Chapter 12

Maximum Likelihood Estimation

12.1 Point estimators and estimates

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from a distribution $f(X_1, X_2, \ldots, X_n; \theta)$ depending on a parameter θ. We wish to 'guess' the value of θ using the experimental (observed) values $(x_1, x_2, \ldots, x_n)$ of the sample $\{X_1, X_2, \ldots, X_n\}$. In order to do that we define a statistic $\hat{\theta} = t(X_1, X_2, \ldots, X_n)$ such that its distribution is 'massed' around the true value θ. Then the observed value of $\hat{\theta}$, i.e. $\hat{\theta} = t(x_1, x_2, \ldots, x_n)$ will generally be close to θ.

Note that $\{X_1, X_2, \ldots, X_n\}$ is a vector of *n* **random variables** while $\{x_1, x_2, \ldots, x_n\}$ is a particular sample from this random vector. The statistic $\hat{\theta} = t(X_1, X_2, \ldots, X_n)$ is an **estimator** of unknown parameter θ. Given the observed sample $(x_1, x_2, \ldots, x_n)$, $\hat{\theta} = t(x_1, x_2, \ldots, x_n)$ takes a specific value. That is $\hat{\theta} = t(x_1, x_2, \ldots, x_n)$ is an observed value of the statistic $\hat{\theta} = t(X_1, X_2, \ldots, X_n)$; it is called a **point estimate** of θ.

Example:

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from a $N(\mu, \sigma^2)$ distribution. Suppose we wish to estimate μ. What estimators could we use? Try the estimator $\hat{\mu} = \overline{X}$. The distribution of $\hat{\mu} = \overline{X}$ is $N\left(\mu, \frac{\sigma^2}{n}\right)$. Thus the middle of the sampling distribution is μ and the sampling distribution has a variance which gets smaller as *n* gets bigger. Hence we expect $\overline{X}$ to be close to the true population mean μ, especially as *n* gets bigger.

Suppose a random sample $\{X_1, X_2, X_3\}$ of size *n*=3 is taken i.e. $(x_1, x_2, x_3) = (4.2, 3.7, 5.0)$. Then $\hat{\mu} = \overline{X}$ is an **estimator** of μ and the observed sample mean $\bar{x} = 4.3$ is a **point estimate** of μ.

We have estimators (which are r.v.'s) and estimates which are observed values (numbers) of the estimators. The estimators are r.v.'s so they have sampling distribution, expectations variance etc. However, in practice we often talk about the expectation and variance, etc, of the estimates (when we understand that we are really

thinking about the underlying r.v.'s). You may note that this is what we did in chapter 1, where we used small *y*'s for r.v.'s.

12.2 Desirable properties of estimators

12.2.1 Unbiasedness

***Definition 1*:** An estimator $\hat{\theta}$ of θ is an unbiased estimator if $E(\hat{\theta}) = \theta$.

Definition 2: The bias of an estimator $\hat{\theta}$ of θ is given by $\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$.

12.2.2 Efficiency

Definition 3: The **mean square error** (m.s.e.) of an estimator $\hat{\theta}$ of θ is $E\left[(\hat{\theta}-\theta)^2\right]$ i.e. the m.s.e. is the average squared distance from $\hat{\theta}$ to the true parameter θ.

Theorem 1: $\text{m.s.e.}(\hat{\theta}) = V(\hat{\theta}) + \left[\text{bias}(\hat{\theta})\right]^2$

Proof:

$$\begin{aligned}
\text{m.s.e.}(\hat{\theta}) &= E\left[(\hat{\theta}-\theta)^2\right] \\
&= E\left[(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2\right] \\
&= E\left[(\hat{\theta}-E(\hat{\theta}))^2\right] + E\left[(E(\hat{\theta})-\theta)^2\right] + 2E\left[(\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\theta)\right] \\
&= V(\hat{\theta}) + \left[\text{bias}(\hat{\theta})\right]^2 + 2\left[E(\hat{\theta})-\theta\right]E\left[\hat{\theta}-E(\hat{\theta})\right] \\
&= V(\hat{\theta}) + \left[\text{bias}(\hat{\theta})\right]^2 \qquad \text{since } E\left[\hat{\theta}-E(\hat{\theta})\right] = 0
\end{aligned}$$

Definition 4: $\hat{\theta}_1$ is a more efficient estimator than $\hat{\theta}_2$ for parameter θ if $\text{m.s.e.}(\hat{\theta}_1) < \text{m.s.e.}(\hat{\theta}_2)$.

