CHAPTER 6

STATISTICAL ANALYSIS

6.1 STATISTICAL SUMMARIES & SAMPLING DISTRIBUTIONS

Definition 6.1.1 A collection of i.i.d. random variables $X_1, ..., X_n$ each of which has distribution defined by cdf F_X (or mass/density function f_X) is a **random sample** of size n from F_X (or f_X).

Definition 6.1.2 A function, T, of a random sample, $X_1, ..., X_n$, that is, $T = t(X_1, ..., X_n)$ that depends only on $X_1, ..., X_n$ is a **statistic**. A statistic is a random variable.
For example, the sample mean $\bar{X} = (X_1 + X_2 + ... + X_n)/n$ is a statistic.

Definition 6.1.3 If $X_1, ..., X_n$ is a random sample from F_X, say, and $T = t(X_1, ..., X_n)$ is a statistic, then F_T (or f_T), the cdf (or mass/density function) of random variable T, is the **sampling distribution** of T.

The objective is to derive the distribution of T from the distribution of $X_1, ..., X_n$.

EXAMPLE: If $X_1, ..., X_n$ are independent random variables, with $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, ..., n$, and $a_1, ..., a_n$ are constants, consider the distribution of random variable Y defined by

$$Y = \sum_{i=1}^{n} a_i X_i$$

Using standard mgf results, the distribution of Y is derived to be normal with parameters

$$\mu_Y = \sum_{i=1}^{n} a_i \mu_i \qquad \sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2.$$

Now consider the special case of this result when $X_1, ..., X_n$ are i.i.d. with $\mu_i = \mu$ and $\sigma_i^2 = \sigma^2$, and where $a_i = 1/n$ for $i = 1, ..., n$. Then

$$Y = \sum_{i=1}^{n} \frac{1}{n} X_i = \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Definition 6.1.4 For a random sample $X_1, ..., X_n$ from a probability distribution, then the **sample variances**, s^2 and S^2, are statistics defined by

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \qquad\qquad S^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

Theorem 6.1.1 *Suppose that $X_1, ..., X_n$ is a random sample from a normal distribution, say $X_i \sim N(\mu, \sigma^2)$. Then*

(1) $\bar{X}$ is independent of $\{X_i - \overline{X}, i = 1, ..., n\}$
(2) $\bar{X}$ and s^2 are independent random variables
(3) The random variable

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

*has a **chi-squared distribution** with $n-1$ degrees of freedom.*

Proof. (1) Sketch: Consider the multivariate transformation to $Y_1, ..., Y_n$ where

$$Y_1 = \bar{X} \qquad Y_i = X_i - \bar{X}, i = 2, ..., n \Longleftrightarrow X_1 = Y_1 - \sum_{i=1}^{n} Y_i \qquad X_i = Y_i + Y_1.$$

Then $f_{Y_1,..,Y_n}(y_1, .., y_n) = f_{Y_2,..,Y_n}(y_2, .., y_n) f_{Y_1}(y_1)$, and therefore Y_1 is independent of $Y_2, ..., Y_n$. Hence $\bar{X}$ is independent of the random variables terms $\{Y_i = X_i - \bar{X}, i = 2, ..., n\}$. Finally, as $X_1 - \bar{X} = -\sum_{i=2}^{n} \left(X_i - \bar{X}\right)$, $\bar{X}$ is also independent of $X_1 - \bar{X}$

(2) s^2 is a function only of $\{X_i - \bar{X}, i = 1, ..., n\}$. As $\bar{X}$ is independent of these variables, $\bar{X}$ and s^2 are also independent.

(3) Can decompose the sums of squares terms that appear in the likelihood as

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 + n\left(\bar{X} - \mu\right)^2 \qquad \text{or} \qquad \sigma^2 V_1 = \sigma^2 V_2 + \sigma^2 V_3$$

say, where the mgfs of V_1, V_2 and V_3 are

$$M_{V_1}(t) = \left(\frac{1/2}{1/2 - t}\right)^{n/2} \qquad M_{V_2}(t) = \left(\frac{1/2}{1/2 - t}\right)^{(n-1)/2} \qquad M_{V_3}(t) = \left(\frac{1/2}{1/2 - t}\right)^{1/2}$$

so that

$$V_1 \sim \chi^2_n \qquad V_2 \sim \chi^2_{n-1} \qquad V_3 \sim \chi^2_1$$

Theorem 6.1.2 *Suppose that $X_1, ..., X_n$ is a random sample from a normal distribution, say $X_i \sim N(\mu, \sigma^2)$. Then the random variable*

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

has a Student-t distribution with $n-1$ degrees of freedom.

