An/12

THE METHOD OF MAXIMUM LIKELIHOOD

By

M. H. A. AGHA

Thesis for M.Sc. degree presented to the University of Glasgow in July, 1962.

CONTENTS

CONTENTS

ACKNOWLEDEMENT

I wish to express my gratitude to Dr. R. A. Robb, for suggesting the problem and his general supervision, his constant advice and encouragement throughout the preparation of this thesis and also Mr. J. Aitchison, for helpful discussion and guidance.

Glasgow, July, 1962 M. H. A. AGHA

CHAPTER I

SINGLE PARAMETER

1. Introduction

Every experiment has one or more unknown parameter. Thus the purpose of the experimental work is to obtain the infomation about these unknown parameters. The outcome of the experiment represents an observation obtained from a population having a certain form of frequency distribution specified by one or more unknown parameter; therefore the outcomes of the repeated experiment will represent a random sample drawn from that population.

The problem of estimation is therefore to estimate the unknown parameters of the population from the observations of the sample which is drawn from that population. Thus it is clear that we require to establish some systematic estimation procedure, in order to estimate the unknown parameters of the population from the information obtained from the sample observations.

There are several methods of estimation; one of them is the maximum likelihood method which is the oldest one. Each of these methods has some optimum properties, but the maximum likelihood method has all the properties of the best method of estimation.

The theory of estimation in fact has been highlighted by R. A. Fisher in his papers "On the Mathematical Foundations of Theoretical Statistics" (1921) and "Theory of Statistical

Estimation" (1925), in which very fruitful work on the maximum likelihood has been done. In recent years Fisher and some other authors, introduced very wide developments in the maximum likelihood method, which has since been widely used in practical applications.

The properties of the best method of estimation:

1.1 <u>Consistency</u>: Let $x_1, \ldots\ldots, x_n$ be a random sample of size n drawn from a population having the probability density function $f(x, \theta)$. Let the statistic $t(x_1, ----, x_n)$ be the estimator of the true value θ_0 of the parameter θ, then t will be said to be a consistent estimator of θ_0, if

$$P_r\left\{|t-\theta_0| > \delta\right\} \longrightarrow 0 \qquad \text{as} \quad n \rightarrow \infty$$

where δ is any arbitrary small positive number.

1.2 <u>Normality</u>: If x is a continuous random variable with probability density function $f(x, \theta, \sigma^2)$ defined by

$$f(x, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\} dx,$$

then x is said to be distributed normally with mean θ and variance σ^2, where $-\infty < \theta < \infty$ and $0 < \sigma^2 < \infty$. This expression is denoted by $N(\theta, \sigma^2)$.

1.3 <u>Unbiasedness</u>: If a statistic t is obtained from the information of the sample observations with probability density function $f(x, \theta)$, then t is said to be an unbiased estimator of the parameter θ, if

$$E(t) = \theta,$$

where E denotes the expectation. That is t is centred on the value of the parameter.

1.4 Efficiency: In some cases there are more than one consistent and unbiased estimator for estimating the true value of the parameter. For example, the median in example 9.7 (The Advanced Theory of Statistics, Vol. 1, page 213) is distributed normally, as the sample size tends to infinity and it is consistent and unbiased. The property which discriminates between these estimators, to show us the best one is called the efficiency.

If t_1 and t_2 are two estimators to the true value of the parameter with variances V_1 and V_2 respectively and the minimum attainable variance is V, then the efficiencies of t_1 and t_2 are respectively defined by

$$E_1 = \frac{V}{V_1} \qquad \text{and} \qquad E_2 = \frac{V}{V_2} .$$

That is, the estimator with smaller variance will be more efficient than the other.

For example in the case of the normal distribution defined by

$$f(x, \theta, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\} ,$$

we have for any n, the variance of the mean is

$$\frac{\sigma^2}{n}$$

and for large n, the variance of the median is

$$\frac{\pi \sigma^2}{2n}$$

The efficiency of the median with respect to the mean is then

$$E = \frac{\sigma^2}{n} \Big/ \frac{\pi\sigma^2}{2n}$$

$$= \frac{2}{\pi} = 0.637 = 63.7\% .$$

1.4(a) <u>The Fisher's Inequality</u>: Let $x_1, ----, x_n$ be a random sample from a population having a probability density function $f(x,\theta)$ and let $t(x_1,---,x_n)$ be an unbiased estimate of $g(\theta)$ a function of the unknown parameter θ. Then the inequality which is defined independently of any method of estimation is

$$V(t) \geqslant \frac{[g'(\theta)]^2}{nE\left(\frac{\partial \log f}{\partial\theta}\right)^2} = \frac{[g'(\theta)]^2}{n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2 f\,dx},$$

where $g'(\theta)$ is the first derivative of $g(\theta)$ and $V(t)$ denotes the variance of the statistic t. This inequality affords the minimum variance and also the amount of information on θ supplied by the sample observations which is defined by

$$n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial\theta}\right)^2 f\,dx .$$

To prove the inequality above we have to consider the following conditions be satisfied.

(a) The range of the stochastic variable is independent of θ.

(b) The probability density function is differentiable under the integral sign.

<u>Proof</u>: Let $G(x,\theta)$ be the joint distribution of the sample values. Then

$$\int --- \int t\,G(x,\theta)\,dx_1 ---- dx_n = g(\theta) .$$

In virtue of condition (b), we have

$$\int \cdots \int t \frac{\partial G}{\partial \theta} \, dx_1 \cdots dx_n = g'(\theta) .$$

The covariance between t and $\frac{1}{G} \frac{\partial G}{\partial \theta}$ is given by

$$\int \cdots \int t \, \frac{1}{G} \frac{\partial G}{\partial \theta} \, G \, dx_1 \cdots dx_n = \int \cdots \int t \, \frac{\partial G}{\partial \theta} \, dx_1 \cdots dx_n = g'(\theta) .$$

We have that

$$\rho^2 = \left[cov\left(t, \frac{1}{G} \frac{\partial G}{\partial \theta}\right)\right]^2 \Big/ var(t) \, var\left(\frac{1}{G} \frac{\partial G}{\partial \theta}\right) ,$$

where ρ is the correlation coefficient between t and $\frac{1}{G} \frac{\partial G}{\partial \theta}$.

That is

$$V(t) \, V\left(\frac{1}{G} \frac{\partial G}{\partial \theta}\right) \geqslant \left[cov\left(t, \frac{1}{G} \frac{\partial G}{\partial \theta}\right)\right]^2 ,$$

since $0 \leqslant \rho \leqslant 1$, where V and C denote the variance and covariance.

Then

$$V(t) \geqslant \frac{[g'(\theta)]^2}{V\left(\frac{1}{G} \frac{\partial G}{\partial \theta}\right)} .$$

Since

$$V\left(\frac{1}{G} \frac{\partial G}{\partial \theta}\right) = V\left(\frac{\partial \log G}{\partial \theta}\right) = E\left(\frac{\partial \log G}{\partial \theta}\right)^2$$

$$= n E\left(\frac{\partial \log f}{\partial \theta}\right)^2 = n \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta}\right)^2 f \, dx ,$$

$\therefore$

$$V(t) \geqslant [g'(\theta)]^2 \Big/ n \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta}\right)^2 f \, dx$$

& $\quad V(t) \geqslant 1 \Big/ n \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta}\right)^2 f \, dx ,\qquad$ when $g(\theta) = \theta$.

1.4(b) Properties of Efficient Estimators:

Let t and t' be two efficient estimators of the same parameter, each one having variance equal to $\frac{\sigma^2}{n}$ and let the correlation coefficient between them be ρ. If t'' is another estimator defined by

$$t'' = \frac{1}{2}(t + t')$$

then t'' will be an efficient estimator with the same variance of t and t'.

We have

$$\rho = \frac{\mathrm{cov}(t, t')}{\sqrt{\mathrm{var}(t)\ \mathrm{var}(t')}} = \frac{\mathrm{cov}(t, t')}{\frac{\sigma^2}{n}},$$

ie.

$$\mathrm{cov}(t, t') = \frac{\sigma^2}{n}\rho .$$

Also we have

$$\begin{aligned}\mathrm{var}(t + t') &= \mathrm{var}(t) + \mathrm{var}(t') + 2\,\mathrm{cov}(t, t') \\ &= \frac{\sigma^2}{n} + \frac{\sigma^2}{n} + 2\frac{\sigma^2}{n}\rho = 2\frac{\sigma^2}{n}(1+\rho),\end{aligned}$$

then

$$\mathrm{var}\,\frac{1}{2}(t+t') = \frac{1}{4}\mathrm{var}(t+t') = \frac{\sigma^2}{n}\left(\frac{1+\rho}{2}\right),$$

ie.

$$\mathrm{var}(t'') = \frac{\sigma^2}{n}\left(\frac{1+\rho}{2}\right)$$

Here ρ can not be less than 1, because $\mathrm{var}(t'') \nless \mathrm{var}(t)$ or $\mathrm{var}(t')$; and since ρ is not greater than 1; therefore $\rho = 1$, and so

$$\mathrm{var}(t'') = \frac{\sigma^2}{n}.$$

That is, for large samples the efficient estimators are equivalent.

1.4(c) <u>Distribution Admitting most Efficient Estimator:</u>

We have from 1.4(a) that

$$V(t)\,V\left(\frac{1}{G}\frac{\partial G}{\partial \theta}\right) \geqslant \left[C\left(t, \frac{1}{G}\frac{\partial G}{\partial \theta}\right)\right]^2 .$$

This inequality may be written as

$$\int [t-g(\theta)]^2\, G(x;\theta)\,dx \cdot \int \left(\frac{1}{G}\frac{\partial G}{\partial \theta}\right)^2 G(x;\theta)\,dx \geqslant \left[\int t\left(\frac{1}{G}\frac{\partial G}{\partial \theta}\right) G(x;\theta)\,dx\right.$$

where $\int$ represents the multiple integral and $dx = dx_1 \cdots dx_n$.
From Schwarz's inequality, the equality occurs when

$$[t - g(\theta)] = \ell\,\frac{1}{G}\frac{\partial G}{\partial \theta}$$

ie.

$$\frac{t}{\ell} - \frac{g(\theta)}{\ell} = \frac{\log G}{\partial \theta}$$

where ℓ is constant dependent on θ. Then

$$\log G = \int \left[\frac{t}{\ell} - \frac{g(\theta)}{\ell}\right] d\theta = K + tX + Y$$

where K is independent of θ, and X and Y are functions of θ.
Hence

$$G = \exp(K + tX + Y)$$
$$= G' \exp(tX + Y),$$

where G' is independent of θ. Since we deduced the last formula from the inequality above which affords the minimum variance when the equality occurs, thus the last formula represents the distribution admitting the most efficient estimator.