Hence if we have to choose one of several estimators, we would generally like to choose the one with the smallest m.s.e. If we consider only unbiased estimators then we would make efficiency comparisons in terms of their variances since, for an unbiased estimator $\hat{\theta}$, $\text{m.s.e.}(\hat{\theta}) = V(\hat{\theta})$.

There are several methods for choosing an estimator. In the next section we will describe the most important one called the method of maximum likelihood.

12.3 Maximum Likelihood Estimation

Definition 5: Let θ be the unknown parameter of the distribution of interest $f(x_i;\theta)$. The **likelihood function,** $L(\theta / x_1, x_2, \ldots, x_n)$ of the realisation $x_1, x_2, \ldots, x_n$ of n random variables $X_1, X_2, \ldots, X_n$ is defined to be the joint density of the n random variables, considered as a function of the parameter θ, i.e.

$$L(\theta / x_1, x_2, \ldots, x_n) = \begin{cases} f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta) & \text{continuous} \\ P(X_1 = x_1 \cap X_2 = x_2 \cap \ldots \cap X_n; \theta) & \text{discrete} \end{cases}$$

Note that in the above definition the *x*'s are treated as fixed while the variable is the parameter θ.

If we have an independent random sample $\{X_1, X_2, \ldots, X_n\}$, (that is the underlying *X*'s are coming from the same distribution $f(X_j;\theta)$), then the Likelihood function of an observed sample $x_1, x_2, \ldots, x_n$ has the form

$$L(\theta / x_1, x_2, \ldots, x_n) = \begin{cases} \prod_{i=1}^{n} f_X(x_i;\theta) & \text{continuous} \\ \prod_{i=1}^{n} P(X = x_i;\theta) & \text{discrete} \end{cases}$$

Definition 6: The **maximum likelihood estimator** (**M.L.E**.) $\hat{\theta}$ of the parameter θ is the value of θ which maximises $L = L(\theta / x_1, x_2, \ldots, x_n)$.

Note that in general the ML.E. $\hat{\theta}$ is a function of the sample $\{X_1, X_2, \ldots, X_n\}$ i.e. $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$. That is $\hat{\theta}$ is a statistic having its own distribution. Given an observed sample $\{x_1, x_2, \ldots, x_n\}$ the quantity $\theta = \theta(x_1, x_2, \ldots, x_n)$ is called the **maximum likelihood estimate** for θ.

In order to find the M.L.E. for θ we would have to differentiate the likelihood function with respect to θ and set it to zero. Usually it is easier to work with the log of the Likelihood function since both functions have the same maximum, (see below).

Definition 7: The log-likelihood function is defined as $l(\theta) = \log L(\theta / x_1, x_2, \ldots, x_n)$

Now $$\frac{d \log L}{d\theta} = \frac{dl}{d\theta} = \frac{1}{L}\frac{dL}{d\theta}$$

$$\therefore \frac{dL}{d\theta} = 0 \Leftrightarrow \frac{dl}{d\theta} = 0.$$

That is the likelihood and the log-likelihood functions have the some maximum.

It is usually easier to set $\frac{dl}{d\theta} = 0$.

Note that any quantity in the distribution $f(x_i;\theta)$ which does not involve the parameter θ is not needed when we try to find the M.L.E for θ. This leads to an equivalent definition of the Likelihood functions.

Definition 5a: The likelihood function is proportional to the p.d.f. of the sample

In the case of independent random sample we have

$$L(\theta / x_1, x_2, \ldots, x_n) \propto \begin{cases} \prod_{i=1}^{n} f_X(x_i;\theta) & \text{continuous} \\ \prod_{i=1}^{n} P(X = x_i;\theta) & \text{discrete} \end{cases}$$

Discrete Example

Suppose we toss a coin 10 times and get exactly 3 heads and we wish to estimate p the probability of getting a head on each toss of the coin. Let X = number of heads from 10 tosses, Then $X \sim B(10,p)$ and

$$P(X = 3) = {}^{10}C_3\, p^3 (1-p)^7, \quad 0 \le p \le 1$$

The Likelihood function for p is

$$L(p) \propto p^3(1-p)^7$$

and the log likelihood is

$$l(p) = \log L(p) = 3\log p + 7\log(1-p) + \textit{Constant not involving } p$$

Both functions are shown in figures 12.1 and 12.2 respectively.