Proof. Consider the random variables

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1) \qquad V = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \qquad \text{and} \qquad \frac{Z}{\sqrt{\dfrac{V}{n-1}}}$$

and use the properties of the normal distribution and related random variables (**NOTE (6), p. 60, section 4.5)**

6.2 HYPOTHESIS TESTING*

Given a sample $x_1, ..., x_n$ from a probability model $f_X(x;\theta)$ depending on parameter θ, we might want to test, say, whether or not there is evidence from the sample that true (but unobserved) value of θ is not equal to a specified value. We use **sampling distributions** to quantify this evidence. We will look at two situations, namely **one sample** and **two sample** experiments.

	Random Variables	Data	Models ?
ONE SAMPLE :	$X_1, ..., X_n \sim N(\mu, \sigma^2)$	$x_1, ...x_n$	$\mu = c_1, \sigma = c_2$
TWO SAMPLE :	$X_1, ..., X_n \sim N(\mu_X, \sigma_X^2)$	$x_1, ...x_n$	$\mu_X = \mu_Y, \sigma_X = \sigma_Y$
	$Y_1, ..., Y_n \sim N(\mu_Y, \sigma_Y^2)$	$y_1, ...y_n$	

6.2.1 TESTS FOR NORMAL DATA I - THE Z-TEST (σ KNOWN)

Recall that, if $X_1, ..., X_n \sim N(\mu, \sigma^2)$ are the i.i.d. outcome random variables of n experimental trials, then

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{and} \qquad \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$

with $\bar{X}$ and S^2 statistically independent. Suppose we want to test the **hypothesis** that $\mu = c$, for some specified constant c, (where, for example, $c = 20.0$) is a plausible model; more specifically, we want to test

$$\begin{aligned} H_0 &: \mu = c & &\text{the \textbf{NULL} hypothesis} \\ H_1 &: \mu \neq c & &\text{the \textbf{ALTERNATIVE} hypothesis} \end{aligned}$$

Specifically, we want to test whether H_0 is true, or whether H_1 is true. Now, we know that, in the case of a Normal sample, the distribution of the estimator $\bar{X}$ is Normal, and

$$\bar{\bar{X}} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Longrightarrow Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

where Z is a **random variable**. Now, when we have observed the data sample, we can calculate $\bar{x}$, and therefore we have a way of testing whether $\mu = c$ is a plausible model; we calculate

$$z = \frac{\bar{x} - c}{\sigma/\sqrt{n}}$$

If H_0 is true, and $\mu = c$, then the **observed** z should be an observation from an $N(0,1)$ distribution (as $Z \sim N(0,1)$), that is, it should be near zero with high probability. In fact, z should lie between -1.96 and 1.96 with probability $1 - \alpha = 0.95$, say, as

$$\text{P}[-1.96 \leq Z < 1.96] = \Phi(1.96) - \Phi(-1.96) = 0.975 - 0.025 = 0.95$$

.If we observe z to be outside of this range, then there is evidence that H_0 is **not true**.

Alternatively, we could calculate the probability p of observing a z value that is **more extreme** than the z we did observe; this probability is given by

$$p = 2\Phi(-|z|)$$

If p is very small, say $p \leq \alpha = 0.05$, then there is evidence that H_0 is **not true**. In summary, if z is a **surprising** observation from an $N(0,1)$ distribution then we can **reject** H_0.

6.2.2 HYPOTHESIS TESTING TERMINOLOGY

There are five crucial components to a hypothesis test, namely

- **TEST STATISTIC**
- **NULL DISTRIBUTION**
- **SIGNIFICANCE LEVEL**, denoted α
- **P-VALUE**, denoted p.
- **CRITICAL VALUE(S)**

In the Normal example given above, we have that

z is the **test statistic**

The distribution of random variable Z if H_0 is true is the **null distribution**

$\alpha = 0.05$ is the **significance level** of the test (we could use $\alpha = 0.01$ if we require a "stronger" test).

p is the **p-value** of the test statistic under the null distribution

The solution C_R of $\Phi(C_R) = 1 - \alpha/2$ ($C_R = 1.96$ above) gives the **critical values** of the test $\pm C_R$.

