-\theta x}}{1-e^{-\theta}}\,\mathrm{d}x = \left[\frac{-e^{-\theta x}}{1-e^{-\theta}}\right]_{\frac{1}{2}}^{1} = \frac{e^{-\frac{1}{2}\theta} - e^{-\theta}}{(1-e^{-\frac{1}{2}\theta})(1+e^{-\frac{1}{2}\theta})} = \frac{e^{-\frac{1}{2}\theta}}{1+e^{-\frac{1}{2}\theta}}.$

[7] This makes sense! The estimate of ϕ is the proportion of cases for which Y_i equals one!

[8] Notice $\dfrac{\mathrm{d}^2 l(\phi;\boldsymbol{y})}{\mathrm{d}\phi^2} = -\dfrac{Y}{\phi^2} - \dfrac{(n-Y)}{(1-\phi)^2}$. Putting $\phi = \widehat{\phi}$ gives, for $0 < Y < n$, $\dfrac{\mathrm{d}^2 l(\phi;\boldsymbol{y})}{\mathrm{d}\phi^2} = \dfrac{-n^3}{Y(n-Y)} < 0$ so $\widehat{\phi}$ does here give the maximum.

However more care needs to be taken in finding the maximum likelihood estimate!

Suppose $Y = 0$. Then $L(\phi; \boldsymbol{y}) = (1-\phi)^n$ and is maximised by $\widehat{\phi} = 0$ (though this is not a stationary point so the maximum cannot be found by differentiation of $L(\phi; \boldsymbol{y})$ or $l(\phi; \boldsymbol{y})$). Here $\widehat{\phi} = 0$ is on the boundary of the allowed region for ϕ.

Suppose now $Y > \frac{1}{2}n$. Using $\widehat{\phi} = Y/n$ would give $\widehat{\phi} > 0.5$, which is outside the allowed region for ϕ. Though $l(\phi; \boldsymbol{y})$ is maximised by $\phi = Y/n$ for $0 < \phi < 1$, the maximum value of $l(\phi; \boldsymbol{y})$ for $0 < \phi < 0.5$ occurs in this case at $\widehat{\phi} = 0.5$. Though the maximum, this is again not a stationary point of the likelihood or log-likelihood function. Here $\widehat{\phi} = 0.5$ is on the boundary of the allowed region for ϕ.

Thus the maximum likelihood estimator for ϕ is

$$\widehat{\phi} = \begin{cases} Y/n & \text{if} \quad 0 \le Y \le \frac{1}{2}n, \\ 0.5 & \text{if} \quad Y > \frac{1}{2}n. \end{cases}$$

Since $\phi = \dfrac{e^{-\frac{1}{2}\theta}}{1 + e^{-\frac{1}{2}\theta}}$ gives a 1–1 relationship between ϕ and θ, it follows the maximum likelihood estimator for θ satisfies

$$\frac{e^{-\frac{1}{2}\widehat{\theta}}}{1 + e^{-\frac{1}{2}\widehat{\theta}}} = \begin{cases} Y/n & \text{if} \quad 0 \le Y \le \frac{1}{2}n, \\ 0.5 & \text{if} \quad Y > \frac{1}{2}n. \end{cases}$$

This gives[9]

$$e^{-\frac{1}{2}\widehat{\theta}} = \begin{cases} \frac{Y}{n-Y} & \text{if} \quad 0 \le Y \le \frac{1}{2}n, \\ 1 & \text{if} \quad Y > \frac{1}{2}n, \end{cases} \qquad \Longrightarrow \qquad \widehat{\theta} = \begin{cases} -2\log\left(\frac{Y}{n-Y}\right) & \text{if} \quad 0 \le Y \le \frac{1}{2}n, \\ 0 & \text{if} \quad Y > \frac{1}{2}n. \end{cases}$$

[9] This estimator for θ will have greater variance than that in the previous question when we knew the X_i values. This is because the estimator here utilises less information than if we know the precise value of the X_i.

If $Y = 0$, then $\widehat{\theta}$ is not defined. This is a major problem if θ is large.