To maximise $L \propto P(X = 3)$ with respect to p, set $\frac{dL}{dp} = 0$.

$$\frac{dl}{dp} = 0 \Rightarrow \hat{p}_{M.L.E.} = \frac{3}{10}.$$

We have plotted the probability of our data ($L \propto P(X = 3)$) against possible values of the parameter p and found the value $\hat{p}_{M.L.E.}$ of p which maximises L. i.e. $\hat{p}_{M.L.E.}$ is the value of p which makes the data we actually obtained (i.e. $X = 3$) most likely.

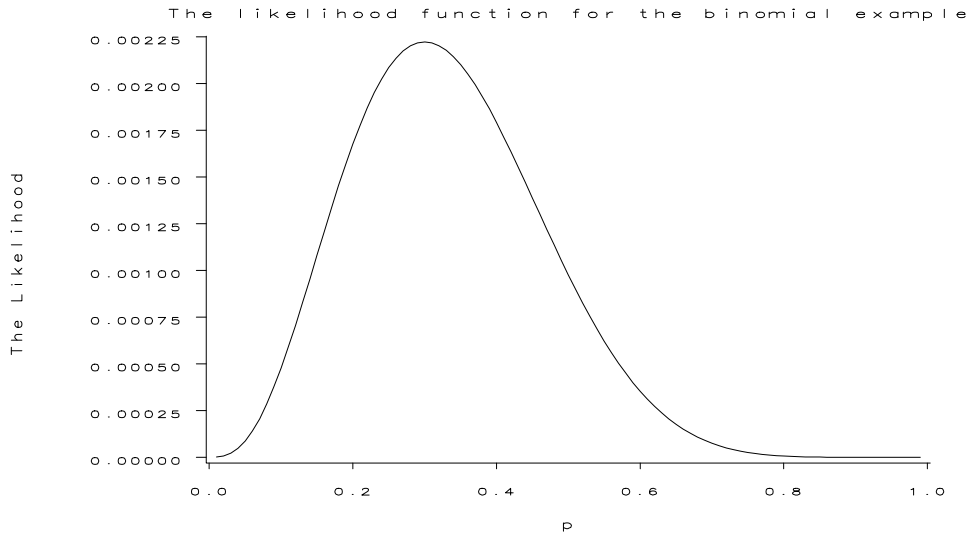

Figure 12.1: Showing the Likelihood function for the discrete binomial example.

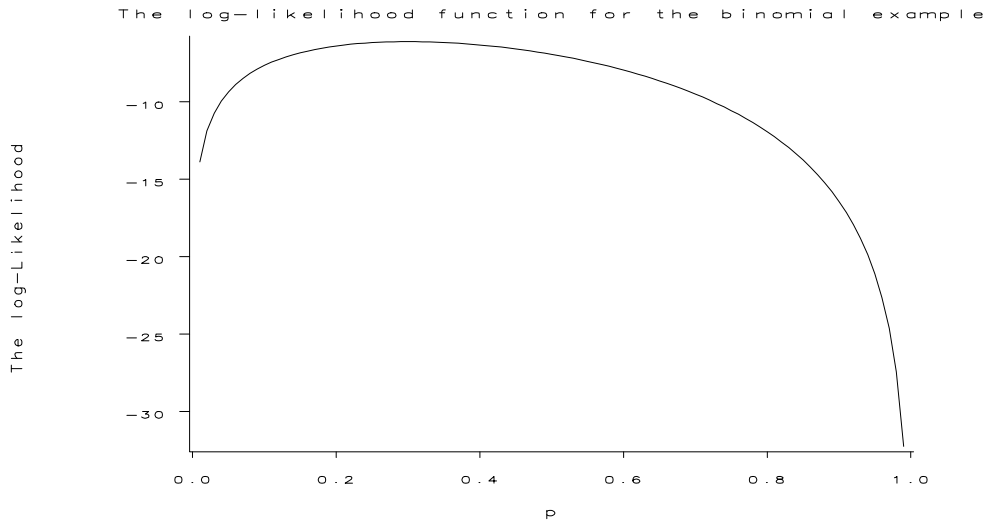

Figure 12.2: Showing the log-likelihood for the discrete binomial example.