<u>Example 1.1</u> Consider the Poisson distribution; the joint frequency function is then

$$G(x;\theta) = \frac{e^{-n\theta}\,\theta^{\Sigma x_i}}{\prod_1^n x_i!}$$
$$= \frac{1}{\prod_1^n x_i!} \exp\left\{(\Sigma x_i)\log\theta - n\theta\right\}.$$

Here

$$G' = \frac{1}{\prod_1^n x_i!}$$

&

$$\exp\{tX+Y\} = \exp\{(\Sigma x_i)\log\theta - n\theta\}.$$

Therefore the distribution admits a most efficient estimator.

<u>Example 1.2</u> Consider the normal distribution with unknown mean μ and known variance σ^2, the joint frequency function is then

$$G(x;\mu) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(x-\mu)^2\right\}$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\Sigma x^2 + \frac{1}{\sigma^2}\mu\Sigma x - \frac{n}{\sigma^2}\mu^2\right\}.$$

Here

$$G' = e^{-\frac{1}{2\sigma^2}\Sigma x^2} \Big/ (2\pi\sigma^2)^{\frac{n}{2}}$$

&

$$\exp\{tX+Y\} = \exp\left\{\frac{1}{2\sigma^2}(2\mu\Sigma x - n\mu^2)\right\}.$$

Therefore the distribution admits a most efficient estimator.

1.5 <u>Sufficiency:</u>

Let $x_1, \ldots, x_n$ be a random sample from a population with probability density function $f(x,\theta)$. Then the necessary condition that the estimator t be sufficient for θ is

$$\prod_1^n f(x_i,\theta) = f_1(t,\theta) f_2(x_1,\ldots,x_n)$$

that is $\prod_1^n f(x_i,\theta)$ factorised into two parts, $f_1(t,\theta)$ dependent on t and θ only and $f_2(x_1,\ldots,x_n)$ is independent of θ. We can extend this property to more than one parameter, Let $t_1, \ldots, t_m$ be estimators of the parameters $\theta_1, \ldots, \theta_m$ then the necessary condition that the estimators t's are sufficient for θ's is

$$\prod_1^n f(x_i, \theta_1, ---, \theta_m) = f_1(t_1, ---, t_m; \theta_1, ---, \theta_m)\, f_2(x) \,.$$

1.5(a) The General Distribution Admitting Sufficient Statistic:

Let $x_1, ---, x_n$ be a random sample and each random variable have density function $f(x, \theta)$ then the joint distribution of the sample values is

$$F(x, \theta) = \prod_1^n f(x_i, \theta) \,. \qquad i = 1, 2, ---, n$$

If there exists a sufficient statistic t, say, as an estimate of the parameter θ, then $F(x, \theta)$ can be factorised as

$$F(x, \theta) = F_1(t, \theta)\, F_2(x)$$

Taking the logarithm and differentiating with respect to θ we get

$$\frac{\partial \log F(x, \theta)}{\partial \theta} = \frac{\partial \log F_1(t, \theta)}{\partial \theta} = H(t, \theta) \quad ,$$

where $H(t, \theta)$ is a function of t and θ. If we substitute any particular value of θ in $H(t, \theta)$ then $H(t, \theta)$ will be a function of t, $h(t)$, say, where $h(t)$ may be put as

$$h(t) = h\{g(x)\} \,.$$

Now

$$\frac{\partial H(t, \theta)}{\partial t} \div \frac{d\, h(t)}{dt} = k(\theta)$$

i.e.

$$\frac{\partial H(t, \theta)}{\partial t} = k(\theta)\, \frac{d\, h(t)}{dt}$$

Integrating with respect to t, we obtain

$$H(t, \theta) = \frac{\partial \log F}{\partial \theta} = k(\theta)\, h(t) + l(\theta) \,.$$

Integrating with respect to θ, we get

$$\log F = K(\theta)\, h(t) + L(\theta) + m(x) \quad ,$$

that is,

$$F = F' \exp\left[K(\theta)\, h(t) + L(\theta)\right]$$

where F' is a function of x and $K(\theta)$ and $L(\theta)$ are functions of θ .

Example 1.3 Consider the normal distribution with mean θ and variance σ^2 , where σ^2 is known, then the frequency function may be put in the following form

$$F(x,\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum x^2\right\} \exp\left\{-\frac{n}{2\sigma^2}(\theta^2 - 2\bar{x}\theta)\right\}.$$

Here

$$F' = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum x^2\right\}$$

&

$$\exp\left\{K(\theta)\, h(t) + L(\theta)\right\} = \exp\left\{\frac{n}{\sigma^2}\bar{x}\theta - \frac{n}{2\sigma^2}\theta^2\right\}.$$

Therefore a sufficient statistic for θ exists.

Example 1.4 Consider the Poisson distribution, then the joint frequency function will be such that

$$F(x,\theta) = \frac{e^{-n\theta}\ \theta^{n\bar{x}}}{\prod_1^n x_i!} = \frac{1}{\prod_1^n x_i!} \exp\left\{\log \theta^{n\bar{x}} - n\theta\right\}.$$

Here

$$F'(x) = \frac{1}{\prod_1^n x_i!}$$

&

$$\exp\left\{K(\theta)\, h(t) + L(\theta)\right\} = \exp\left\{n(\bar{x}\log\theta - \theta)\right\}.$$

Thus a sufficient statistic of θ exists.

2. The Principle of Maximum Likelihood

If $f(x,\theta)$ is the frequency function of a population, then the likelihood function of a sample of size n drawn from that population, is defined by

$$L(x,\theta) = \prod_{1}^{n} f(x_i,\theta)$$

where θ is the parameter of the population.

Now, if the statistic $t(x_1, ---, x_n)$ maximises the likelihood function $L(x,\theta)$ for variations of θ, then $t(x_1, ---, x_n)$ is called the maximum likelihood estimator of θ.

In virtue of the foregoing, the statistic $t(x_1,--,x_n)$ will be the solution of the equation

$$\frac{\partial L(x,\theta)}{\partial \theta} = 0,$$

or the solution of the equation

$$\frac{1}{L}\frac{\partial L(x,\theta)}{\partial \theta} = 0,$$

ie.

$$\frac{\partial \log L(x,\theta)}{\partial \theta} = 0,$$

since $L(x,\theta) \neq 0$.

More frequently the last equation was found to be easier and more convenient in the practical work than the first one.

3. Consistency of Maximum Likelihood Estimator.

Let $x_1, -----, x_n$ be a random sample from a population with probability density function $f(x,\theta)$ and let θ^x denote the maximum likelihood estimator of the true value θ_0 of the parameter θ. Then we have to prove that

$$Pr\left\{|\theta^x - \theta_0| > \delta\right\} \to 0 \qquad \text{as } n \to \infty$$

where δ is any arbitrary small positive number. The following conditions are supposed to be satisfied:

(a) The derivatives

$$\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0} \quad \text{and} \quad \left(\frac{\partial^2 \log f}{\partial \theta^2}\right)_{\theta_0}$$

are finite and integrable over $(-\infty, \infty)$, and

$$\frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0} = E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right)_{\theta_0} .$$

(b) $\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f \, dx$ is finite and positive.

(c) $\sum_{i=2}^{\infty} \frac{1}{i!}\left(\theta^* - \theta_0\right)^i \left(\frac{\partial^{i+1} \log L}{\partial \theta^{i+1}}\right)_{\theta_0} \longrightarrow 0$.

<u>Proof</u>: By Taylor's theorem we have

$$\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta^*} = \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} + (\theta^* - \theta_0)\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0} + \text{-----}$$

In virtue of condition (c) we obtain

$$0 = \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} + (\theta^* - \theta_0)\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0}$$

ie.

$$\theta^* - \theta_0 = -\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} \Big/ \left(\frac{\partial^2 \log L}{\partial \theta^2}\right) \quad \text{------} \quad (1)$$

Now

$$E\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0} = \int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0} f \, dx = 0$$

&

$$V\left(\frac{\partial \log f}{\partial \theta}\right) = E\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} = k^2, \quad \text{say} .$$

Also we can show that

$$-E\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} = E\left(\frac{\partial^2 \log f}{\partial \theta}\right)_{\theta_0} .$$

By condition (a) we have

$$\frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0} = E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right)_{\theta_0} = -V\left(\frac{\partial \log f}{\partial \theta}\right) = -k^2,$$

we can write

$$\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} = \sum_1^n \left(\frac{\partial \log f_i}{\partial \theta}\right)_{\theta_0} .$$

From (1) we have

$$\Pr\left\{ |\theta^x - \theta_0| > \delta \right\} = \Pr\left\{ \left| - \frac{\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0}}{\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0}} \right| > \delta \right\}$$

$$= \Pr\left\{ \left| - \frac{\frac{1}{\sqrt{n}} \sum_1^n \left(\frac{\partial \log f_i}{\partial \theta}\right)_{\theta_0}}{\frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0}} \right| > \sqrt{n}\,\delta \right\}$$

$$\doteq \Pr\left\{ \left| \frac{\frac{1}{\sqrt{n}} \sum_1^n \left(\frac{\partial \log f_i}{\partial \theta}\right)_{\theta_0}}{k} \right| > \sqrt{n}\,k\,\delta \right\}.$$

In virtue of the central limit theorem we have

$$\frac{1}{k\sqrt{n}} \sum_1^n \left(\frac{\partial \log f_i}{\partial \theta}\right)_{\theta_0} \quad \text{distributed as } N(0,1).$$

Then

$$\Pr\left\{ |\theta^x - \theta_0| > \delta \right\} = \Pr\left\{ N(0,1) > k\delta\sqrt{n} \right\}$$

since n is large, then $k\delta\sqrt{n}$ will be large enough to make

$$\Pr\left\{ N(0,1) > k\delta\sqrt{n} \right\} \longrightarrow 0 .$$

Then

$$\Pr\left\{ |\theta^x - \theta_0| > \delta \right\} \longrightarrow 0 .$$

That is, θ^x is a consistent estimator of θ_0 .

