SAMPLING AND STATISTICAL INFERENCE
SECTION 2

Method of Moments (A SIMPLE METHOD FOR FINDING ESTIMATORS)

Here, we equate population moments to corresponding sample moment — and then solve for the parameters.

One Parameter Case

Usually, equate pop. mean to sample mean (assuming that the pop mean involves the parameter)

e.g. $X_i \sim \text{Poisson}(m)$; sample of size n

$E(X_i) = m$; equate this to $\bar{X}$

$\Rightarrow \hat{m} = \bar{X}$

Sim $X_i \sim \text{Negve Exp}(\lambda)$

$E(X_i) = \frac{1}{\lambda}$, equate this to $\bar{X}$

$\Rightarrow \hat{\lambda} = \frac{1}{\bar{X}}$

Geometric: $P(X = r) = q^{r-1}p$

$E(X) = \frac{1}{p}$

Data: $X_1, X_2, X_3, \ldots$

(repeated observations of # trials to 1st success)

Set $\frac{1}{p} = \bar{X}$

$\hat{p} = \frac{1}{\bar{X}}$ (Sim to negve exp.)

Two Parameter Case

Here we need 2 eqns: Set 1st & 2nd Pop moments = 1st & 2nd Sample mom

1st Moment: $E(X_i) = \bar{X}$ as before

2nd Moment: Could use $EX_i^2 = \frac{\sum X_i^2}{n}$

or. $E(X_i - \mu)^2 = \frac{\sum (X_i - \bar{X})^2}{n}$

EQUIVALENT

example Binomial: $X_i \sim B(n,p)$: OBSERVE $x_1, x_2, \ldots, x_N$

$E(X_i^2)$
$np(1-p)+(np)^2$

$EX_i = np$ Set $= m_1$

$E(X_i^2) = np(1-p) + (np)^2$ Set $= m_2 = \frac{\sum X_i^2}{N}$

$np - \frac{(np)^2}{n} + (np)^2 = m_2$ $\quad m_1 - \frac{m_1^2}{n} + m_1^2 = m_2$

$m_1 - m_2 + m_1^2 = \frac{m_1^2}{n}$

$\Rightarrow \hat{n} = \frac{m_1^2}{m_1 - m_2 + m_1^2}$

EXAMPLE $X_i \sim \text{Gamma}(\alpha, \beta)$

Data $x_1, x_2, \ldots x_n$

Set $E(X_i) = \alpha\beta = \bar{X}$ $\quad M_1'$

and $V(X_i) = \alpha\beta^2 = \dfrac{\sum(X_i - \bar{X})^2}{n} \leftarrow M_2'$

$$\Rightarrow \quad \hat{\beta} = \frac{M_2'}{\bar{X}}$$

$$\Rightarrow \quad \hat{\alpha} = \frac{\bar{X}}{\hat{\beta}} = \frac{(\bar{X})^2}{M_2'}$$

METHOD OF MOMENTS ESTIMATORS

MSE (MEAN SQUARED ERROR)

$$MSE[g(X)] = E[g(X) - \theta]^2$$

Easy to show that

$$MSE = \text{Variance} + (\text{Bias})^2$$

$$MSE = E[g(X) - E(g) + E(g) - \theta]^2$$

$$= E[g(X) - E(g)]^2 + E[E(g) - \theta]^2$$

$$+ 2E[g(X) - E(g)][E(g) - \theta]$$

$$= V[g] + (\text{Bias})^2$$

THE METHOD OF MAXIMUM LIKELIHOOD

This is regarded as a very good method for finding estimators.

We consider random variables $X_1, X_2, \ldots, X_n$ having joint pdf $f(x_1, x_2, \ldots, x_n)$ — or, if discrete, having joint probability function $f(x_1, x_2, \ldots, x_n)$ — which contains one or more parameters θ.