Generalisation of discrete example

Suppose now n people each toss the same coin 10 times and each count the number of heads $x_1, x_2, \ldots, x_n$, then the probability of getting this data is $P(X_1 = x_1 \cap X_2 = x_2 \cap \ldots \cap X_n) = \prod_{i=1}^{n} P(X_i = x_i)$ and since $X_i \sim B(10, p)$ for $i = 1, 2, \ldots, 10$ and all X_i's are independent we have

$$L(p) = \prod_{i=1}^{n} {}^{10}C_{x_i} p^{x_i} (1-p)^{10-x_i} \propto p^{\sum x_i} (1-p)^{\sum (10-x_i)}$$

The log likelihood is given by

$$l(p) = \sum x_i \log p + \sum (10 - x_i) \log(1-p) + \textit{Constant}$$

To find the M.L.E

$$\frac{dl}{dp} = \frac{\sum x_i}{p} - \frac{\sum (10 - x_i)}{1-p}$$

$$\frac{dl}{dp} = 0 \Rightarrow \frac{\sum x_i}{\hat{p}} = \frac{\sum (10 - x_i)}{1-\hat{p}}$$

$$\Rightarrow \hat{p}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} 10} = \frac{\sum_{i=1}^{n} x_i}{10n} = \frac{\bar{x}}{10}$$

e.g. if $(x_1, x_2, \ldots, x_5) = (3,5,2,1,6)$ then $\hat{p}_{MLE} = \frac{3+5+2+1+6}{50} = 0.34$

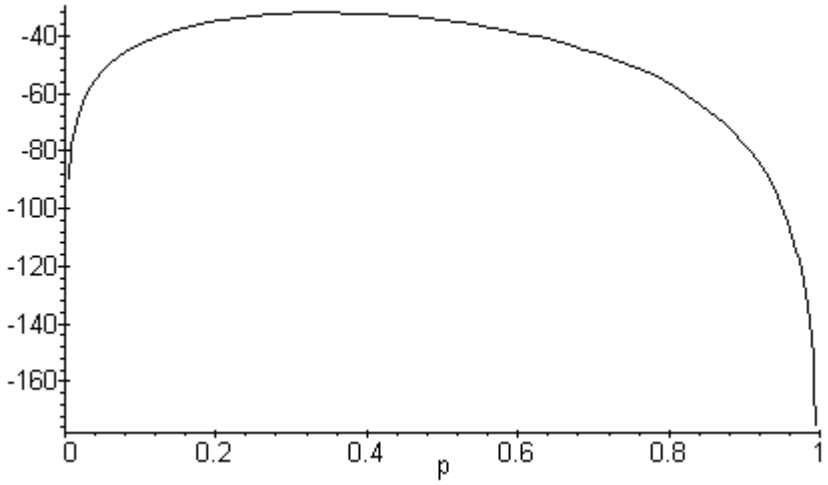

Figure 12.3: Showing the log-likelihood for the above discrete binomial example.

Continuous Example

Let $\{X_1, X_2, ..., X_n\}$ be a random sample from an $Ex(\lambda)$ distribution. Find the M.L.E. of λ. The likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{\lambda \sum x_i}$$

The log-likelihood is

$$l(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i$$

To find the M.L.E. of λ set

$$\frac{dl}{d\lambda} = 0 \Rightarrow \frac{n}{\hat{\lambda}} = \sum_{i=1}^{n} x_i \Rightarrow \hat{\lambda}_{MLE} = \frac{1}{\bar{x}}$$

Hence the **maximum likelihood estimate** of λ is $\frac{1}{\bar{x}}$.

The maximum likelihood estimator is obtained by replacing the x_i by the corresponding random variable X_i hence the **maximum likelihood estimator** of λ is $\frac{1}{\bar{X}}$.

Check also that $\frac{d^2l}{d\lambda^2}$ is negative at $\lambda = \frac{1}{\bar{x}}$.

$$\frac{d^2l}{d\lambda^2} = -\frac{n}{\lambda^2} < 0 .$$

e.g. if $(x_1, x_2, .., x_5) = (3.1, 5.5, 2.3, 1.9, 6.3)$ then $\bar{x} = \frac{3.1 + 5.5 + 2.3 + 1.9 + 6.3}{5} = 3.82$ so

$\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{3.82} = 0.2617$. (see also figure 12.4 for the log likelihood).