6.2.3 TESTS FOR NORMAL DATA II - THE T-TEST (σ UNKNOWN)

If the variance σ^2 is **unknown**, we proceed as follows; the sampling distributions of $\bar{X}$ and s^2 are known from Theorem 6.1.1, and that the two estimators are statistically independent. Now, from the properties of the Normal distribution, if we have independent random variables $Z \sim N(0,1)$ and $Y \sim \chi^2_\nu$, then we know that random variable T defined by

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

has a Student-t distribution with ν degrees of freedom. Using this result, and recalling the sampling distributions of $\bar{X}$ and s^2, we see that

$$T = \frac{\bar{\bar{X}} - \mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)s^2/\sigma^2}{(n-1)}} = \frac{(\bar{X} - \mu)}{s/\sqrt{n}} \sim t_{n-1}$$

and T has a Student-t distribution with $n-1$ degrees of freedom, denoted $St(n-1)$. Thus we can repeat the procedure used in the σ known case, but use the sampling distribution of T rather than that of Z to assess whether the test statistic is "surprising" or not. Specifically, we calculate

$$t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}}$$

and find the critical values for a $\alpha = 0.05$ significance test by finding the ordinates corresponding to the 0.025 and 0.975 percentiles of a Student-t distribution, $St(n-1)$ (rather than a $N(0,1)$) distribution. Finally, we can calculate the probability p of observing a T that is **more extreme** than the t we did observe; this probability is given by

$$p = 2F_{St(n-1)}(-|t|)$$

If p is very small, say $p \leq \alpha = 0.05$, then, again, there is evidence that H_0 is **not true**.

6.2.4 TESTS FOR NORMAL DATA III - TESTING σ

The Z-test and T-test are both tests for the parameter μ. Suppose that we wish to test a hypothesis about σ, for example

$$\begin{aligned} H_0 &: \sigma^2 = c \\ H_1 &: \sigma^2 \neq c \end{aligned}$$

We construct a test based on the estimate of variance, s^2. In particular, we saw from Theorem 6.1.1 that the random variable Q, defined by

$$Q = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

if the data have an $N(\mu, \sigma^2)$ distribution. Hence if we define test statistic q by

$$q = \frac{(n-1)s^2}{c}$$

then we can compare q with the critical values derived from a χ^2_{n-1} distribution; we look for the 0.025 and 0.975 quantiles - note that the Chi-squared distribution is not symmetric, so we need two distinct critical values

$$C_{R_1} = F_{\chi^2_{n-1}}(0.025) \qquad\qquad C_{R_2} = F_{\chi^2_{n-1}}(0.975)$$

If q is a surprising observation from a χ^2_{n-1} distribution, and we cannot reject H_0.

6.2.5 TWO SAMPLE TESTS

It is straightforward to extend the ideas from the previous sections to two sample situations where we wish to compare the distributions underlying two data samples. Typically, we consider sample one, $x_1, ..., x_{n_X}$, from a $N(\mu_X, \sigma_X^2)$ distribution, and sample two, $y_1, ..., y_{n_Y}$, independently from a $N(\mu_Y, \sigma_Y^2)$ distribution, and test the equality of the parameters in the two models. Suppose that the sample mean and sample variance for samples one and two are denoted $(\bar{x}, s_X^2)$ and $(\bar{y}, s_Y^2)$ respectively.

First, consider testing the hypothesis

$$\begin{aligned} H_0 &: \mu_X = \mu_Y \\ H_1 &: \mu_X \neq \mu_Y \end{aligned}$$

when $\sigma_X = \sigma_Y = \sigma$ is known. Now, we have from the sampling distributions theorem we have

$$\bar{X} \sim N\left(\mu_X, \frac{\sigma^2}{n_X}\right) \qquad \bar{Y} \sim N\left(\mu_Y, \frac{\sigma^2}{n_Y}\right) \quad \Longrightarrow \bar{X} - \bar{Y} \sim N\left(0, \frac{\sigma^2}{n_X} + \frac{\sigma^2}{n_Y}\right)$$

and hence

$$Z = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \sim N(0,1)$$

giving us a test statistic z defined by

$$z = \frac{\bar{x} - \bar{y}}{\sigma\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

which we can compare with the standard normal distribution; if z is a surprising observation from $N(0,1)$, and lies outside of the critical region, then we can reject H_0. This procedure is the **Two Sample Z-Test**.