The LIKELIHOOD of θ as a function of $x_1, x_2, \ldots, x_n$ is defined as

$$L(\theta) = f(x_1, x_2, \cdots, x_n ; \theta)$$

Here, we shall confine our consideration to random samples $X_1, X_2, \cdots, X_n$, so that

$$L(\theta) = \prod_{i=1}^{n} f(x_i ; \theta)$$

where $X_1, X_2, \cdots, X_n$ are i.i.d. from a population (or process) with pdf (or probability function — if discrete) $f(x ; \theta)$

NOTE that the likelihood function gives the probability of observing the given values $(x_1, x_2, \cdots, x_n)$, as a function of the parameter θ, in the discrete case — and in the continuous case the likelihood function is proportional to the probability of observing values in the neighbourhood of the given values.

The maximum likelihood estimate (MLE) of θ is that value that maximizes the likelihood — i.e. that would make the observed data the most likely to be observed.

It is usually more convenient to maximize the natural log of the likelihood function — known as the log likelihood.

The process of maximization is often done by differentiation (with respect to θ), setting the derivative(s) to zero, and solving for θ.

<u>Some Examples</u> : A random sample from Poisson(m)

$$L(m) = \prod_{i=1}^{n} \frac{(m)^{X_i} e^{-m}}{(X_i)!}$$

$$\log L(m) = \sum_{i=1}^{n} \left[(X_i \log m) - (m) - \log(X_i!) \right]$$

$$\frac{\partial}{\partial m}\left[\operatorname{Log} L(m)\right] = \frac{1}{m}\sum_{i=1}^{n} X_i - n$$

Setting this to zero:

$$\Longrightarrow \quad \hat{m} = \bar{X}$$

(same as for Method of Moments.)

To check that this $\hat{m}$ yields a maximum of the likelihood function, differentiate again — and check the sign (for $m = \bar{X}$) — turns out to be negative, indicating that this is indeed a maximum.

<u>Continuous case</u> : A random sample from Negve Exp.(λ)

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda X_i}$$

$$\log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} X_i$$

$$\frac{\partial}{\partial \lambda}\left[\log L(\lambda)\right] = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i$$

Setting this to zero:

$$\Longrightarrow \quad \hat{\lambda} = \frac{1}{\bar{X}}$$

(same as for method of moments)

ML Estimation for Two Parameters

There are some important two-parameter distributions such as $N(\mu, \sigma^2)$ and $\text{Gamma}(\alpha, \beta)$.

Consider a sample $(X_1, X_2, \ldots, X_n)$ from $N(\mu, \sigma^2)$

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2}$$

$$\log L(\mu, \sigma^2) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$$

Get the partial derivatives with respect to μ and σ

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^3}$$

Setting $\frac{\partial \log L}{\partial \mu} = 0$, and solving for μ

$$\Longrightarrow \sum_{i=1}^{n}(X_i-\mu) = 0 \Longrightarrow \hat{\mu} = \bar{X}$$

Setting $\frac{\partial \log L}{\partial \sigma} = 0$, and then substituting in for μ,

$$\Longrightarrow \frac{n}{\sigma} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$$

$$\Longrightarrow \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}}$$

(Same as for the Method of Moments)

To check that $(\hat{\mu}, \hat{\sigma})$ maximizes $L(\mu, \sigma^2)$ one could consider the Hessian matrix (of second partial derivatives). We shall not deal with this — but it's been shown that $(\hat{\mu}, \hat{\sigma})$ values do provide a maximum in the two-dimensional likelihood function.

ML Estimation of (α, β) from the Gamma Distribution

We have a sample $(X_1, X_2, \cdots, X_n)$ from Gamma(α, λ)

$$\log L(\alpha, \lambda) = \log\left[\prod_{i=1}^{n}\left\{\frac{1}{\Gamma(\alpha)}(\lambda X_i)^{\alpha-1} e^{-\lambda X_i}\,\lambda\right\}\right]$$

$$= \sum_{i=1}^{n}\left[\alpha \log\lambda - \log\Gamma(\alpha) + (\alpha-1)\log X_i - \lambda X_i\right]$$

$$= n\alpha\log\lambda - n\log\Gamma(\alpha) + (\alpha-1)\sum_{i=1}^{n}\log X_i$$

$$- \lambda\sum_{i=1}^{n} X_i$$

Taking partial derivatives of $\log L$:

$$\frac{\partial \operatorname{Log} L}{\partial\alpha} = n\log\lambda + \sum_{i=1}^{n}\log X_i - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$\frac{\partial \operatorname{Log} L}{\partial\lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} X_i$$