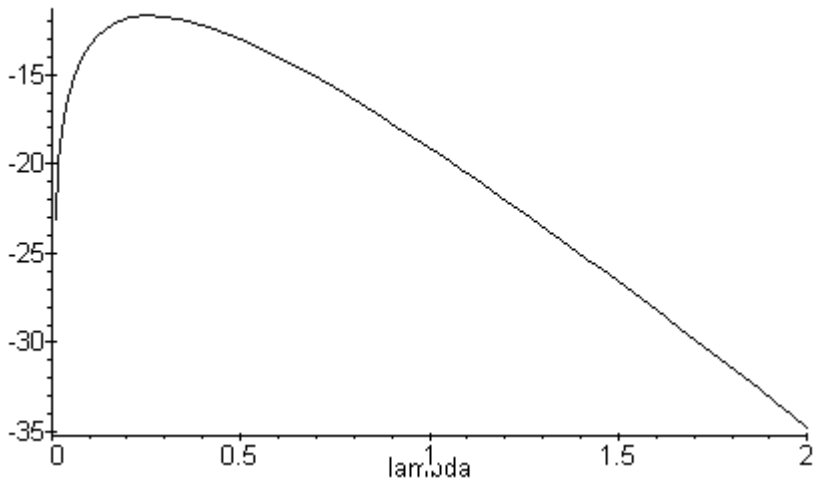

Figure 12.4: Showing the log-likelihood for the above continuous exponential distribution example.

Maximum Likelihood Estimation when the range depends on a parameter

Let be a random sample from a Uniform $(0,\theta)$ distribution with p.d.f.

$$f(x) = \frac{1}{\theta} = \frac{hv(\theta - x)}{\theta} \qquad 0 \le x \le \theta, \qquad \theta > 0$$

where hv(z) is the heaviside function defined as 1 if z>0 and zero elsewhere

The likelihood for n observations is

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \quad \text{for } 0 \le x_1 \le \theta, \quad 0 \le x_2 \le \theta \quad \ldots \quad 0 \le x_n \le \theta$$

and zero if any x_i is outside $[0,\theta]$. This can be written as

$$\mathrm{L}(\theta) = \frac{\mathrm{hv}(\theta - \mathrm{x}_1)}{\theta} \cdots \frac{\mathrm{hv}(\theta - \mathrm{x}_\mathrm{n})}{\theta} = \frac{\mathrm{hv}(\theta - \max(\mathrm{x}))}{\theta}$$

The likelihood function with n=4 and max(x)=20 is shown in the figure 4.5 below.

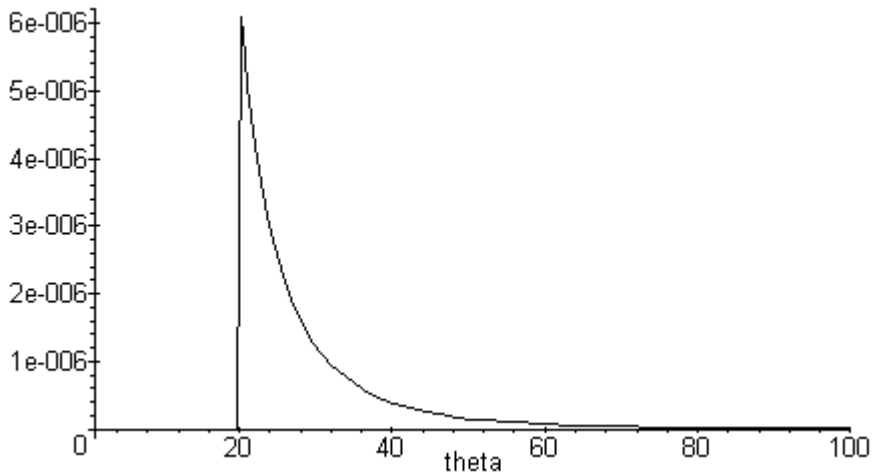

Figure 4.5: Showing the log-likelihood for the above uniform distribution example.

It can be seen that the likelihood decreases when $\theta \to \infty$ and that the likelihood is zero if θ is less that 20 (i.e. zero if less than the maximum of x_i). Hence L is at a maximum when θ is equal to the largest value of x_i, i.e. $\hat{\theta} = \max(x_i)$. The likelihood

is discontinuous at the maximum point $\hat{\theta}$; therefore the likelihood derivative does not exist there. That is the M.L.E is not a solution to $\frac{\partial \mathrm{L}}{\partial \theta} = 0$.

Important: whenever the range of *X* depends on the parameter it is important to see whether the maximum of the likelihood is achieved at an extreme rather than at a solution of $\frac{\partial \mathrm{L}}{\partial \theta} = 0$.

12.4 Optimum properties of maximum likelihood estimators

12.4.1 The invariance property

Let $\hat{\theta}$ be the M.L.E. of then the M.L.E. of $\phi = g(\theta)$ is $\hat{\phi} = g(\hat{\theta})$.