If $\sigma_X = \sigma_Y = \sigma$ is unknown, we parallel the one sample T-test by replacing σ by an estimate in the two sample Z-test. First, we obtain an estimate of σ by "pooling" the two samples; our estimate is the **pooled estimate**, s_P^2, defined by

$$s_P^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}$$

which we then use to form the test statistic t defined by

$$t = \frac{\bar{x} - \bar{y}}{s_P\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

It can be shown that, if H_0 is true then t should be an observation from a Student-t distribution with $n_X + n_Y - 2$ degrees of freedom. Hence we can derive the critical values from the tables of the Student-t distribution.

If $\sigma_X \neq \sigma_Y$, but both parameters are known, we can use a similar approach to the one above to derive test statistic z defined by

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

which has an $N(0,1)$ distribution if H_0 is true.

Clearly, the choice of test depends on whether $\sigma_X = \sigma_Y$ or otherwise; we may test this hypothesis formally; to test

$$\begin{aligned} H_0 &: \sigma_X = \sigma_Y \\ H_1 &: \sigma_X \neq \sigma_Y \end{aligned}$$

we compute the test statistic

$$q = \frac{s_X^2}{s_Y^2}$$

which has a null distribution known as the **Fisher** or F distribution with $(n_X - 1, n_Y - 1)$ degrees of freedom; this distribution can be denoted $F(n_X - 1, n_Y - 1)$, and its quantiles are tabulated. Hence we can look up the 0.025 and 0.975 quantiles of this distribution (the F distribution is not symmetric), and hence define the critical region; informally, if the test statistic q is very small or very large, then it is a surprising observation from the F distribution and hence we reject the hypothesis of equal variances.

6.2.6 CONFIDENCE INTERVALS

The procedures above allow us to test specific hypothesis about the parameters of probability models. We may complement such tests by reporting a **confidence interval**, which is an interval in which we believe the "true" parameter lies with high probability. Essentially, we use the sampling distribution to derive such intervals.

For example, in a one sample Z-test, we saw that

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

that is, that, for critical values $\pm C_R$ in the test at the 5% Significance level

$$P[\ -C_R \leq Z \leq C_R\ ] = P\left[-C_R \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq C_R\right] = 0.95$$

Now, from tables we have $C_R = 1.96$, so re-arranging this expression we obtain

$$P\left[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right] = 0.95$$

from which we deduce a **95 % Confidence Interval** for μ based on the sample mean $\bar{x}$ of

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

We can derive other confidence intervals (corresponding to different significance levels in the equivalent tests) by looking up the appropriate values of the critical values. The general approach for construction of confidence interval for generic parameter θ proceeds as follows. From the modelling assumptions, we derive a **pivotal quantity**, that is, a statistic, T_{PQ}, say, (usually the test statistic random variable) that depends on θ, but whose sampling distribution is "parameter-free" (that is, does not depend on θ). We then look up the critical values C_{R_1} and C_{R_2}, such that

$$P\left[C_{R_1} \leq T_{PQ} \leq C_{R_2}\right] = 1 - \alpha$$

where α is the significance level of the corresponding test. We then rearrange this expression to the form

$$P\left[c_1 \leq \theta \leq c_2\right] = 1 - \alpha$$

where c_1 and c_2 are functions of C_{R_1} and C_{R_2} respectively. Then a $1-\alpha$ % Confidence Interval for θ is $[c_1, c_2]$.

HYPOTHESIS TESTING SUMMARY

In general, to test a hypothesis, consider a statistic calculated from the sample data. Derive the probability distribution of the statistic when the hypothesis is true, and compare the actual value of the statistic with the hypothetical probability distribution. Assess whether the value is a likely observation from this probability distribution. If it is not, then reject the hypothesis.