Example 1.5 Let $x_1, \ldots, x_n$ be a random sample from a population distributed normally with unknown mean θ and variance σ^2 . Then the likelihood function is

$$L(x;\theta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum (x_i - \theta)^2 \right\}.$$

Taking the logarithm of both sides and differentiating with respect to θ , in order to estimate the value of θ , we get

$$\frac{\partial \log L}{\partial \theta} = -\frac{2}{2\sigma^2} \sum (x_i - \theta) .$$

The solution of

$$\frac{\partial \log L}{\partial \theta} = -\frac{1}{\sigma^2}\sum(x_i-\theta) = 0$$

is $\hat{\theta} = \bar{x}$; ie. the sample mean is the estimate of the population mean. Now we want to show that $\bar{x}$ is a consistent estimate to θ. We have by definition that if

$$\Pr\{|\bar{x}-\theta| > \delta\} \longrightarrow 0 \qquad \text{as} \quad n \longrightarrow \infty$$

where δ is a small positive number, then $\bar{x}$ is a consistent estimate to θ. Here

$$\Pr\{|\bar{x}-\theta| > \delta\} = \Pr\left\{\left(\frac{\bar{x}-\theta}{\frac{\sigma}{\sqrt{n}}}\right)^2 > \left(\frac{\delta\sqrt{n}}{\sigma}\right)^2\right\}$$

$$= \Pr\left\{\chi^2_{[1]} > \frac{\delta^2 n}{\sigma^2}\right\}$$

where $\frac{\sigma^2}{n}$ is the variance of $\bar{x}$. Since $\frac{\delta^2 n}{\sigma^2}$ is sufficiently large then

$$\Pr\left\{\chi^2_{[1]} > \frac{\delta^2 n}{\sigma^2}\right\} \longrightarrow 0 ,$$

ie.

$$\Pr\{|\bar{x}-\theta| > \delta\} \longrightarrow 0 .$$

Or

$$\Pr\{|\bar{x}-\theta| > \delta\} = \Pr\left\{\left|\frac{\bar{x}-\theta}{\frac{\sigma}{\sqrt{n}}}\right| > \frac{\delta\sqrt{n}}{\sigma}\right\} = \Pr\left\{N(0,1) > \frac{\delta\sqrt{n}}{\sigma}\right\}$$

Since $\frac{\delta\sqrt{n}}{\sigma}$ is sufficiently large then

$$\Pr\{|\bar{x}-\theta| > \delta\} \longrightarrow 0$$

Hence $\bar{x}$ is a consistent estimate to θ.

To find the maximum likelihood estimate of σ^2, the population variance we have to differentiate the logarithm of the likelihood function with respect to σ^2. Here we have

$$\log L = \text{constant} - \frac{1}{2} n \log \sigma^2 - \frac{1}{2\sigma^2}\sum(x-\bar{x})^2 ,$$

then

$$\frac{\partial \log L}{\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum(x-\bar{x})^2}{2\sigma^4} .$$

Equating the last equation to zero we get

$$\overset{x}{\sigma}{}^2 = \frac{\Sigma(x-\bar{x})^2}{n} ,$$

and since

$$S_x^2 = \frac{\Sigma(x-\bar{x})^2}{n-1} ,$$

is the unbiased estimate of σ^2 then

$$S_x^2 = \frac{n}{n-1}\overset{x}{\sigma}{}^2$$

where S_x^2 is the sample variance.

Now we must show that

$$Pr\left\{|S_x^2 - \sigma^2| > \delta\right\} \longrightarrow 0 .$$

Here

$$Pr\left\{|S_x^2 - \sigma^2| > \delta\right\} = Pr\left\{\left|\frac{S_x^2-\sigma^2}{\sqrt{\frac{2\sigma^4}{n}}}\right| > \frac{\delta\sqrt{n}}{\sigma^2\sqrt{2}}\right\}$$

$$= Pr\left\{N(0,1) > \frac{\delta\sqrt{n}}{\sigma^2\sqrt{2}}\right\} \longrightarrow 0 ,$$

since $\frac{\delta\sqrt{n}}{\sigma^2\sqrt{2}}$ is sufficiently large. Hence S_x^2 is a consistent estimate of σ^2 .

4. <u>The Maximum Likelihood Estimators are Asymptotically, Most Efficient, Normally Distributed and Unbiased.</u>

Let $x_1, ----, x_n$ be a random sample from a population having a probability density function $f(x,\theta)$ where θ is the parameter. Let $\overset{x}{\theta}$ be the maximum likelihood estimator of the true value θ_0 of the parameter θ .

To prove the properties mentioned above, the following conditions must be satisfied.

(a) $\overset{x}{\theta}$ is consistent .

(b) This condition is the estension of condition (a); that is

$$\sum_{i=2}^{\infty} \frac{1}{i!}(\overset{x}{\theta} - \theta_0)^i \left(\frac{\partial^{i+1}\log L}{\partial\theta^{i+1}}\right)_{\theta_0} \longrightarrow 0 ,$$

where L is the likelihood function.

(c) $\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2 f\,dx$ is finite and positive in some interval containing the true value θ_0 .

(d) The attainable minimum variance proved independently of any method of estimation is defined by

$$V(\theta_0) = 1 \Big/ n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f\,dx \; .$$

(e) $\frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right) = E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right)$ in some interval containing the true value θ_0 .

(f) $\frac{\partial \log f}{\partial \theta}$ and $\frac{\partial^2 \log f}{\partial \theta^2}$ are integrable over $(-\infty, \infty)$.

By Taylor's theorem we have

$$(\log L)_{\theta^*} = (\log L)_{\theta_0} + (\theta^* - \theta_0)\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} + \sum_{i=2}^{\infty}\frac{1}{i!}(\theta^* - \theta_0)^i\left(\frac{\partial^i \log L}{\partial \theta^i}\right)_{\theta_0}$$

Differentiating with respect to θ , we get

$$\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta^*} = \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} + (\theta^* - \theta_0)\left(\frac{\partial^2 \log L}{\partial \theta^2}\right) + \sum_{i=2}^{\infty}\frac{1}{i!}(\theta^* - \theta_0)^i\left(\frac{\partial^{i+1} \log L}{\partial \theta^{i+1}}\right)_{\theta_0}$$

In virtue of condition (b) and since $\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta^*} = 0$, we obtain

$$0 = \left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} + (\theta^* - \theta_0)\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0}$$

Then

$$\theta^* - \theta_0 = -\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} \Big/ \left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0} \qquad \text{------ (A)}$$

That is

$$\sqrt{n}\,(\theta^* - \theta_0) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial \log f_i}{\partial \theta}\right)_{\theta_0} \Big/ \frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0}$$

Here

$$E\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0} = \int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0} f\,dx = \int_{-\infty}^{\infty}\left(\frac{\partial f}{\partial \theta}\right)_{\theta_0} dx = 0$$

&

$$\frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0} = E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right)_{\theta_0} = \int_{-\infty}^{\infty}\left(\frac{\partial^2 \log f}{\partial \theta^2}\right)_{\theta_0} f\,dx$$

ie. $\frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0} = -\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f\,dx = -V\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0}$

In virtue of condition (d), we get

$$\frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0} = -\frac{1}{nV(\theta_0)} .$$

Then rewriting (A), we have

$$\sqrt{n}\,(\theta^x - \theta_0) = -\frac{\frac{1}{\sqrt{n}}\left[\sum_1^n\left(\frac{\partial \log f_i}{\partial \theta}\right) - 0\right]}{-\frac{1}{nV(\theta_0)}}$$

That is

$$\frac{1}{\sqrt{V(\theta_0)}}\,(\theta^x - \theta_0) = \sum_1^n\left(\frac{\partial \log f_i}{\partial \theta}\right)_{\theta_0} \Big/ \sqrt{n}\sqrt{\frac{1}{nV(\theta_0)}}$$

By central limit theorem

$$\sum_1^n\left(\frac{\partial \log f_i}{\partial \theta}\right)_{\theta_0} \Big/ \sqrt{n}\sqrt{\frac{1}{nV(\theta_0)}}$$

is distributed normally with zero mean and unit variance .

Then $\frac{1}{\sqrt{V(\theta_0)}}\,(\theta^x - \theta_0)$ is distributed as $N(0,1)$.

That is θ^x is distributed as $N(\theta_0, V(\theta_0))$.

That is means that θ^x is distributed normally. Since θ_0 is the mean of θ^x, then θ^x is unbiased. And since $V(\theta_0)$ is the minimum variance, then θ^x is most efficient.

We can show that the maximum likelihood estimators are unbiased and most efficient by proof differ but modified from the proof above.

We start from equation (A) above; ie.

$$\theta^x - \theta_0 = -\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0} \Big/ \left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0}$$

In virtue of condition (e) we have

$$\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0} = n E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right)_{\theta_0} = \text{constant} = C \, ,$$

then

$$E\left(\theta^{x}-\theta_0\right) = -\frac{E\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0}}{C} = -\frac{1}{C} \sum \left\{ E\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0} \right\} ,$$

and since

$$E\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0} = 0 \, ,$$

then

$$E\left(\theta^{x}\right) = E\left(\theta_0\right) = \theta_0$$

ie. θ^{x} is unbiased estimate of θ_0 .

Now to show that θ^{x} is most efficient, square both sides of the first equation above: we get

$$\left(\theta^{x}-\theta_0\right)^2 = \frac{\left[\left(\frac{\partial \log L}{\partial \theta}\right)_{\theta_0}\right]^2}{\left[\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\theta_0}\right]^2}$$

$$= \left[\sum\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0}\right]^2 \Big/ \left[n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f \, dx\right]^2 ,$$

then

$$E\left(\theta^{x}-\theta_0\right)^2 = E\left[\sum\left(\frac{\partial \log f}{\partial \theta}\right)_{\theta_0}\right]^2 \Big/ \left[n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f \, dx\right]^2 ,$$

$$= \sum\left[E\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0}\right] \Big/ \left[n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f \, dx\right]^2$$

$$= n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f \, dx \Big/ \left[n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f \, dx\right]^2$$

$$= 1 \Big/ n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\theta_0} f \, dx \, .$$

Since the R.H.S. is the minimum variance then the maximum likelihood estimate has the minimum variance; ie. it is most efficient.

Example 1.6: Let $x_1, ----, x_n$ be a random sample from a population distributed normally with mean θ and variance σ^2. We want to find the maximum likelihood estimate to the parameter θ and show that that estimate is unbiased, normally distributed and has the minimum variance. Here the likelihood function is

$$L(x;\theta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_1^n (x_i-\theta)^2\right\}.$$

By taking the logarithm and differentiating with respect to θ we get

$$\frac{\partial \log L}{\partial \theta} = -\frac{1}{\sigma^2}\sum_1^n (x_i-\theta).$$

The solution of

$$\frac{\partial \log L}{\partial \theta} = 0$$

is $\theta^* = \bar{x}$.