Example:

For a random sample from $N\left(\mu,\sigma^2\right)$, the M.L.E. of μ is $\overline{X}$. By the invariance property, the M.L.E. of $\phi = \frac{1}{\mu}$ is $\hat{\phi} = \frac{1}{\overline{X}}$. Similarly, the M.L.E. of μ^2 is $\overline{X}^2$

12.4.2. Asymptotic Normality

Let $\{X_1, X_2, ..., X_n\}$ be a random sample from a distribution with p.d.f. $f(x;\theta)$ depending on an unknown parameter θ. If $\hat{\theta}$ is the M.L.E. of θ then under certain regularity conditions, $\hat{\theta}$ is asymptotically Normally distributed with mean θ and variance $\frac{1}{E\left[\left(\frac{d \log L}{d\theta}\right)^2\right]}$.

i.e. distribution of $\hat{\theta} \rightarrow N\left(\theta, \frac{1}{i(\theta)}\right)$ as $n \rightarrow \infty$.

where $i(\theta) = E\left[\left(\frac{d \log L}{d\theta}\right)^2\right] = -E\left[\frac{d^2 \log L}{d\theta^2}\right]$ is called the **expected information** for θ (it is defined under certain regularity conditions).

The quantity $\frac{1}{i(\theta)}$ is the Cramer-Rao lower bound for the variance of unbiased estimators of θ.

Alternatively the quantity $I(\theta) = -\left[\frac{d^2 \log L}{d\theta^2}\right]$ can be used instead of $i(\theta)$ in the above approximation. $I(\theta)$ is called the **information** for θ.

Note that in general the evaluation of both $i(\theta)$ and $I(\theta)$ depends on the parameter θ which is usually unknown. An estimate of the information or the expected information can be obtained by replacing θ by its M.L. estimate $\hat{\theta}$. The quantities $i(\hat{\theta})$ and $I(\hat{\theta})$ are then called the **observed expected information** and the **observed information** respectively.

If $g(.)$ is differentiable, then $g\left(\hat{\theta}\right) \overset{asymptotically}{\sim} N\left(g(\theta), \left[\frac{dg(\theta)}{d\theta}\right]^2 \Big/ i(\theta)\right)$

Example

The Exponential distribution p.d.f. is given by

$$f(x) = \theta e^{-\theta x} \quad 0 < x < \infty$$

Let us assume that we have a random sample of size n from an exponential distribution. The Likelihood function and the log-likelihood are

$$L(\theta) = \prod_{i=1}^{n} f\left(x_i\right) = \prod_{i=1}^{n} \theta\, e^{-\theta x_i}$$
$$= \theta^n e^{-\theta \sum x_i}$$

and

$$l(\theta) = \log L(\theta) = n \log \theta - \theta \sum_{i=1}^{n} x_i \ .$$

By differentiating the log-likelihood and setting it to zero we have

$$\frac{d \log L}{d\theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$

$$\frac{d \log L}{d\theta} = 0 \Rightarrow \hat{\theta} = \frac{1}{\overline{X}}$$

From the asymptotic normality property we have,

$$\hat{\theta} \overset{asymptotically}{\sim} N\left(\theta, \frac{1}{i(\theta)}\right)$$

where $i(\theta) = -E\left[\frac{d^2 \log L}{d\theta^2}\right] = -E\left[-\frac{n}{\theta^2}\right] = \frac{n}{\theta^2} = I(\theta)$

That is, in this case $i(\theta)$ and $I(\theta)$ are the identical since the second derivative does not involve the random variables, X_i.

$$\therefore \hat{\theta} \overset{asymptotically}{\sim} N\left(\theta, \frac{\theta^2}{n}\right)$$

Hence we can construct an approximate $100(1-\alpha)\%$ C.I. for θ. We have $\hat{\theta} \overset{approx.}{\sim} N\left(\theta, \frac{\theta^2}{n}\right)$ for n sufficiently large so $Z = \frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta^2}{n}}} \overset{approx.}{\sim} N(0,1)$

We require

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

i.e. $$P\left(-z_{\frac{\alpha}{2}} < \frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta^2}{n}}} < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\sqrt{n}\left(\hat{\theta} - \theta\right)}{\theta} < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(-z_{\frac{\alpha}{2}} < \sqrt{n}\left(\frac{\hat{\theta}}{\theta} - 1\right) < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} < \frac{\hat{\theta}}{\theta} < 1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}\right) = 1-\alpha$$

$$P\left(\frac{\hat{\theta}}{1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}} < \theta < \frac{\hat{\theta}}{1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}\right) = 1-\alpha$$

Thus an approximate $100(1-\alpha)\%$ C.I. for θ is $\left(\frac{\hat{\theta}}{1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}, \frac{\hat{\theta}}{1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}\right)$.