Setting these partial derivatives to zero, the second one

$$\Rightarrow \quad \hat{\lambda} = \frac{n\hat{\alpha}}{\sum X_i} = \frac{\hat{\alpha}}{\bar{X}}$$

We now substitute this into the first equation $\left(\frac{\partial}{\partial\alpha}\log L = 0\right)$

$$n\log\left(\frac{\hat{\alpha}}{\bar{X}}\right) + \sum_{i=1}^{n}\log X_i - n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = 0$$

This is a non-linear equation for $\hat{\alpha}$, and it cannot be solved explicitly (i.e. in closed form).

It is necessary to use numerical (iterative) methods to solve for $\hat{\alpha}$ (and hence to find $\hat{\lambda}$).

The (simpler) method of moments could be used to get a starting value for $\hat{\alpha}$ (to start the iterative method).

INVARIANCE PROPERTY FOR ML ESTIMATORS

IF $\hat{\theta}$ is the MLE for θ, then $g(\hat{\theta})$ is the MLE for $g(\theta)$

e.g. for the case of estimation of μ, σ for data from $N(\mu, \sigma^2)$ we found that

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

We can see that the MLE for σ^2 is $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

APPLICABILITY TO TRUNCATED OR CENSORED DATA

This is a very useful aspect of ML estimation, which will be used in a later course on SURVIVAL ANALYSIS. The data may consist of n_1 observations of a variable X, and a further n_2 observations where we know only that the observation is greater than a value t. Then the likelihood function is

$$L(\theta) = \left[\prod_{i=1}^{n_1} f(x_i, \theta)\right]\left[P(X > t)\right]^{n_2}$$

and the MLE for θ could be found by maximizing $L(\theta)$.

BIASED ESTIMATORS AND MEAN SQUARED ERROR

Remember that if $E(\hat{\theta}) = \theta$, then $\hat{\theta}$ is an unbiased estimator. A measure of performance for a biased estimator is the MEAN SQUARED ERROR (MSE).

$$MSE(\hat{\theta}) = E[\hat{\theta} - \theta]^2$$

THIS IS $\underline{\underline{\text{NOT}}}$ THE VARIANCE OF $\hat{\theta}$

which is the second moment of $\hat{\theta}$ about θ.

Now

$$\begin{aligned} MSE(\hat{\theta}) &= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2 \\ &= E[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 + 2\underbrace{E(\hat{\theta} - E\hat{\theta})}_{0}(E\hat{\theta} - \theta) \\ &= \text{Variance}(\hat{\theta}) + (\text{Bias})^2 \end{aligned}$$

(The bias $= (E\hat{\theta}) - \theta$)

A BIASED ESTIMATOR CAN HAVE SMALLER MSE THAN AN UNBIASED ESTIMATOR.

THE DISTRIBUTION OF AN ML ESTIMATOR FOR LARGE n

(This is referred to as the asymptotic distribution of an MLE.)

Suppose that $\hat{\theta}$ is the MLE for θ

It can be shown that, as $n \to \infty$

(i) $\hat{\theta}$ is unbiased

(ii) $\hat{\theta}$ is Normally distributed

(iii) $V(\hat{\theta}) = \dfrac{-1}{E\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right]}$ or alternatively $\dfrac{1}{E\left[\frac{\partial}{\partial\theta}\log L(\theta)\right]^2}$

The expression given for $V(\hat{\theta})$ is known as the CRAMER-RAO LOWER BOUND (CRLB).

IN FACT, NO UNBIASED ESTIMATOR FOR θ CAN HAVE A SMALLER VARIANCE THAN THE CRLB.

(THE CRLB does not apply when the range of values for the distribution is related to θ, e.g. the uniform distrib $U(0,\theta)$.)