The moment generating function of x is defined by

$$M_x(t) = E(e^{xt}) = \int_{-\infty}^{\infty} e^{xt}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-\theta)^2\right\}$$

$$= \exp\left\{\tfrac{1}{2}t^2\sigma^2 + t\theta\right\}.$$

Since x is distributed normally with mean θ and variance σ^2, then

$$\exp\left\{\tfrac{1}{2}t^2\sigma^2 + t\theta\right\}$$

represents a normal distribution formula of a random variable with mean θ and variance σ^2. Then

$$M_{\bar{x}}(t) = M_{\frac{1}{n}\Sigma x}(t)$$

$$= \exp\left\{\tfrac{1}{2}t^2\frac{\sigma^2}{n^2} + -----+ \text{to } n \text{ terms} + \frac{n\theta t}{n}\right\}$$

$$= \exp\left\{\frac{1}{2}t^2\frac{\sigma^2}{n} + \theta t\right\}.$$

That is $\bar{x}$ is distributed normally with mean θ (unbiased) and variance $\frac{\sigma^2}{n}$. Now we have to show that $\frac{\sigma^2}{n}$ is the minimum variance. The following equality affords the minimum variance

$$V(\bar{x}) = \frac{1}{n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial\theta}\right)^2_{\theta_0=\bar{x}} f\,dx} .$$

Here

$$n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial\theta}\right)^2_{\theta_0=\bar{x}} f\,dx = n\int_{-\infty}^{\infty}\left\{-\frac{1}{\sigma^2}(x-\bar{x})\right\}^2 f\,dx$$

$$= \frac{n}{\sigma^4}\int_{-\infty}^{\infty}(x-\bar{x})^2 f\,dx$$

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

then

$$V(\bar{x}) = \frac{\sigma^2}{n} .$$

So $\bar{x}$ is a most efficient estimate to θ.

5. Successive Approximations to Efficient Estimators Using Maximum Likelihood

It often happens that maximum likelihood equations are difficult to be solved directly. In such cases we have to find by some inefficient method an initial estimate of the maximum likelihood. Then by successive approximations we obtain the efficient (maximum likelihood) estimate. Now we deduce the formula which is used to find the approximations.

Let θ^{x} be the maximum likelihood estimate and $\theta^{x(0)}$ be the initial estimate to θ^{x}, so $\theta^{x(i)}$ $i=1,2,\ldots,$ will be denoted the successive approximated estimates. Then we have by Taylor's theorem that

$$\left(\frac{\partial \log L}{\partial \theta}\right)_{\overset{x}{\theta}} \simeq \left(\frac{\partial \log L}{\partial \theta}\right)_{\overset{x}{\theta}^{(0)}} + (\overset{x}{\theta} - \overset{x}{\theta}^{(0)})\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\overset{x}{\theta}^{(0)}} .$$

Since the left hand side equal zero, then

$$\overset{x}{\theta} - \overset{x}{\theta}^{(0)} \simeq - \frac{\left(\frac{\partial \log L}{\partial \theta}\right)_{\overset{x}{\theta}^{(0)}}}{\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\overset{x}{\theta}^{(0)}}} .$$

If n is large, then by the law of large numbers we have

$$\frac{1}{n}\left(\frac{\partial^2 \log L}{\partial \theta^2}\right)_{\overset{x}{\theta}^{(0)}} = E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right)_{\overset{x}{\theta}^{(0)}} = -E\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\overset{x}{\theta}^{(0)}} = -\frac{I_{\overset{x}{\theta}^{(0)}}}{n} .$$

Hence

$$\overset{x}{\theta} - \overset{x}{\theta}^{(0)} = \left(\frac{\partial \log L}{\partial \theta}\right)_{\overset{x}{\theta}^{(0)}} \Big/ nE\left(\frac{\partial \log f}{\partial \theta}\right)^2_{\overset{x}{\theta}^{(0)}} ,$$

that is

$$\overset{x}{\theta} = \overset{x}{\theta}^{(0)} + I^{-1}_{\overset{x}{\theta}^{(0)}}\left(\frac{\partial \log L}{\partial \theta}\right)_{\overset{x}{\theta}^{(0)}} .$$

The last formula, may be written in general as

$$\overset{x}{\theta}^{(k+1)} = \overset{x}{\theta}^{(k)} + I^{-1}_{\overset{x}{\theta}^{(k)}}\left(\frac{\partial \log L}{\partial \theta}\right)_{\overset{x}{\theta}^{(k)}} .$$

When $\overset{x}{\theta}^{(0)}$ is very near from $\overset{x}{\theta}$, $I^{-1}_{\overset{x}{\theta}^{(0)}}$ will be used for all the approximations.

6. The Maximum Likelihood Estimator is Sufficient:

If there exists a sufficient estimator $t(x_1, ---, x_n)$, say, to the true value θ_0 of the parameter θ, then

$$L(x,\theta) = \prod_1^n f(x_i,\theta) = \prod_1^n \left\{ f_1(t,\theta)\ f_2(x) \right\}$$

$$= L_1(t,\theta)\, L_2(x) .$$

Then

$$\log L(x,\theta) = \log L_1(t,\theta) + \log L_2(x) .$$

Differentiating with respect to θ we get

$$\frac{\partial \log L(x,\theta)}{\partial \theta} = \frac{\partial \log L_1(t,\theta)}{\partial \theta} .$$

Here the solution of $\frac{\partial \log L}{\partial \theta} = 0$ is the solution of $\frac{\partial \log L_1}{\partial \theta} = 0$, and since $\frac{\partial \log L_1}{\partial \theta}$ involves the sufficient statistic $t(x_1, \ldots, x_n)$, then it will be the solution of $\frac{\partial \log L}{\partial \theta} = 0$. Since the maximum likelihood estimator is the solution of

$$\frac{\partial \log L}{\partial \theta} = 0 ,$$

therefore the maximum likelihood estimator is sufficient.

<u>Example 1.7</u> Consider the normal distribution with unknown mean θ and known variance σ^2. The likelihood function will be so that

$$L(x,\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum (x-\theta)^2 \right\} .$$

We can show that $\bar{x}$ is themaximum likelihood estimate of θ . In order to show that $\bar{x}$ is sufficient we must show that $L(x, \theta)$ can be factorised into two parts, one part dependent on $\bar{x}$ and θ , and the other part independent of θ . That is we have to show that

$$L(x, \theta) = L_1(\bar{x}, \theta) L_2(x),$$

or

$$\log L(x, \theta) = \log L_1(\bar{x}, \theta) + \log L_2(x).$$

We have

$$\log L(x, \theta) = C - \frac{1}{2\sigma^2} \sum (x - \theta)^2$$

where

$$C = \log\left\{1/(2\pi\sigma^2)^{\frac{n}{2}}\right\} \qquad \text{independent of } \theta .$$

Then

$$\log L(x, \theta) = \left(C - \frac{1}{2\sigma^2} \sum x^2\right) + \frac{1}{2\sigma^2}\left(2n\bar{x}\theta - n\theta^2\right),$$

we notice here that the first term in the R.H.S. isindependent of θ , and the second term is dependent on θ and $\bar{x}$.

Therefore $\bar{x}$ is a sufficient estimate for θ.

There is another way to show that the maximum likelihood estimator is sufficient. In our foregoing discussion about the sufficiency of the maximum likelihood estimator, we mentioned that if there exists a sufficient estimator, then

$$\frac{\partial \log L}{\partial \theta} = 0$$

will afford it; ie. $\frac{\partial \log L}{\partial \theta}$ must be dependent on θ and t (the sufficient statistic). Therefore our criterion of sufficiency is to show that

$$\frac{\partial \log L}{\partial \theta}$$

is dependent on θ and the statistic t.

In our example

$$\frac{\partial \log L}{\partial \theta} = \frac{1}{\sigma^2} n(\bar{x}-\theta)$$

which is dependent on θ and $\bar{x}$, therefore $\bar{x}$ is a sufficient estimator for θ.

<u>Example 1.8</u> Consider the distribution of Poisson, the parameter θ is unknown, then the likelihood function will be such that

$$L(x,\theta) = e^{-n\theta} \theta^{\sum_1^n x_i} \frac{1}{\prod_1^n x_i!}$$

$$= e^{-n\theta} \theta^{n\bar{x}} \frac{1}{\prod_1^n x_i!}.$$

Taking the logarithm and differentiating with respect to θ we get

$$\frac{\partial \log L}{\partial \theta} = n\left(\frac{\bar{x}}{\theta} - 1\right)$$

Hence

$$\frac{\partial \log L}{\partial \theta}$$

is dependent only on θ and $\bar{x}$, therefore $\bar{x}$ is a sufficient estimator for θ .

Example 1.9 In case of Binomial distribution with the parameter p , the likelihood function then will be such that

$$L(x,p) = \binom{n}{x} p^x (1-p)^{n-x} .$$

Taking the logarithm and differentiating with respect to p we get

$$\frac{\partial \log L}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p}$$

$$= \frac{x-np}{p(1-p)}$$

$$= \frac{n(\bar{x}-p)}{p(1-p)} .$$

Since $\bar{x}$ is the maximum likelihood estimator of p , then $\bar{x}$ is sufficient because

$$\frac{\partial \log L}{\partial p}$$

is dependent only on $\bar{x}$ and p .

Example 1.10 Consider the distribution of Type III to estimate the parameter α , where the parameter λ is known. The distribution of Type III is defined by

$$f(x,\alpha) = x^{\lambda-1} e^{-x/\alpha} \Big/ \Gamma(\lambda)\, \alpha^{\lambda} \qquad 0 \leqslant x \leqslant \infty$$

The likelihood function is then

$$L(x,\alpha) = \prod_{1}^{n} x_i^{(\lambda-1)} e^{-n\bar{x}/\alpha} \Big/ [\Gamma(\lambda)]^n \alpha^{n\lambda}$$

Taking the logarithm and differentiating with respect to α , we get

$$\frac{\partial \log L}{\partial \alpha} = n\lambda \left\{ \frac{\bar{x}/\lambda}{\alpha^2} - \frac{1}{\alpha} \right\}$$

where $\frac{\bar{x}}{\lambda}$ is the maximum likelihood estimate of α . Since $\frac{\partial \log L}{\partial \alpha}$ is dependent on $\frac{\bar{x}}{\lambda}$ and α only then the maximum likelihood estimate $\frac{\bar{x}}{\lambda}$ is sufficient.