12.5 Maximum Likelihood Estimation of k parameters

If the likelihood function contains k parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, then the maximum likelihood estimators of the parameters are obtained by solving the k equations

$$\frac{\partial \log L}{\partial \theta_1} = 0, \quad \frac{\partial \log L}{\partial \theta_2} = 0, \quad \ldots \quad \frac{\partial \log L}{\partial \theta_k} = 0, \qquad \text{simultaneously.}$$

Also

$$\mathbf{i}(\theta) = E\left[\left(\frac{\partial \log L}{\partial \theta_i}\right)\left(\frac{\partial \log L}{\partial \theta_j}\right)\right]_{ij} = \left\{-E\left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right]\right\}_{ij}$$

is the *k* x *k* **expected information matrix** for the parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ and

$$\mathbf{I}(\theta) = \left\{-\left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right]\right\}_{ij}$$

is the *k* x *k* **information matrix.**

Example: M.L.E. of 2 parameters

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample from a Normal distribution with unknown mean μ and variance σ^2. Find the M.L.E. of μ and σ^2.

The likelihood function is

$$L(\mu,\sigma^2)=\prod_{i=1}^{n} f(x_i;\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{(2\pi\sigma^2)^{1/2}}e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}$$
$$=(2\pi\sigma^2)^{-n/2}e^{-\frac{1}{2}\frac{\sum(x_i-\mu)^2}{\sigma^2}}$$

By dropping the constant $(2\pi)^{-\frac{n}{2}}$ and letting $\theta=\sigma^2$ the log-likelihood is

$$l(\mu,\theta)=-\frac{n}{2}\log(\theta)-\frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2$$

Differentiate the log-likelihood with respect to μ and θ we have

$$\frac{\partial l}{\partial \mu}=\frac{\sum_{i=1}^{n}(x_i-\hat{\mu})}{\hat{\theta}}=0$$

$$\frac{\partial l}{\partial \theta}=-\frac{n}{\hat{\theta}}+\frac{\sum_{i=1}^{n}(x_i-\hat{\mu})^2}{\hat{\theta}^2}=0$$

resulting to the MLE's

$$\hat{\mu}=\bar{x}=\sum_{i=1}^{n}x_i/n \qquad \text{(the usual sample mean)}$$

and

$$\hat{\theta}=\hat{\sigma}^2=\sum_{i=1}^{n}(x_i-\hat{\mu})^2/n=\sum_{i=1}^{n}(x_i-\bar{x})^2/n$$

This is not the usual sample variance which is defined by

$$s^2=\sum_{i=1}^{n}(x_i-\bar{x})^2/n-1.$$

The M.L.E $\hat{\sigma}^2$ is a biased estimator for σ^2 since $E(\sum_{i=1}^{n}(x_i-\bar{x})^2)=(n-1)\sigma^2$ while s^2 is unbiased.

Example:

Let the sample $(x_1, x_2, \ldots, x_{10}) = (11.3, 8, 15.5, 6.8, 13.6, 16.7, 9.9, 20.7, 11.3, 9.2)$

We have

$$\hat{\mu} = \bar{x} = 12.3$$

and

$$\hat{\sigma}^2 = 168/10 = 16.8.$$

The usual sample variance is $s^2 = 168/9 = 18.66$.

The log-likelihood function for the above data in terms of μ and σ^2 is shown below in figure 12.6. The log-likelihood function has quadratic shape for the parameter μ and is very flat in terms of σ^2.

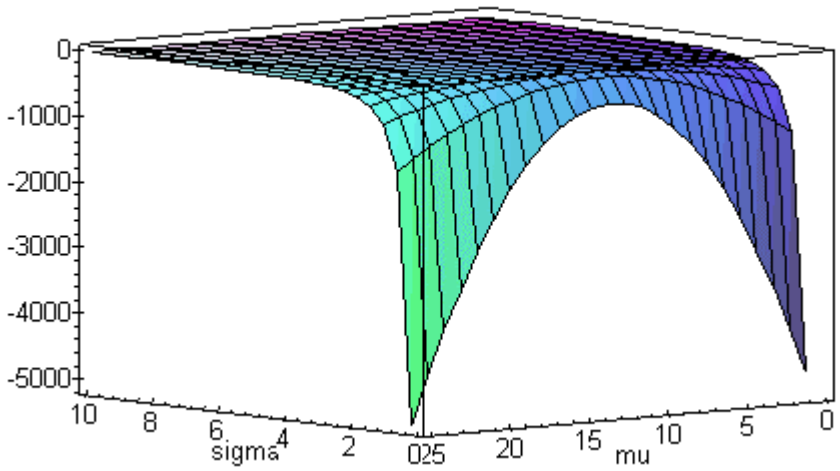

Figure 12.6: Showing the log-likelihood function from the above Normal distribution example.