EXAMPLES OF CRLB EVALUATION

We found that for a random sample $X_1, X_2, \dots, X_n$ from a Poisson distribution, with mean m:

The MLE for $m = \bar{X}$

Now we know that $E(\hat{m}) \equiv E(\bar{X}) = m$

and $V(\hat{m}) = V(\bar{X}) = \frac{m}{n}$ $\boxed{\text{HOW DO WE KNOW THIS?}}$

Let's find the CRLB here:

(FROM PAGE 2)

$$\log L(m) = \sum_{i=1}^{n}(X_i \log m) - mn - (\log X_i!)$$

$$\frac{\partial^2 \log L(m)}{\partial m^2} = -\frac{1}{m^2}\sum_{i=1}^{n} X_i$$

So that
$$E\left[\frac{\partial^2 \log L(m)}{\partial m^2}\right] = -\frac{1}{m^2}\sum_{i=1}^{n} E(X_i) = -\frac{n}{m}$$

and thus, the CRLB $= \dfrac{m}{n}$

We can say that the sample mean $\bar{X}$ attains the CR Lower bound for $V(\hat{m})$ in this case.

ANOTHER EXAMPLE: THE CRLB FOR THE ML ESTIMATOR OF σ^2 FOR A SAMPLE FROM $N(\mu, \sigma^2)$

The expression for $\log L(\mu, \sigma^2)$ is on page 3. Let's express it in terms of $\theta = \sigma^2$.

$$\Rightarrow \log L(\mu, \theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(X_i - \mu)^2$$

$$\frac{\partial}{\partial\theta}\left[\log L(\mu, \theta)\right] = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(X_i - \mu)^2$$

$$\frac{\partial^2}{\partial\theta^2}\left[\log L(\mu, \theta)\right] = \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^{n}(X_i - \mu)^2$$

$$= \frac{n}{2\theta^2} - \frac{1}{\theta^2}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

$$E\left[\frac{\partial^2}{\partial\theta^2}\log L(\mu, \theta)\right] = \frac{n}{2\theta^2} - \frac{n}{\theta^2}$$

$$= -\frac{n}{2\theta^2}$$

Thus the CRLB for estimation of σ^2 from a random sample from $N(\mu, \sigma^2)$:

$$= \frac{2\theta^2}{n} = \frac{2\sigma^4}{n}$$

Thus for any unbiased estimator $\hat{\theta}$ for σ^2,

$$V(\hat{\theta}) \geq \frac{2\sigma^4}{n}$$

EXAMPLE: A random sample $\{x_i\}$ from the pdf $f(x) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}$ for $x \geq 0$
$= 0$, else.

$$\log L(x_1, x_2, \dots x_n) = \sum_1^n \log x_i - n\log\theta - \frac{\sum_1^n x_i^2}{2\theta}$$

$$\frac{\partial}{\partial\theta}\log L = -\frac{n}{\theta} + \frac{\sum_1^n x_i^2}{2\theta^2}$$

Setting to zero,

$$\Longrightarrow \quad \sum_1^n x_i^2 = 2n\theta$$

and

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i^2}{2n}$$

$$E(\hat{\theta}) = \frac{\sum_{i=1}^n \int_0^\infty x^2 \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} dx}{2n}$$

using the hint, introduce $u = \frac{x^2}{2\theta}$ so that $du = \frac{x}{\theta}dx$
and $x^2 = 2\theta u$

$$E(\hat{\theta}) = \frac{\sum_{i=1}^n \int_0^\infty 2\theta u e^{-u} du}{2n}$$

$$= \frac{\sum_{i=1}^n 2\theta \Gamma(1)}{2n} \qquad \Gamma(1) = 1$$

$$= \theta$$

CRLB $= \dfrac{-1}{E\left[\frac{\partial^2}{\partial\theta^2}\log L\right]}$

$$\frac{\partial^2}{\partial\theta^2}\log L = \frac{n}{\theta^2} - \frac{\sum_1^n x_i^2}{\theta^3}$$

$$E\left[\frac{\partial^2}{\partial\theta^2}\log L\right] = \frac{n}{\theta^2} - \frac{\sum_1^n}{\theta^3}\int_0^\infty x^2 \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} dx$$

$$= \frac{n}{\theta^2} - \frac{\sum_1^n 2\theta}{\theta^3}$$

$$= -\frac{n}{\theta^2}$$

$$\Rightarrow \text{CRLB} = \theta^2/n$$