CHAPTER II

SEVERAL PARAMETERS

1. Introduction:

In chapter I we discussed the problem for a single parameter. In this chapter we are dealing with several parameters; as a model let $x_1, ----, x_n$ be a random sample drawn from a population with joint frequency function $F(x_1, ---, x_n; \theta_1, --, \theta_m)$; that is there are m parameters to be required. Hereafter we denote $F(\underset{\sim}{x}; \underset{\sim}{\theta})$ instead of $F(x_1, ----, x_n; \theta_1, -----, \theta_m)$ and sometimes $\underset{\sim}{\theta}$ may be written as a column vector such that

$$\underset{\sim}{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}$$

2. The Amount of Information:

We have shown in Chapter I (1.4(a)) that the amount of information about the parameter θ supplied from the ~~statistic~~ observations is given by

$$n \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta} \right)^2 f \, dx$$

where f is the density function of a single observation and n is the sample size. In present case where θ is as several parameters the amount of information about these parameters supplied from the corresponding estimators is as a square matrix of order m whose (i,j)th element is

$$-\frac{1}{n} \underset{\text{divided by } n}{E} \left(\frac{\partial^2 \log F}{\partial \theta_i \partial \theta_j} \right) . \qquad i, j = 1, 2, ---, m$$

The inverse of this matrix is called the variance-covariance matrix of the estimators of the parameters $\theta_1, \theta_2, ----, \theta_m$.

3. Successive Approximations to Efficient Estimators Using M.L.:

In Chapter I section 5 we have shown that the formula used for the successive approximations is given by

$$\overset{\times}{\theta}^{(k+1)} = \overset{\times}{\theta}^{(k)} + I^{-1}_{\overset{\times}{\theta}^{(k)}} \left(\frac{\partial \log L}{\partial \theta} \right)_{\overset{\times}{\theta}^{(k)}}$$

In case of several parameters the formula becomes such that

$$\begin{bmatrix} \overset{\times}{\theta}_1^{(k+1)} \\ \vdots \\ \overset{\times}{\theta}_m^{(k+1)} \end{bmatrix} = \begin{bmatrix} \overset{\times}{\theta}_1^{(k)} \\ \vdots \\ \overset{\times}{\theta}_m^{(k)} \end{bmatrix} + \underset{\sim}{I}^{-1}_{\underset{\sim}{\overset{\times}{\theta}}^{(k)}} \begin{bmatrix} \frac{\partial \log F}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log F}{\partial \theta_m} \end{bmatrix}_{\underset{\sim}{\overset{\times}{\theta}}^{(k)}}$$

where F is itself the likelihood function and $\underset{\sim}{I}^{-1}_{\underset{\sim}{\overset{\times}{\theta}}^{(k)}}$ is a square matrix of order m whose (i,j)th element is

$$-E\left(\frac{\partial^2 \log F}{\partial \theta_i \partial \theta_j} \right). \qquad i, j = 1, 2, ---, m$$

If $\underset{\sim}{\overset{\times}{\theta}}^{(0)}$ the initial estimate is very near to $\underset{\sim}{\overset{\times}{\theta}}$ the maximum likelihood estimate, then $\underset{\sim}{I}^{-1}_{\underset{\sim}{\overset{\times}{\theta}}^{(k)}}$ will be replaced by $\underset{\sim}{I}^{-1}_{\underset{\sim}{\overset{\times}{\theta}}^{(0)}}$ for all the process of the approximations. Note: The application will be shown in chapter 3.

4. Distribution Admitting Sufficient Statistics:

Koopman (1936) has shown that if the distribution function $h(x_1, ---, x_n; \theta_1, ------, \theta_m)$ is continuous and not zero over some continuous range of the θ's, and $\frac{\partial h}{\partial x}$ exists, then the necessary and sufficient form of the function h to admit the sufficient statistics, is

$$h = \exp\left\{ P_1(\theta) q_1(x) + --- + P_m(\theta) q_m(x) + P_0(\theta) + q_0(x) \right\},$$

where $P_i(\theta)$ and $q_i(x)$, $i=0,1,2,\ldots,m$ are functions of θ and x respectively.

Example 2.1 Consider the normal distribution with unknown mean θ and variance σ^2. The joint frequency function is then

$$F(x;\theta,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2\sigma^2}\sum(x-\theta)^2\right\},$$

ie.

$$F(x;\theta,\sigma^2)=\frac{1}{(2\pi)^{\frac{n}{2}}\ (\sigma^2)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2\sigma^2}(n-1)S_x^2-\frac{1}{2\sigma^2}n(\bar{x}-\theta)^2\right\}$$

where

$$S_x^2=\frac{\sum(x-\bar{x})^2}{n-1},$$

and

$$\bar{x}=\frac{1}{n}\sum x$$

Here

$$P_1(\theta)\,q_1(x)=-\frac{1}{2\sigma^2}n(\bar{x}-\theta)^2,$$

$$P_2(\theta)\,q_2(x)=-\frac{1}{2\sigma^2}(n-1)S_x^2,$$

$$P_0(\theta)=-\frac{1}{2}n\log\sigma^2,$$

$$q_0(x)=-\frac{1}{2}n\log 2\pi.$$

Therefore the normal distribution with unknown mean θ and variance σ^2 admits sufficient estimators for θ and σ^2.

Example 2.2 Consider the Type III distribution

$$f(x;\rho,\sigma)=\frac{1}{\sigma\Gamma(\rho)}\left(\frac{x-\alpha}{\sigma}\right)^{\rho-1}\exp\left\{-\left(\frac{x-\alpha}{\sigma}\right)\right\},$$

where α is known and $\alpha \leqslant x \leq \infty$. Then

$$F(x;\rho,\sigma) = \frac{1}{\sigma^n}\frac{1}{[\Gamma(\rho)]^n}\left(\frac{x-\alpha}{\sigma}\right)^{n(\rho-1)} \exp\left\{-\sum\left(\frac{x-\alpha}{\sigma}\right)\right\}$$

Here

$$P_1(\theta)\, q_1(x) = -\sum\left(\frac{x-\alpha}{\sigma}\right)$$

$$P_2(\theta)\, q_2(x) = n(\rho-1)\log(x-\alpha)$$

$$P_0(\theta) = -n\rho\log\sigma - n\log\Gamma(\rho)$$

Therefore there are sufficient estimators to ρ and σ. In this distribution it is clear that if α is unknown there are no sufficient estimators, even if σ and ρ are known.

5. Maximum Likelihood Estimators are sufficient

Let $x_1, ----, x_n$ be a random sample from a population with probability density function $f(x;\theta_1,---,\theta_m)$ and let $t_1, ------, t_m$ be sufficient estimators to $\theta_1, ------, \theta_m$ respectively. Then the likelihood function will be factorised such that

$$L(\underset{\sim}{x};\underset{\sim}{\theta}) = L_1(\underset{\sim}{t};\underset{\sim}{\theta})\, L_2(\underset{\sim}{x})$$

where $L_1(\underset{\sim}{t};\underset{\sim}{\theta})$ is dependent on θ and t only, and $L_2(\underset{\sim}{x})$ is independent of θ.

Differentiating with respect to θ_i we get

$$\frac{\partial L(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial\theta_i} = L_2(\underset{\sim}{x})\,\frac{\partial L_1(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial\theta_i}, \qquad i=1,--,m$$

since the solution of the equations

$$\frac{\partial L_1(\underset{\sim}{t};\underset{\sim}{\theta})}{\partial \theta_i} = 0$$

affords sufficient estimators then the solution of the equations

$$\frac{\partial L(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial \theta_i} = 0$$

affords sufficient estimators too. Since the solution of the equations

$$\frac{\partial L(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial \theta_i} = 0$$

affords the maximum likelihood estimators, therefore they are sufficient.

Example 2.3 Consider the normal distribution with unknown mean μ and variance σ^2. The likelihood function is then

$$L(x;\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum(x-\mu)^2\right\}$$

Differentiating the logarithm of both sides with respect to μ we get

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum(x-\mu) \; ,$$

since $\sigma^2 = S_x^2 = \dfrac{\sum(x-\bar{x})^2}{n-1}$, then

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{S_x^2}\sum(x-\mu)$$

That is

$$\frac{\partial \log L}{\partial \mu}$$

is dependent only on $\bar{x}$ and μ, therefore the maximum likelihood estimate $\bar{x}$ is sufficient for μ.

Now differentiating the logarithm of both sides with respect to σ^2 we get

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x-\bar{x})^2 ,$$

that is

$$\frac{\partial \log L}{\partial \sigma^2}$$

is dependent only on S^2_x and σ^2, therefore S^2_x is sufficient estimator for σ^2. Finally $\bar{x}$ and S^2_x are sufficient estimators for μ and σ^2.

6. Simultaneous Estimation of Several Parameters

We have shown in Chapter I section 2 that if $L(x,\theta)$ is the likelihood function then the estimator of θ will be the solution of the equation

$$\frac{\partial \log L(x,\theta)}{\partial \theta} = 0 ,$$

so in the case of several parameters the estimators of these parameters will be the solution of the equations

$$\frac{\partial \log F(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial \theta_i} = 0 \qquad i = 1, ---, m$$

where $F(\underset{\sim}{x};\underset{\sim}{\theta})$ itself represents the likelihood function as defined in section I of this chapter.

Example 2.4 Consider the normal distribution with unknown mean α and variance σ^2. The likelihood function is then

$$F(x;\alpha,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum(x-\alpha)^2\right\}$$

Then

$$\log F = \text{constant} - \tfrac{1}{2} n \log \sigma^2 - \frac{1}{2\sigma^2} \sum (x-\alpha)^2 .$$

Differentiating with respect to α we get

$$\frac{\partial \log F}{\partial \alpha} = \frac{1}{\sigma^2} \sum (x-\alpha) ,$$

then the solution of

$$\frac{\partial \log F}{\partial \alpha} = 0$$

is $\overset{\times}{\alpha} = \bar{x}$; ie. the maximum likelihood estimate of α is the sample mean $\bar{x}$. Now we differentiate with respect to σ^2

$$\frac{\partial \log F}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (x-\alpha)^2$$

$$= -\frac{1}{2\sigma^2} \left[n - \frac{1}{\sigma^2} \sum (x-\bar{x})^2 \right]$$

Equating to zero we obtain

$$\overset{\times}{\sigma}^2 = \frac{1}{n} \sum (x-\bar{x})^2 .$$

It is worth while to find the amount of information on the parameters α and σ^2 supplied from the maximum likelihood estimators as illustration to section 2, chapter II. The (i,j)th element of the matrix which represents the amount of information is given by

$$-\frac{1}{n} E\left(\frac{\partial^2 \log F}{\partial \theta_i \partial \theta_j} \right) \qquad i,j = 1,2 .$$

Here

$$-E\left(\frac{\partial^2 \log F}{\partial \alpha^2} \right) = \frac{n}{\sigma^2} ,$$

$$-E\left(\frac{\partial^2 \log F}{\partial\alpha\,\partial\sigma^2}\right) = \frac{1}{\sigma^4}\sum(x-\alpha) = 0$$

$$-E\left(\frac{\partial^2 \log F}{\partial(\sigma^2)^2}\right) = -\left[\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum(x-\bar{x})^2\right]$$

$$= -\left[\frac{n}{2\sigma^4} - \frac{n}{\sigma^4}\right]$$

$$= \frac{n}{2\sigma^4}\,.$$

Then the amount of information is given by

$$\frac{1}{n}\begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

and the variance-covariance matrix is then

$$\frac{1}{n}\begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}^{-1} = \frac{1}{n}\left(\frac{1}{2\sigma^6}\right)^{-1}\begin{bmatrix} \frac{1}{2\sigma^4} & 0 \\ 0 & \frac{1}{\sigma^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

that is the variances of $\overset{\times}{\alpha}$ and $\overset{\times}{\sigma}{}^2$ are $\frac{\sigma^2}{n}$ and $\frac{2\sigma^4}{n}$ respectively and the covariance of $\overset{\times}{\alpha}$ and $\overset{\times}{\sigma}{}^2$ is zero, ie. the correlation coefficient between $\overset{\times}{\alpha}$ and $\overset{\times}{\sigma}{}^2$ is zero.

