Maximum Likelihood Estimation, Bias, Mean Squared Error

A. Stepanov
D. Simpson
S. Culpepper

Definition: Maximum Likelihood Estimator (MLE)

p.m.f. or p.d.f. $f(x;\theta)$, $\theta \in \Omega$. Ω – parameter space.

Likelihood function for a sample of i.i.d. $X_1, \dots, X_n$,

$$L(\theta;\mathbf{x}) = L(\theta;x_1,\dots,x_n) = f(x_1,\dots,x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

where $\mathbf{x} = (x_1, \dots, x_n)'$ is a vector of sample observations.

It is often easier to consider the log-likelihood,

$$\ell(\theta;\mathbf{x}) = \ln[L(\theta;x_1,\dots,x_n)] = \sum_{i=1}^{n} \ln[f(x_i;\theta)]$$

Assumptions (Regularity Conditions):

(R0) The pdfs are distinct; i.e., $\theta \neq \theta' \Rightarrow f(x_i;\theta) \neq f(x_i;\theta')$.

(R1) The pdfs have common support for all θ.

(R2) The true unknown point θ_0 is an interior point in Ω.

Theorem 6.1.1. Let θ_0 be the true parameter. Under assumptions (R0) and (R1),

$$\lim_{n\to\infty} P[L(\theta_0;\mathbf{x}) > L(\theta;\mathbf{x})] = 1 \;\forall\; \theta_0 \neq \theta.$$

Asymptotically the likelihood function is maximized at the true value θ_0.

Let $\hat{\theta}$ be the maximum likelihood estimate (m.l.e.) of θ,

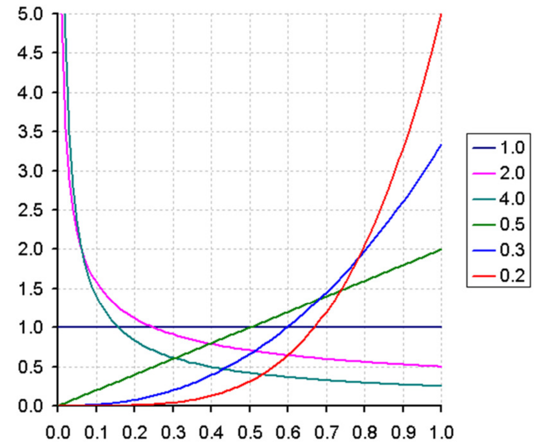
$$\hat{\theta} = Argmax\; L(\theta;\mathbf{x})$$

Maximum Likelihood Estimation, Bias, Mean Squared Error

A. Stepanov
D. Simpson
S. Culpepper

Example 1. Let $X_1, \ldots, X_n$ be a random sample of size n from the distribution with probability density function,

$$f_X(x;\theta) = \begin{cases} \frac{1}{\theta} x^{\frac{1-\theta}{\theta}}, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}$$
$$0 < \theta < \infty$$



a) Obtain the method of moments estimator of θ, $\tilde{\theta}$.

$$E(X) = \int_{-\infty}^{\infty} x f_X(x;\theta)dx = \frac{1}{\theta}\int_0^1 x^{\frac{1}{\theta}}dx = \frac{1}{\theta}\left[\frac{1}{\frac{1}{\theta}+1} x^{\frac{1}{\theta}+1}\right]\Bigg|_0^1 = \frac{1}{1+\theta}$$

$$\bar{X} = \frac{1}{1+\theta} \Rightarrow \tilde{\theta} = \frac{1-\bar{X}}{\bar{X}}.$$

b) Obtain the maximum likelihood estimator of θ, $\hat{\theta}$.