6.3 ESTIMATION

Definition 6.3.1 Let $X_1, ..., X_n$ be a random sample from a distribution with mass/density function f_X that depends on (possibly vector) parameter θ, that is, $f_{X_1}(x_1) = f_X(x_1;\theta)$, so that

$$f_{X_1,...,X_k}(x_1, ..., x_k) = \prod_{i=1}^{k} f_X(x_i;\theta)$$

then a statistic $T = t(X_1, ..., X_n)$ that is used to represent a function $\tau(\theta)$ of θ based on the sample $x_1, ..., x_n$ is an **estimator**, and $t = t(x_1, ..., x_n)$ is an **estimate**, $\hat{\tau}(\theta)$, of $\tau(\theta)$.

6.3.1 ESTIMATION TECHNIQUES I: METHOD OF MOMENTS

Suppose that $X_1, ..., X_n$ is a random sample from a probability distribution with mass/density function f_X that depends on vector parameter θ of dimension k, and suppose that a sample $x_1, ..., x_n$ has been observed. Let the jth moment of f_X be denoted μ_j, and let the jth sample moment be denoted m_j for $j = 1, ..., k$. Then

$$m_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j \qquad \text{is an } \textbf{estimate} \text{ of } \mu_j \qquad\qquad M_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j \qquad \text{is an } \textbf{estimator} \text{ of } \mu_j^{'}$$

Interpretation : This method of estimation involves matching the theoretical moments to the sample moments, giving (in most cases) k equations in the k elements of vector θ which may be solved simultaneously to find the parameter estimates. Intuitively, and recalling the Weak Law of Large Numbers, it is reasonable to suppose that there is a close relationship between the theoretical properties of a probability distribution, and large sample derived estimates; for example, we know that, for large n, the sample mean converges in probability to the theoretical expectation.

6.3.2 ESTIMATION TECHNIQUES II: MAXIMUM LIKELIHOOD

Definition 6.3.2 Let random variables $X_1, ..., X_n$ have joint mass or density function, denoted $f_{X_1,...,X_k}$, that depends on vector parameter $\theta = (\theta_1, ..., \theta_k)$. Then the joint/mass density function evaluated at fixed (possibly observed) values of the variables, $x_1, ..., x_n$, and viewed as a function of θ is the **likelihood function**, $L(\theta)$,

$$L(\theta) = f_{X_1,...,X_n}(x_1, ..., x_n;\theta)$$

If $X_1, ..., X_n$ represents a random sample from joint/mass density function f_X

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

Definition 6.3.3 Let $L(\theta)$ be the likelihood function derived from the joint/mass density function of random variables $X_1, ..., X_n$, where $\theta \in \Theta \subseteq \mathbb{R}^k$, say, and Θ is termed the parameter space. Then for a fixed set of observed values $x_1, ..., x_n$ of the variables, the estimate of θ termed the maximum likelihood estimate of θ, $\hat{\theta}$, is defined by

$$\hat{\theta} = \arg\max_{\theta\in\Theta} L(\theta)$$

that is, the value of θ for which $L(\theta)$ is maximized in the parameter space Θ.

Interpretation : This method of estimation involves finding the value of θ for which $L(\theta)$ is maximized by setting the first partial derivatives of $L(\theta)$ with respect to θ_j equal to zero, for $j = 1, ..., k$, and solving the resulting k simultaneous equations. Suppose a sample $x_1, ..., x_n$ has been obtained from a probability model specified by mass or density function $f_X(x;\theta)$ depending on parameter(s) θ lying in parameter space Θ. The maximum likelihood estimate (m.l.e.) is produced as follows;

THE FOUR STEP PROCEDURE

- Write down the likelihood function, $L(\theta)$.
- Take the natural log of the likelihood, collect terms involving θ.
- Find the value of $\theta \in \Theta$, $\hat{\theta}$, for which $\log L(\theta)$ is maximized, for example by differentiation. Note that, if parameter space Θ is a bounded interval, then the maximum likelihood estimate may lie on the boundary of Θ. If the parameter is a k vector, the maximization involves evaluation of partial derivatives.
- Check that the estimate $\hat{\theta}$ obtained in STEP 3 truly corresponds to a maximum in the (log) likelihood function by inspecting the second derivative of $\log L(\theta)$ with respect to θ. In the single parameter case, if the second derivative of the log -likelihood is negative at $\theta = \hat{\theta}$, then $\hat{\theta}$ is confirmed as the m.l.e. of θ (other techniques may be used to verify that the likelihood is maximized at $\hat{\theta}$).