Example 2.5 Consider the distribution of the bivariate normal form, ie.

$$F(x;\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$$

$$= \frac{1}{\left[2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}\right]^n} \exp\left\{-\frac{1}{2(1-\rho^2)}\sum\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right.$$

then

$$\log F = \text{constant} - \tfrac{1}{2}n\log\sigma_1^2 - \tfrac{1}{2}n\log\sigma_2^2 - \frac{1}{2(1-\rho^2)}\sum\left\{\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right\} - \tfrac{1}{2}n\log(1-\rho^2).$$

It can be shown that the solution of the equations

$$\frac{\partial \log F}{\partial \theta_i} = 0 \qquad i = 1, 2, ---, 5$$

where $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ are $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ respectively, gives us the following estimators

$$\overset{\times}{\mu}_1 = \bar{x}, \qquad \overset{\times}{\mu}_2 = \bar{y}, \qquad \overset{\times}{\sigma}_1^2 = \frac{1}{n}\sum(x-\bar{x})^2,$$

$$\overset{\times}{\sigma}_2^2 = \frac{1}{n}\sum(y-\bar{y})^2, \qquad \overset{\times}{\rho}\,\overset{\times}{\sigma}_1\,\overset{\times}{\sigma}_2 = \frac{1}{n}\sum(x-\bar{x})(y-\bar{y}).$$

To obtain the amount of information we must find the elements of the representative matrix

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial \mu_1^2}\right) = \frac{1}{\sigma_1^2(1-\rho^2)}, \qquad -\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial \mu_2^2}\right) = \frac{1}{\sigma_2^2(1-\rho^2)},$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial(\sigma_1^2)^2}\right)=\frac{1}{\sigma_1^4}\,\frac{1-2\rho^2+3\rho^4}{1-\rho^2}\,,$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial(\sigma_2^2)^2}\right)=\frac{1}{\sigma_2^4}\,\frac{1-2\rho^2+3\rho^4}{1-\rho^2}\,,$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial(\rho^2)}\right)=\frac{(1+12\rho^2+3\rho^4)}{(1-\rho^2)^3}\,,$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\mu_1\partial\mu_2}\right)=\frac{2\rho}{\sigma_1\sigma_2}\,,\qquad -\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\mu_1\partial\sigma_1^2}\right)=0\,,$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\mu_1\partial\sigma_2^2}\right)=0\,,\qquad -\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\mu_1\partial\rho}\right)=0\,,$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\mu_2\partial\sigma_2^2}\right)=0\,,\qquad -\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\mu_2\partial\sigma_1^2}\right)=0\,,$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\mu_2\partial\rho}\right)=0\,,\qquad -\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\sigma_1^2\partial\sigma_2^2}\right)=\frac{\rho^2}{\sigma_1^2\sigma_2^2}\,,$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\sigma_1^2\partial\rho}\right)=-\frac{\rho^3(2-\rho^2)}{\sigma_1^2(1-\rho^2)^2}\,,\qquad -\frac{1}{n}E\left(\frac{\partial^2 \log F}{\partial\sigma_2^2\partial\rho}\right)=-\frac{\rho^3(2-\rho^2)}{\sigma_2^2(1-\rho^2)^2}\,.$$

Then the amount of information is given by the following symmetrical square matrix

$$
\begin{bmatrix}
\frac{1}{\sigma_1^2(1-\rho^2)} & \frac{2\rho}{\sigma_1\sigma_2} & 0 & 0 & 0 \\
\frac{2\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2(1-\rho^2)} & 0 & 0 & 0 \\
0 & 0 & \frac{1-2\rho^2+3\rho^4}{\sigma_1^4(1-\rho^2)} & \frac{\rho^2}{\sigma_1^2\sigma_2^2} & -\frac{\rho^3(2-\rho^2)}{\sigma_1^2(1-\rho^2)^2} \\
0 & 0 & \frac{\rho^2}{\sigma_1^2\sigma_2^2} & \frac{1-2\rho^2+3\rho^4}{\sigma_2^4(1-\rho^2)} & -\frac{\rho^3(2-\rho^2)}{\sigma_2^2(1-\rho^2)^2} \\
0 & 0 & -\frac{\rho^3(2-\rho^2)}{\sigma_1^2(1-\rho^2)^2} & -\frac{\rho^3(2-\rho^2)}{\sigma_2^2(1-\rho^2)^2} & \frac{1+12\rho^2+3\rho^4}{(1-\rho^2)^3}
\end{bmatrix}
$$

The variance-covariance matrix of the estimators $\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\rho}$ is given by the inverse of the matrix above divided by n .

7. Wald Technique:

The Wald technique for solving the maximum likelihood equations is related to his test. This test is used to know whether the unrestricted estimates of the unknown parameters satisfy some relationships which specify the null hypothesis. Thus the idea of Wald technique is to estimate the unrestricted parameters of maximum likelihood equations.

Let $x_1, ---, x_n$ be a random sample from a population with probability density function $f(x; \theta_1, ----, \theta_m)$, where $\theta_1, \theta_2, ----, \theta_m$ are unknown parameters. Then the estimates of the unrestricted parameters will be the solution of the equations

$$\frac{\partial \log L}{\partial \theta_i} = 0, \qquad i = 1, 2, ----, m$$

where L denotes the likelihood function. If these equations are difficult to solve we apply the successive approximation proccesses (section 3, chapter II) to find the maximum likelihood estimates.

If the restrictions $k (< m)$ which specify the null hypothesis are

$$h_1(\theta) = h_2(\theta) = ----- = h_k(\theta) = 0$$

then the Wald test which determines whether the unrestricted maximum likelihood estimates satisfy these restrictions, is based on the statistic

$$n\, h'(\hat{\theta}) \left[H'_{\hat{\theta}} \left(\tfrac{1}{n} \underset{\sim}{I}_{\hat{\theta}} \right)^{-1} H_{\hat{\theta}} \right]^{-1} h(\hat{\theta})$$

which is distributed as $\chi^2_{[k]}$, where $\left(\frac{1}{n} \underset{\sim}{I}_{\hat{\theta}} \right)$ is the information matrix whose (i,j)th element is $-\frac{1}{n} E\left(\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right)$ i, j = $1, ----, m$; $h(\theta)$ is the $k-$ column vector whose ith element is $h_i(\theta)$ and H_θ is the $m \times k$ matrix whose (i,j)th element is $\partial h_j(\theta) / \partial \theta_i$. If $\chi^2_{[k]} < \chi^2$ we accept the null hypothesis and we reject it otherwise, where χ^2 is obtainable from the statistical tables with the corresponding degrees of freedom.

8. Lagrange Multiplier Technique

This technique is related to the test of the null hypothesis which says whether the restricted estimates of the unknown parameters nearly maximize the likelihood functions. In virtue of the foregoing mentioned the idea of the Lagrange-multiplier technique will be the procedure for estimating the restricted parameters of the restricted likelihood equations.

Let $x_1, x_2, -----, x_m$ be a random sample from a population with probability density function $f(x; \theta_1, ---, \theta_m)$ where $\theta_1, \theta_2, --, \theta_{m-1}$ and θ_m are unknown parameters, and let there be $k (< m)$ restrictions in the form

$$h_1(\theta) = h_2(\theta) = --- = h_k(\theta) = 0$$

then the estimates of the restricted parameters will be the solution of the equations

$$\frac{1}{n} \frac{\partial \log L}{\partial \theta_i} + \sum \lambda_j \frac{\partial h_j(\theta)}{\partial \theta_i} = 0 \qquad (i = 1, 2, --, m),$$

$$h_j(\theta) = 0 \qquad (j = 1, 2, --, k),$$

where $\lambda_1, \lambda_2, --, \lambda_k$ are Lagrange multipliers, and L is the likelihood function.

Usually, in practice, these equations are difficult to solve, so in such cases we use the successive approximations procedure (section 3, chapter II) to calculate the maximum likelihood estimates. Here the successive approximation form will be such that

$$\begin{bmatrix} \overset{\otimes}{\theta}{}_1^{\ell+1} \\ \vdots \\ \overset{\otimes}{\theta}{}_m^{\ell+1} \\ \lambda_1^{\ell+1} \\ \vdots \\ \lambda_k^{\ell+1} \end{bmatrix} = \begin{bmatrix} \overset{\otimes}{\theta}{}_1^{\ell} \\ \vdots \\ \overset{\otimes}{\theta}{}_m^{\ell} \\ \lambda_1^{\ell} \\ \vdots \\ \lambda_k^{\ell} \end{bmatrix} + \begin{bmatrix} \frac{1}{n} \underset{\sim}{I}_\theta & -H_\theta \\ -H'_\theta & 0 \end{bmatrix}^{-1}_{\overset{(0)}{\theta}} \begin{bmatrix} \frac{1}{n} \frac{\partial \log L}{\partial \theta_1} + \sum_j \lambda_j \frac{\partial h_j}{\partial \theta_1} \\ \vdots \\ \frac{1}{n} \frac{\partial \log L}{\partial \theta_m} + \sum_j \lambda_j \frac{\partial h_j}{\partial \theta_m} \\ h_1(\theta) \\ \vdots \\ h_k(\theta) \end{bmatrix}_{\overset{\otimes}{\theta}{}^{\ell}}$$

where $\ell \geq 1$

where $\frac{1}{n}\underset{\sim}{I}_\theta$ and H_θ are as defined in section 7 of this chapter. For if

$$\begin{bmatrix} \frac{1}{n}\underset{\sim}{I}_\theta & -H_\theta \\ -H_\theta & 0 \end{bmatrix}^{-1}_{\overset{(0)}{\theta}} = \begin{bmatrix} A_\theta & B_\theta \\ B'_\theta & C_\theta \end{bmatrix}_{\overset{(0)}{\theta}}$$

then $\frac{1}{n} A_{\overset{(0)}{\theta}}$ will be the variance-covariance matrix of the restricted maximum likelihood estimates.