Likelihood function:

$$L(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta) = \frac{1}{\theta^n}\left[\prod_{i=1}^{n} x_i\right]^{\frac{1-\theta}{\theta}}$$

$$\ell(\theta;\mathbf{x}) = -n\ln\theta + \left(\frac{1-\theta}{\theta}\right)\sum_{i=1}^{n}\ln x_i = -n\ln\theta + \left(\frac{1}{\theta}-1\right)\sum_{i=1}^{n}\ln x_i$$

$$\frac{d\ell(\theta;\mathbf{x})}{d\theta} = -\frac{n}{\theta} - \frac{1}{\theta^2}\sum_{i=1}^{n}\ln x_i = 0$$

$$\Rightarrow \hat{\theta} = -\frac{1}{n}\sum_{i=1}^{n}\ln x_i.$$

c) Suppose $n = 3$, and $x_1 = 0.2$, $x_2 = 0.3$, $x_1 = 0.5$. Compute the values of the method of moments estimate and the maximum likelihood estimate for θ.

$$\bar{X} = \frac{0.2 + 0.3 + 0.5}{3} = \frac{1}{3}$$

$$\tilde{\theta} = \frac{1 - \bar{X}}{\bar{X}} = \frac{1 - \frac{1}{3}}{\frac{1}{3}} = 2$$

$$\hat{\theta} = -\frac{1}{3}\sum_{i=1}^{3}\ln x_i = -\frac{1}{3}(\ln 0.2 + \ln 0.3 + \ln 0.5) = -\frac{1}{3}\ln 0.03 \approx 1.16885$$

Def An estimator $\hat{\theta}$ is said to be **unbiased for θ** if $E\left(\hat{\theta}\right) = \theta$.

Example 2. Reconsider the prior pdf. Are $\hat{\theta}$ and $\tilde{\theta}$ unbiased estimators?

a) Is $\hat{\theta}$ unbiased for θ? That is, does $E\left(\hat{\theta}\right) = \theta$?

$$E[\ln(X)] = \int_0^1 \ln(x)\frac{1}{\theta}x^{\frac{1-\theta}{\theta}}dx.$$

Integration by parts: $\int_a^b u\,dv = uv|_a^b - \int_a^b v\,du$

$$u = \ln x, \qquad dv = \frac{1}{\theta} x^{\frac{1-\theta}{\theta}} dx,$$
$$du = \frac{1}{x} dx \qquad v = x^{\frac{1}{\theta}}$$

$$E[\ln(X)] = \int_0^1 \ln(x) \frac{1}{\theta} x^{\frac{1-\theta}{\theta}} dx = \left(\ln(x)\, x^{\frac{1}{\theta}}\right)\Big|_0^1 - \int_0^1 \frac{1}{x} x^{\frac{1}{\theta}} dx$$
$$= -\int_0^1 x^{\frac{1}{\theta}-1} dx = -\left(\frac{1}{\frac{1}{\theta}} x^{\frac{1}{\theta}}\right)\Bigg|_0^1 = -\theta$$

Therefore,

$$E(\hat{\theta}) = -\frac{1}{n}\sum_{i=1}^{n} E[\ln(x_i)] = -\frac{1}{n}\sum_{i=1}^{n}(-\theta) = \theta$$

that is, $\hat{\theta}$ is an unbiased estimator for θ.

OR

$$F_X(x) = x^{\frac{1}{\theta}}, 0 < x < 1.$$

Let $Y_i = -\ln(X_i), i = 1, \dots, n$.

$$F_Y(y) = P(Y \le y) = P(X \ge e^{-y})$$

$$= 1 - F_X(e^{-y}) = 1 - e^{-\frac{y}{\theta}}, y > 0$$

$$\Rightarrow Y_1, \dots, Y_n \; are\; i.i.d.\, Exponential(\theta)$$

Then $\hat{\theta} = \bar{Y}$. $E(\hat{\theta}) = E(\bar{Y}) = E(Y) = \theta$, that is, $\hat{\theta}$ is an unbiased estimator for θ.

b) Is $\tilde{\theta}$ unbiased for θ? That is, does $E(\tilde{\theta}) = \theta$?