Exercise 12.1

1. A manufacturing process produces fibres of varying lengths. The length of a fibre is a continuous variable with p.d.f.

$$f(x) = \theta^{-2}\, x\, e^{-x/\theta} \qquad \text{for} \qquad x > 0$$

where $\theta > 0$ is an unknown parameter. Suppose that *n* randomly selected fibres have lengths $x_1, x_2, \ldots, x_n$. Find expressions for the MLE for θ.

2. Suppose that $x_1, x_2, \ldots, x_n$ are independent values from a Normal distribution $N(\mu, 1)$. Find the MLE of μ.

3. Suppose that $x_1, x_2, \ldots, x_n$ are independent values from Normal distribution $N(0, \sigma^2)$. Find the MLE of σ. Is the estimate unbiased?

Exercise 12.2

a) Let X_i be a random variable having a binomial distribution with probability density function given by

$$f(x_i) = \binom{n_i}{x_i} p^{x_i} (1-p)^{n_i - x_i} \quad \text{for } x_i = 0,1,\ldots,n_i$$

where and $0 < p < 1$.

Let also $X_1, X_2, \ldots, X_k$ be an independent random sample of observations from the above distribution

i) Write down the Likelihood function $L(p; x_1 \ldots x_k)$ and show that the log likelihood function $l(p; x_1 \ldots x_k)$ for the parameter p is given by:

$$l(p; x_1 \ldots x_k) = \sum_{i=1}^{k} x_i \log(p) + \sum_{i=1}^{k} (n_i - x_i) \log(1-p)$$

ii) Find the maximum likelihood estimator of p.

iii) Show that the Expected information for the parameter p is given by

$$i(p) = -E\left[\frac{d^2 \log L}{dp^2}\right] = \frac{\sum_{i=1}^{k} n_i}{p} + \frac{\sum_{i=1}^{k} n_i}{(1-p)}$$

(Note that $E(X_i) = n_i p$)

b) The following data concern teenage pregnancy in k=13 north central Florida counties during the 3 year period 1989-91. Let n_i denote the total number of births in county i during the reporting period and let x_i denote the numbers involving underage mothers (i.e. women 17 or under).

	x_i	n_i
1	91	2848
2	16	344
3	36	1617
4	34	688
5	7	160
6	2	133
7	13	171
8	2	66
9	10	360
10	80	2753
11	43	982
12	12	351
13	7	135

Assuming that the rate (p) of teenage pregnancies is the same in all the 13 Florida counties calculate the Maximum Likelihood estimator for p and give an asymptotic 95% Confidence Interval for p by using its Observed Expected information.

Exercise 12.3

a) Let X_i be a random variable having a Poisson distribution with probability density function given by

$$f(x_i) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i !} \qquad \text{for } x_i = 0,1,...$$

where $0 < \lambda < \infty$. Let also $x_1, x_2, ..., x_n$ be an independent random sample of observations from the above distribution.

i) Write down the Likelihood function $L(\lambda; x_1 ... x_n)$ and show that the log likelihood function $\ell(\lambda; x_1 ... x_n)$ for the parameter λ is given by:

$$\ell(\lambda; x_1 ... x_n) = \sum_{i=1}^{n} x_i \, log(\lambda) - n\lambda - \sum_{i=1}^{n} log(x_i !)$$

ii) Find the maximum likelihood estimator of λ

iii) Show that the Expected information for the parameter λ is given by

$$i(\lambda) = -E\left[\frac{d^2 \ell}{d\lambda^2}\right] = \frac{n}{\lambda}$$

(Note $E(X_i) = \lambda$. You do not need to prove this.)

b) Ten families were randomly selected in a certain district, with the aim to determine the average number of children per family in the district. The numbers children for the ten families was as follows

2, 4, 0, 1, 1, 3, 6, 0, 1, 2

Assuming that the number of children per family follows a Poisson distribution with mean λ, calculate the Maximum Likelihood estimator for λ and give an asymptotic 95% Confidence Interval for λ.