There is a very useful method to find the inverse of the matrix

$$\begin{bmatrix} \frac{1}{n}\underset{\sim}{I} & -H \\ -H' & 0 \end{bmatrix}$$

<u>The Procedure:</u>

1) Obtain $(\frac{1}{n}\underset{\sim}{I})^{-1}$.
2) Compute $H'(\frac{1}{n}\underset{\sim}{I})^{-1}$ and $H'(\frac{1}{n}\underset{\sim}{I})^{-1}H$.
3) Obtain $[H'(\frac{1}{n}\underset{\sim}{I})^{-1}H]^{-1} = -C$
4) Compute $B' = C[H'(\frac{1}{n}\underset{\sim}{I})^{-1}]$
5) Compute $A = (\frac{1}{n}\underset{\sim}{I})^{-1} + B[H'(\frac{1}{n}\underset{\sim}{I})^{-1}]$; The last matrix is symmetrical, and this property gives us good check on our computation.

The Lagrange-multiplier test is defined by the statistic

$$\frac{1}{n}\begin{bmatrix} \frac{\partial \log L}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L}{\partial \theta_m} \end{bmatrix}'_{\overset{\otimes}{\theta}} \left(\frac{1}{n}\underset{\sim}{I}_{\overset{\otimes}{\theta}}\right)^{-1} \begin{bmatrix} \frac{\partial \log L}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L}{\partial \theta_m} \end{bmatrix}_{\overset{\otimes}{\theta}}$$

where $\overset{\circledR}{\theta}$ is the restricted maximum likelihood estimate. This statistic is distributed as $\chi^2_{[k]}$, therefore if $\chi^2_{[k]} < \chi^2$ we accept the null hypothesis and we reject it otherwise, where χ^2 is obtainable from the statistical table with the corresponding degrees of freedom.

9. Singular Information Matrices:

In both of the previous techniques the information matrix was non-singular because it is related to the identifiability parameters. But some-times the information matrix is singular in a case when the unknown parameters is identifiable by some imposed restrictions. In such cases we have to do some modifications to make a non-singular matrix.

Let $\theta_1, \theta_2, ---, \theta_m$ be unknown parameters with k restrictions in the form

$$h_1(\theta) = h_2(\theta) = ---- = h_k(\theta) = 0$$

and let there be $d(<k)$ restrictions which make the m parameters identifiable, then the $(k-d)$ restrictions will specify the null hypothesis. Now, the $m \times k$ matrix H_θ whose (i,j)th element is $\frac{\partial h_j(\theta)}{\partial \theta_i}$ could be partitioned into $[H_{1\theta} \; H_{2\theta}]$ where $H_{1\theta}$ is $m \times d$ matrix whose (i,j)th element is $\frac{\partial h_j(\theta)}{\partial \theta_i}$ then the matrix $[\frac{1}{n} \underset{\sim}{I}_\theta + H_{1\theta} H'_{1\theta}]$ will be non-singular. Therefore in such cases we have to replace $[\frac{1}{n} \underset{\sim}{I}_\theta + H_{1\theta} H'_{1\theta}]$ instead of $\frac{1}{n} \underset{\sim}{I}_\theta$ and so the successive approximations procedure will be in the following forms

$$\begin{bmatrix} \overset{\times}{\theta}{}_1^{\ell+1} \\ \vdots \\ \overset{\times}{\theta}{}_m^{\ell+1} \end{bmatrix} = \begin{bmatrix} \overset{\times}{\theta}{}_1^{\ell} \\ \vdots \\ \overset{\times}{\theta}{}_m^{\ell} \end{bmatrix} + \left[\frac{1}{n} \underset{\sim}{I}_\theta + H_{1\theta} H'_{1\theta} \right]^{-1}_{\overset{(0)}{\theta}} \begin{bmatrix} \frac{\partial \log L}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L}{\partial \theta_m} \end{bmatrix}_{\overset{\times}{\theta}{}^{\ell}}$$

for the Wald technique, and

$$\begin{bmatrix} \overset{\otimes}{\theta}{}_1^{\ell+1} \\ \vdots \\ \overset{\otimes}{\theta}{}_m^{\ell+1} \\ \lambda_1^{\ell+1} \\ \vdots \\ \lambda_k^{\ell+1} \end{bmatrix} = \begin{bmatrix} \overset{\otimes}{\theta}{}_1^{\ell} \\ \vdots \\ \overset{\otimes}{\theta}{}_m^{\ell} \\ \lambda_1^{\ell+1} \\ \vdots \\ \lambda_k^{\ell+1} \end{bmatrix} + \begin{bmatrix} \frac{1}{n} \underset{\sim}{I}_\theta + H_{1\theta} H'_{1\theta} & -H_{1\theta} \\ -H'_{1\theta} & 0 \end{bmatrix}^{-1}_{\overset{(0)}{\theta}} \begin{bmatrix} \frac{1}{n} \frac{\partial \log L}{\partial \theta_1} + \sum_j \lambda_j \frac{\partial h_j}{\partial \theta_1} \\ \vdots \\ \frac{1}{n} \frac{\partial \log L}{\partial \theta_m} + \sum_j \lambda_j \frac{\partial h_j}{\partial \theta_m} \\ h_1(\theta) \\ \vdots \\ h_k(\theta) \end{bmatrix}$$

for the Lagrange-multiplier technique.

The statistics of Wald and Lagrange-multiplier tests for the null hypothesis, which says whether the unknown parameters satisfy the $(k-d)$ restrictions, will become

$$n\, h'(\overset{\times}{\theta}) \left[H'_{\overset{\times}{\theta}} \left(\frac{1}{n} \underset{\sim}{I}_{\overset{\times}{\theta}} + H_{1\overset{\times}{\theta}} H'_{1\overset{\times}{\theta}} \right)^{-1} H_{\overset{\times}{\theta}} \right]^{-1} h(\overset{\times}{\theta}) \,,$$

and

$$\frac{1}{n} \begin{bmatrix} \frac{\partial \log L}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L}{\partial \theta_m} \end{bmatrix}'_{\overset{\otimes}{\theta}} \left[\frac{1}{n} \underset{\sim}{I}_\theta + H_{1\theta} H'_{1\theta} \right]^{-1}_{\overset{\otimes}{\theta}} \begin{bmatrix} \frac{\partial \log L}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L}{\partial \theta_m} \end{bmatrix}_{\overset{\otimes}{\theta}}$$

respectively, and each statistic is distributed as $\chi^2_{[k-d]}$.

The estimate of the variance-covariance matrix of $\overset{\times}{\theta}_1, ---, \overset{\times}{\theta}_m$ will be given by $\frac{1}{n}\left[\frac{1}{n}\underset{\sim}{I}_\theta + H_{1\theta}H'_{1\theta}\right]^{-1}_{\theta^{(0)}}$ and so $\frac{1}{n}\left[\frac{1}{n}\underset{\sim}{I}_\theta + H_{1\theta}H'_{1\theta}\right]^{-1}_{\overset{\times}{\theta}}$ will be a better such estimate. If

$$\begin{bmatrix} \frac{1}{n}\underset{\sim}{I}_\theta + H_{1\theta}H'_{1\theta} & -H_\theta \\ -H'_\theta & 0 \end{bmatrix}^{-1}_{\theta^{(0)}} = \begin{bmatrix} \bar{A}_\theta & \bar{B}_\theta \\ \bar{B}'_\theta & \bar{C}_\theta \end{bmatrix}_{\theta^{(0)}}$$

then $\frac{1}{n}\bar{A}_{\theta^{(0)}}$ will be the estimate of the variance-covariance matrix of $\theta^{\otimes}_1, ---, \theta^{\otimes}_m$ and $\frac{1}{n}\bar{A}_{\theta^{\otimes}}$ will be a better such estimate.

10. Maximum Likelihood Estimates of the Mean and Variance of Normal Populations from Truncated Samples.

Let μ and σ^2 be the mean and variance of a normal population. Let x_0 be the truncated point measured on the original scale of the variate x (the variate of the complete distribution) and ξ be the truncation point measured in standard units of the complete distribution. Then we can write μ such that

$$\mu = x_0 - \sigma\xi$$

that is

$$x_0 = \mu + \sigma\xi \quad .$$

Then the probability density function of the variate $x'(= x - x_0)$ in the truncated normal distribution will be such that

$$f(x') = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \div I_0(\xi)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x'+\sigma\xi}{\sigma}\right)^2} \div I_0(\xi) \ ,$$

where

$$I_0(\xi) = \int_{\xi}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \quad .$$

Hereafter we will abbreviate $I_n(\xi)$ to I_n. The likelihood function of x' is then

$$L(x') = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2}\sum_1^n\left(\frac{x'+\sigma\xi}{\sigma}\right)^2} \div [I_0]^n \ ,$$

where n is the number of the known measured observations x_i'; $i=1,---,n$. Then

$$\log L(x') = \text{constant} - n\log\sigma - \frac{1}{2}\sum_1^n\left(\frac{x'+\sigma\xi}{\sigma}\right)^2 - n\log I_0 \quad .$$

Differentiating with respect to ξ and σ we get

$$\frac{\partial \log L}{\partial \xi} = -\sum_1^n\left(\frac{x'+\sigma\xi}{\sigma}\right) - \frac{n}{I_0}\frac{\partial I_0}{\partial \xi}$$

$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum_1^n\left(\frac{x'^2}{\sigma} + x'\xi\right) \quad .$$

Then the maximum likelihood estimates of ξ and σ will be the solution of

$$\sum_1^n\left(\frac{x'+\sigma\xi}{\sigma}\right) + \frac{n}{I_0}\frac{\partial I_0}{\partial \xi} = 0 \qquad \text{--------} \quad (1)$$

$$\frac{1}{\sigma^2}\sum_1^n\left(\frac{x'^2}{\sigma} + x'\xi\right) - \frac{n}{\sigma} = 0 \qquad \text{--------} \quad (2)$$