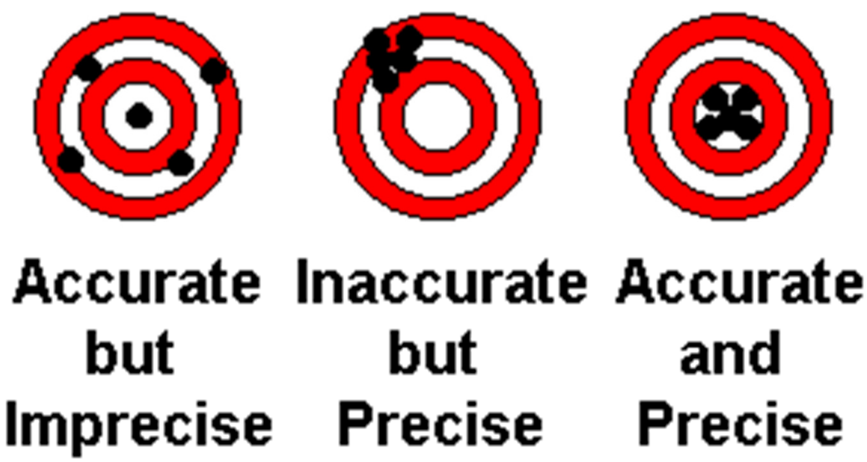
Since $g(x) = \frac{1-x}{x} = \frac{1}{x} - 1,\ 0 < x < 1$, is strictly convex, and $\bar{X}$ is not a constant random variable, by Jensen's Inequality (Theorem 1.10.5),

$$E(\tilde{\theta}) = E[g(\bar{X})] > g[E(\bar{X})] = \theta.$$

$\tilde{\theta}$ is NOT an unbiased estimator for θ.

Def For an estimator $\hat{\theta}$ of θ, define the **Mean Squared Error** of $\hat{\theta}$ by,

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = [E(\hat{\theta}) - \theta]^2 + Var(\hat{\theta}) = [bias(\hat{\theta})]^2 + Var(\hat{\theta})$$



unbiased, large variance	biased, small variance	unbiased, small variance

Note Chebyshev's inequality implies,

$$P(|\hat{\theta} - \theta| \geq \epsilon) \leq \frac{E\left[(\hat{\theta} - \theta)^2\right]}{\epsilon^2} = \frac{MSE(\hat{\theta})}{\epsilon^2}$$

Example 3. What is the **Mean Squared Errors** for $\hat{\theta}$. That is, find $MSE(\hat{\theta})$.

Let $Y_i = -\ln X_i \,, i = 1, \dots, n$. Then $E(Y) = \theta$, $Var(Y) = \theta^2$.

$$\hat{\theta} = \bar{Y}$$

$$\Rightarrow Var(\hat{\theta}) = Var(\bar{Y}) = \frac{Var(Y)}{n} = \frac{\theta^2}{n}$$

$$MSE(\hat{\theta}) = \left[bias(\hat{\theta})\right]^2 + Var(\hat{\theta}) = \frac{\theta^2}{n}$$

Def Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for θ. $\hat{\theta}_1$ is said to be **more efficient** than $\hat{\theta}_2$ if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$.

The **relative efficiency** of $\hat{\theta}_1$ with respect to $\hat{\theta}_2$ is $Var(\hat{\theta}_2)/Var(\hat{\theta}_1)$.

Example 4. What is the relative *asymptotic* efficiency of $\hat{\theta}$ to $\tilde{\theta}$.

Recall, from the convergence part 2 notes that the asymptotic variance of $\tilde{\theta}$ is,

$$\frac{\theta^2(1+\theta)^2}{(1+2\theta)n}$$

The asymptotic relative efficiency is,

$$\frac{\frac{\theta^2(1+\theta)^2}{(1+2\theta)n}}{\frac{\theta^2}{n}} = \frac{(1+\theta)^2}{1+2\theta}.$$

$\tilde{\theta}$ is asymptotically less efficient than $\hat{\theta}$.

Example 5. Let $\lambda > 0$ and let $X_1, \dots, X_n$ be a random sample from the distribution with the probability density function,

$$f(x;\theta) = 2\lambda^2 x^3 e^{-\lambda x^2}, \qquad x > 0$$

a) Find $E(X^k)$, $k > -4$.