EXAMPLE Suppose a sample $x_1, ..., x_n$ is modelled by a Poisson distribution with parameter denoted λ, so that

$$f_X(x;\theta) \equiv f_X(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \qquad x = 0, 1, 2, ...$$

for some $\lambda > 0$. To estimate λ by maximum likelihood, proceed as follows.

<u>**STEP 1**</u> Calculate the likelihood function $L(\lambda)$.

$$L(\lambda) = \prod_{i=1}^{n} f_X(x_i;\lambda) = \prod_{i=1}^{n} \left\{ \frac{\lambda^{x_i}}{x_i!}e^{-\lambda} \right\} = \frac{\lambda^{x_1+...+x_n}}{x_1!....x_n!}e^{-n\lambda}$$

for $\lambda \in \Theta = R^+$.

<u>**STEP 2**</u> Calculate the log-likelihood $\log L(\lambda)$.

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

<u>**STEP 3**</u> Differentiate $\log L(\lambda)$ with respect to λ, and equate the derivative to zero to find the m.l.e..

$$\frac{d}{d\lambda}\{\log L(\lambda)\} = \sum_{i=1}^{n} \frac{x_i}{\lambda} - n = 0 \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Thus the maximum likelihood estimate of λ is $\hat{\lambda} = \bar{x}$

STEP 4 Check that the second derivative of $\log L(\lambda)$ with respect to λ is negative at $\lambda = \hat{\lambda}$.

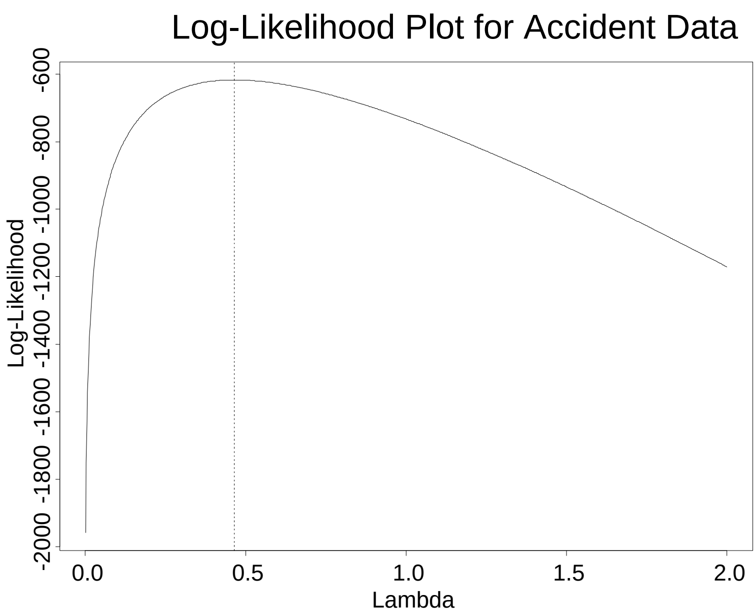
$$\frac{d^2}{d\lambda^2}\{\log L(\lambda)\} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0 \quad \text{at } \lambda = \hat{\lambda}$$

EXAMPLE: The following data record the numbers of domestics accidents observed to occur per household in one year. For this experiment, the total number of household studied is $n = 647$

Number of accidents	Frequency
0	447
1	132
2	42
3	21
4	3
5	2

Using the maximum likelihood procedure, the estimate of λ if a Poisson model is assumed is

$$\hat{\lambda}_{ML} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{(447 \times 0) + (132 \times 1) + (42 \times 2) + (21 \times 3) + (3 \times 4) + (2 \times 5)}{647} = 0.465$$



Note that, here, the maximum likelihood estimator is of the same form as the method of moments estimator, that is, $\widehat{\lambda}_{MM} = \widehat{\lambda}_{ML} = \overline{X}$

6.4 PROPERTIES OF ESTIMATORS*

Having constructed an estimator using the methods described above, we then seek to assess the properties of the estimator that might lead us to use it in preference to other estimators.

- **INVARIANCE** : An estimator, T, is an **invariant** estimator of parameter θ if, for a function $\tau(\theta)$ of θ, $\tau(T)$ is an estimator of $\tau(\theta)$.

- **UNBIASEDNESS** : An estimator, T, is an **unbiased** estimator of function $\tau(\theta)$ of parameter θ if

$$E_{f_T}[T] = \tau(\theta)$$

where f_T is the sampling distribution of T. The **bias**, $b(T)$, of an estimator T of $\tau(\theta)$ is defined by

$$b(T) = E_{f_T}[T] - \tau(\theta)$$

and the **Mean Squared Error**, or MSE, of T is defined by

$$MSE(T) = E_{f_T}[(T - \tau(\theta))^2]$$

- **ASYMPTOTIC UNBIASEDNESS** A sequence $T_1, ..., T_n, ...$ of estimators of function $\tau(\theta)$ of parameter θ are **asymptotically unbiased** if,

$$\lim_{n \longrightarrow \infty} E_{f_{T_n}}[T_n] = \tau(\theta)$$

for **every** $\theta \in \Theta$.

- **SIMPLE CONSISTENCY** : A sequence $T_1, ..., T_n, ...$ of estimators of function $\tau(\theta)$ of parameter θ. The sequence of estimators are **consistent** estimators of $\tau(\theta)$ if, for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} P[|T_n - \tau(\theta)| < \epsilon] = 1$$

for every $\theta \in \Theta$, that is, if

$$T_n \stackrel{p}{\longrightarrow} \tau(\theta)$$

and T_n converges in probability to $\tau(\theta)$.

- **MSE CONSISTENCY** : A sequence $T_1, ..., T_n, ...$ of estimators of function $\tau(\theta)$ of parameter θ are **mean squared error consistent** if,

$$\lim_{n \longrightarrow \infty} E_{f_{T_n}}[(T_n - \tau(\theta))^2] = 0$$

for every $\theta \in \Theta$.

- **EFFICIENCY**: It is desirable that the sampling distribution of an estimator is concentrated around the true value of the parameter. For example, if T_1 and T_2 are two possible estimators of a function $\tau(\theta)$ of parameter θ, then we would prefer T_1 to T_2 if T_1 is more concentrated about $\tau(\theta)$ than is T_2, that is, if, for all $\epsilon > 0$,

$$P[\tau(\theta) - \epsilon < T_1 < \tau(\theta) + \epsilon] \geq P[\tau(\theta) - \epsilon < T_2 < \tau(\theta) + \epsilon]$$

where the left and right hand side depend on the sampling distributions of T_1 and T_2 respectively. Furthermore, if T is an unbiased estimator of $\tau(\theta)$, then the **Chebychev Inequality** suggests that

$$P[\tau(\theta) - \epsilon < T < \tau(\theta) + \epsilon] \geq 1 - Var_{f_T}[T]/\epsilon^2$$

for all $\epsilon > 0$, as $\mathrm{E}_{f_T}[T] = \tau(\theta)$. Hence an unbiased estimator whose sampling distribution has a small variance will be concentrated around $\tau(\theta)$, and thus will be preferable to estimators with larger variance. If this is true for all possible values of parameter θ, then this estimator might be regarded as the "best" estimator of $\tau(\theta)$.

An unbiased T_1 estimator is more **efficient** than another estimator T_2 if

$$Var_{f_{T_1}}[T_1] \leq Var_{f_{T_2}}[T_2]$$

ASYMPTOTIC PROPERTIES OF MLES

It can be shown that, under certain regularity conditions, maximum likelihood estimators have desirable properties. In particular if $\hat{\theta}_n$ is the maximum likelihood estimator of parameter θ derived from a random sample of size n, then

(i) $\hat{\theta}_n$ exists and is unique
(ii) $\hat{\theta}_n$ is consistent
(iii) $\hat{\theta}_n$ is asymptotically normally distributed, that is,

$$\hat{\theta}_n \xrightarrow{d} N\left(\theta, \frac{1}{nE_{f_X}\left[\{\partial(X;\theta)\}^2\right]}\right)$$

where

$$\partial(x;\theta) = \frac{\partial}{\partial\theta}\{\log f_X(x;\theta)\}$$

is the **score function,** that is, $\hat{\theta}_n$ is asymptotically unbiased, and it can be shown that $\hat{\theta}_n$ has minimum variance in the class of unbiased estimators.