Hint 1: Consider $u = \lambda x^2$ or $u = x^2$.
Hint 2: $\Gamma(a) = \int_0^\infty u^{a-1} e^{-u} du, a > 0$.

$$E(X^k) = \int_0^\infty x^k 2\lambda^2 x^3 e^{-\lambda x^2} dx, \qquad u = \lambda x^2, du = 2\lambda x\, dx$$

$$= \lambda \int_0^\infty \left(\frac{u}{\lambda}\right)^{\frac{k}{2}+1} e^{-u} dx = \lambda^{-\frac{k}{2}} \int_0^\infty u^{\frac{k}{2}+1} e^{-u} dx$$

$$= \lambda^{-\frac{k}{2}} \Gamma\left(\frac{k}{2} + 2\right).$$

b) Obtain a method of moments estimator of λ, $\tilde{\lambda}$.

$$E(X) = \lambda^{-\frac{1}{2}} \Gamma\left(\frac{1}{2} + 2\right) = \lambda^{-\frac{1}{2}} \Gamma\left(\frac{1}{2} + 2\right) = \frac{3}{2} \lambda^{-\frac{1}{2}} \Gamma\left(\frac{3}{2}\right) = \frac{3}{4} \lambda^{-\frac{1}{2}} \Gamma\left(\frac{1}{2}\right) = \frac{3}{4}\sqrt{\frac{\pi}{\lambda}}$$

$$\bar{X} = \frac{3}{4}\sqrt{\frac{\pi}{\lambda}} \Rightarrow \tilde{\lambda}_1 = \frac{9\pi}{16(\bar{X})^2}.$$

OR

$$E(X^2) = \lambda^{-\frac{2}{2}} \Gamma\left(\frac{2}{2} + 2\right) = \lambda^{-1} \Gamma(3) = \frac{2}{\lambda}.$$

$$\overline{X^2} = \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{2}{\lambda} \Rightarrow \tilde{\lambda}_2 = \frac{2n}{\sum_{i=1}^n X_i^2}.$$

c) Obtain the maximum likelihood estimator of λ, $\hat{\lambda}$.

$$L(\lambda;\mathbf{x}) = \prod_{i=1}^{n}\left(2\lambda^2 x_i^3 e^{-\lambda x_i^2}\right) = 2^n \lambda^{2n} \exp\left(-\lambda \sum_{i=1}^{n} x_i^2\right)\left(\prod_{i=1}^{n} x_i\right)^3.$$

$$\ell(\lambda;\mathbf{x}) = \ln[L(\lambda;\mathbf{x})] = n\ln(2) + 2n\ln(\lambda) - \lambda\sum_{i=1}^{n} x_i^2 + 3\sum_{i=1}^{n}\ln(x_i).$$

$$\ell'(\lambda;\mathbf{x}) = \frac{2n}{\lambda} - \sum_{i=1}^{n} x_i^2 = 0 \Rightarrow \hat{\lambda} = \frac{2n}{\sum_{i=1}^{n} x_i^2}.$$

d) Suppose $x_1 = 0.6$, $x_2 = 1.1$, $x_3 = 2.7$, $x_4 = 3.3$, $x_5 = 4.5$. Compute $\hat{\lambda}, \tilde{\lambda}_1$, and $\tilde{\lambda}_2$.

Note that,

$$\bar{X} = \frac{12.2}{5} = 2.44, \sum_{i=1}^{5} x_i^2 = 40$$

$$\hat{\lambda} = \frac{2n}{\sum_{i=1}^{n} x_i^2} = \frac{10}{40} = 0.25$$

$$\tilde{\lambda}_1 = \frac{9\pi}{16(\bar{X})^2} = \frac{9\pi}{16(2.44)^2} \approx 0.29682$$

$$\tilde{\lambda}_2 = \frac{10}{40} = 0.25$$

e) What is the probability distribution of $W = X^2$?

$$W = X^2 \Rightarrow X = v(W) = \sqrt{W}$$

$$\frac{dX}{dW} = v'(W) = \frac{1}{2\sqrt{w}}$$

$$f_W(w) = f_X[v(W)]|v'(W)| = \lambda^2 w e^{-\lambda w}, \qquad w > 0.$$

$$\Rightarrow W \sim Gamma\left(\alpha = 2, \theta = \frac{1}{\lambda}\right).$$

f) What is the probability distribution of $Y = \sum_{i=1}^{n} X_i^2$?

The iid assumption implies,

$$M_Y(t) = E(e^{Yt}) = E\left(e^{(\sum_{i=1}^{n} X_i^2)t}\right) = \prod_{i=1}^{n} E(e^{W_i t}) = \left(\frac{1}{1-\theta t}\right)^{2n}$$

$$\Rightarrow Y = \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} W_i \sim Gamma\left(\alpha = 2n, \theta = \frac{1}{\lambda}\right)$$

g) Let $Y = \sum_{i=1}^{n} X_i^2$. Find $E\left(\frac{1}{Y}\right)$.

Recall,

$$Y \sim Gamma\left(\alpha = 2n, \theta = \frac{1}{\lambda}\right)$$

$$E\left(\frac{1}{Y}\right) = \int_0^\infty \frac{1}{y} \frac{\lambda^{2n}}{\Gamma(2n)} y^{2n-1} e^{-\lambda y} dy$$

$$= \frac{\lambda^{2n}}{\Gamma(2n)} \frac{\Gamma(2n-1)}{\lambda^{2n-1}} \int_0^\infty \frac{\lambda^{2n-1}}{\Gamma(2n-1)} y^{2n-1-1} e^{-\lambda y} dy$$

$$= \frac{\lambda}{2n-1}$$

h) Is the maximum likelihood estimator of λ, $\hat{\lambda}$, an unbiased estimator of λ? If not, construct an unbiased estimator of λ based on $\hat{\lambda}$.

$$E(\hat{\lambda}) = E\left(\frac{2n}{Y}\right) = \frac{2n}{2n-1}\lambda$$

$\hat{\lambda}$ is not an unbiased estimator of λ.

Consider,

$$\hat{\hat{\lambda}} = \frac{2n-1}{2n}\hat{\lambda} = \frac{2n-1}{\sum_{i=1}^{n} x_i^2}.$$

Then,

$$E\left(\hat{\hat{\lambda}}\right) = E\left(\frac{2n-1}{Y}\right) = \lambda$$

i) Show that $\hat{\lambda}$ and $\tilde{\lambda}_1$ are consistent estimators of λ.

$$\hat{\lambda} = \frac{2n}{\sum_{i=1}^{n} x_i^2} = \frac{2}{\frac{\sum_{i=1}^{n} x_i^2}{n}} \xrightarrow{P} \frac{2}{E(X^2)} = \frac{2}{2/\lambda} = \lambda$$

$$\tilde{\lambda}_1 = \frac{9\pi}{16(\bar{X})^2} \xrightarrow{P} \frac{9\pi}{16[E(X)]^2} = \frac{9\pi}{16\left[\frac{3}{4}\sqrt{\frac{\pi}{\lambda}}\right]^2} = \lambda.$$

j) Find $MSE(\hat{\lambda}) = E\left[\left(\hat{\lambda} - \lambda\right)^2\right] = Var(\hat{\lambda}) + \left[E(\hat{\lambda}) - \lambda\right]^2$.

$$E(\hat{\lambda}) - \lambda = \frac{\lambda}{2n-1}$$

$$E\left[\left(\frac{1}{Y}\right)^2\right] = \int_0^\infty \frac{1}{y^2} \frac{\lambda^{2n}}{\Gamma(2n)} y^{2n-2-1} e^{-\lambda y} dy$$

$$= \frac{\lambda^{2n}}{\Gamma(2n)} \frac{\Gamma(2n-2)}{\lambda^{2n-2}} \int_0^\infty \frac{\lambda^{2n-2}}{\Gamma(2n-2)} y^{2n-2-1} e^{-\lambda y} dy = \frac{\lambda^2}{(2n-1)(2n-2)}$$

$$\Rightarrow Var(\hat{\lambda}) = \frac{\lambda^2}{(2n-1)(2n-2)} - \frac{\lambda^2}{(2n-1)^2} = \frac{\lambda^2}{(2n-1)^2(2n-2)}$$

$$Var(\hat{\lambda}) = \frac{4n^2\lambda^2}{(2n-1)^2(2n-2)}$$

$$MSE(\hat{\lambda}) = \frac{4n^2\lambda^2}{(2n-1)^2(2n-2)} + \frac{\lambda^2}{(2n-1)^2} = \frac{(2n+2)\lambda^2}{(2n-1)(2n-2)}$$