Since, by definition

$$I_n = \frac{1}{\sqrt{2\pi}}\int_\xi^\infty \frac{(t-\xi)^n}{n!}e^{-\frac{1}{2}t^2}\,dt \quad ,$$

we get

$$(n+1)I_{n+1} + \xi I_n - I_{n-1} = 0 \qquad \text{-------} \quad (3)$$

and

$$\frac{\partial I_n}{\partial \xi} = -I_{n-1}$$

ie.

$$\frac{\partial I_0}{\partial \xi} = -I_{-1} \quad .$$

Hence the equation (1) and (2) will be such that

$$\frac{1}{\sigma}\sum_1^n x' + n\xi - \frac{n}{I_0} I_{-1} = 0 \qquad \text{------ (1)}'$$

$$\frac{1}{\sigma^3}\sum_1^n x'^2 + \frac{\xi \sum_1^n x'}{\sigma^2} - \frac{n}{\sigma} = 0 \qquad \text{------ (2)}'$$

From equation (3) using $I_{-1} = I_1 + \xi I_0$, we get

$$n\bar{x}' - \sigma n \frac{I_1}{I_0} = 0 \qquad \text{------- (1)}''$$

$$\sum_1^n x'^2 + \sigma \xi n \bar{x}' - n\sigma^2 = 0 \qquad \text{------ (2)}''$$

From equation $(1)''$ we get

$$\bar{x}' = \sigma \frac{I_1}{I_0} \qquad \text{------- (4)}$$

Substituting the value of $\bar{x}'$ in equation $(2)''$ we get

$$\sum_1^n x'^2 + \sigma^2 \xi n \frac{I_1}{I_0} - n\sigma^2 = 0 \; .$$

Hence

$$\sum_1^n x'^2 = \frac{1}{I_0} n\sigma^2 (I_0 - \xi I_1)$$

From equation (3) using $2I_2 = I_0 - \xi I_1$, we get

$$\sum_1^n x'^2 = \frac{2I_2}{I_0} n\sigma^2 \qquad \text{------- (5)}$$

Substituting the value of σ obtained from equation (4), in equation (5) we get

$$\sum_1^n x'^2 = \frac{2I_2}{I_0} n\bar{x}'^2 \frac{I_0^2}{I_1^2}$$

$$= \bar{x}'^2 \frac{2I_2 I_0}{I_1^2}$$

ie.

$$\frac{n\sum_1^n x'^2}{\left(\sum_1^n x'\right)^2} = \frac{2I_2I_0}{I_1^2}$$

Since the quantity in the left side is known, then the value of ξ corresponding to $\frac{2I_2I_0}{I_1^2}$ will be obtainable from the "Mathematical Tables" Vol. 1 of the British Association for the Advancement of Science. Also from the tables mentioned above we find the values of I_0 and I_1 corresponding to the value of ξ. By substituting the values of I_0 and I_1 in equation (4) we obtain the value of σ. Finally substituting the values of x_0, σ and ξ in

$$\mu = x_0 - \sigma\xi$$

we get the value of μ.

The variance-covariance matrix of ξ and σ is given by

$$\begin{bmatrix} -E\left(\frac{\partial^2 \log L}{\partial\sigma^2}\right) & -E\left(\frac{\partial^2 \log L}{\partial\sigma\,\partial\xi}\right) \\ -E\left(\frac{\partial^2 \log L}{\partial\sigma\,\partial\xi}\right) & -E\left(\frac{\partial^2 \log L}{\partial\xi^2}\right) \end{bmatrix}^{-1}$$

Here

$$-E\left(\frac{\partial^2 \log L}{\partial\sigma^2}\right) = -E\left(\frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_1^n x'^2 - \frac{2}{\sigma^3}\sum_1^n \xi x'\right)$$

$$= \frac{n}{\sigma^2}\left(\frac{3\sum_1^n x'^2}{n\sigma^2} + \frac{2\xi\bar{x}'}{\sigma} - 1\right)$$

$$-E\left(\frac{\partial^2 \log L}{\partial\sigma\,\partial\xi}\right) = -E\left(\frac{\sum_1^n x'}{\sigma^2}\right) = -\frac{n\bar{x}'}{\sigma^2}$$

$$-E\left(\frac{\partial^2 \log L}{\partial \xi^2}\right) = -E\left(-n + \frac{n}{I_0^2}\left(\frac{\partial I_0}{\partial \xi}\right)^2 - \frac{n}{I_0}\frac{\partial^2 I_0}{\partial \xi^2}\right)$$

$$= -E\left(-n + \frac{n I_{-1}^2}{I_0^2} - \frac{n I_{-2}}{I_0}\right) = n\left(1 + \frac{I_0 I_{-2} - I_{-1}^2}{I_0^2}\right).$$

Hence the variance-covariance matrix of ξ and σ is

$$\underset{\sim}{I}^{-1} = \begin{bmatrix} \frac{n}{\sigma^2}\left(\frac{3\sum_1^n x'^2}{n\sigma^2} + \frac{2\xi\bar{x}'}{\sigma} - 1\right) & -\frac{n\bar{x}'}{\sigma^2} \\ -\frac{n\bar{x}'}{\sigma^2} & n\left(1 + \frac{I_0 I_{-2} - I_{-1}^2}{I_0^2}\right) \end{bmatrix}^{-1}$$

$$= \frac{1}{\Delta}\begin{bmatrix} n\left(1 + \frac{I_0 I_{-2} - I_{-1}^2}{I_0^2}\right) & \frac{n\bar{x}'}{\sigma^2} \\ \frac{n\bar{x}'}{\sigma^2} & \frac{n}{\sigma^2}\left(\frac{3\sum_1^n x'^2}{n\sigma^2} + \frac{2\xi\bar{x}'}{\sigma} - 1\right) \end{bmatrix}$$

where Δ is the determinant of $\underset{\sim}{I}^{-1}$.

COHEN, A.C., has discussed in his paper, Ann. Math. Stat. Vol. 21, 1950 pp.(557-569), the maximum likelihood estimates of the mean and variance of normal populations from singly and doubly truncated samples having known truncation points. In doubly truncated samples he discussed three cases: (i) when the number of the unmeasured observations is unknown; (ii) when the number of the unmeasured observations in each 'tail' is known; and (iii) when the total number of unmeasured observations known, but not the number in each 'tail'. Some numerical examples are given in this paper.

CHAPTER III

APPLICATIONS OF MAXIMUM LIKELIHOOD METHOD

1. SINGLE PARAMETER:

It is worth while to apply some other methods of estimation in example 3.1 and 3.2 to show that the maximum likelihood method is the best. The methods are:

(a) Maximum Likelihood method

(b) Minimum χ^2 method

(c) Product method

(d) Weighted mean method

(e) Additive method, also called Emerson's formula

Example 3.1: (Carver, Genetics, XII. (415-440) 1927), showing linkage between the sugary factor in maize and a factor for white base leaf. The case was oneof repulsion, and the numbers of seedlings counted were as in the following table

	Starchy		Sugary		Total
	Green	White	Green	White	
Observed	1997	906	904	32	3839
Expected	$\frac{n}{4}(2+P)$	$\frac{n}{4}(1-P)$	$\frac{n}{4}(1-P)$	$\frac{n}{4}P$	n

Here $P = p^2$ is the linkage value, and p is the recombination value. The parameter will be estimated is p.

(a) Maximum likelihood method:

Procedure: Let n_1, n_2, n_3, n_4 denote the observed values then the likelihood function is

$$L = \left\{\tfrac{1}{4}(2+P)\right\}^{n_1} \left\{\tfrac{1}{4}(1-P)\right\}^{n_2} \left\{\tfrac{1}{4}(1-P)\right\}^{n_3} \left\{\tfrac{1}{4}P\right\}^{n_4}$$

and

$$\log L = \text{constant} + n_1 \log(2+P) + n_2 \log(1-P) + n_3 \log(1-P) + n_4 \log$$

The estimate of the parameter P will be the solution of

$$\frac{\partial \log L}{\partial P} = 0$$

Here we have

$$\frac{\partial \log L}{\partial P} = \frac{n_1}{2+P} - \frac{n_2+n_3}{1-P} + \frac{n_4}{P}$$

Hence P will be the solution of

$$\frac{n_1}{2+P} - \frac{n_2+n_3}{1-P} + \frac{n_4}{P} = 0$$

By substituting the observed values we get

$$\frac{1997}{2+P} - \frac{1810}{1-P} + \frac{32}{P} = 0$$

Solving the equation we get

$$P = 0.035712$$

Hence

$$p = \sqrt{P} = 0.18898$$

We have from 1.4(a) chapter I that the variance of P will be given by

$$V(p) = \frac{1}{nE\left(\frac{\partial \log f}{\partial p}\right)^2},$$

then

$$V(\hat{P}) = 1 \Big/ \frac{n}{4}\left(\frac{1}{2+P} + \frac{2}{1-P} + \frac{1}{P}\right)$$

$$= \frac{2P(1-P)\ (2+P)}{n(1+2P)}$$

$$= \frac{2 \times 0.035712 \times 0.964288 \times 2.035712}{3839 \times 1.071424} = 0.34005 \times 10^{-4}$$

From Appendix I we have

$$V\hat{p} = \frac{VP}{4P},$$

then the standard error of $\hat{p}$ is

$$\sqrt{V\hat{p}} = \sqrt{\frac{VP}{4P}} = 0.01542 \ .$$

(b) Minimum χ^2 Method:

Procedure: The method of minimum χ^2 is expressed in the equation

$$\chi^2 = \frac{4}{n}\left(\frac{n_1^2}{2+P} + \frac{n_2^2}{1-P} + \frac{n_3^2}{1-P} + \frac{n_4^2}{P}\right) - n$$

The best estimate of P should make χ^2 a minimum and this will lead us to the equation of the 4th degree such that

$$\frac{\partial \chi^2}{\partial P} = \frac{4}{n}\left(-\frac{n_1^2}{(2+P)^2} + \frac{n_2^2+n_3^2}{(1-P)^2} - \frac{n_4^2}{P^2}\right) = 0 \ .$$

By substituting the observed values and solving the equation for P we get

$$P = 0.035785$$

Hence

$$\hat{p} = \sqrt{P} = 0.18917$$

The variance of P will be given by the same formula of method

(a) above; ie.

$$V_P = \frac{2P(1-P)\ (2+P)}{n(1+2P)}$$

$$= 0.3415 \times 10^{-4}$$

Then the standard error of the recombination p is

$$\sqrt{V_p} = \sqrt{\frac{V_P}{4P}} = 0.01547 .$$

(c) <u>Product Method:</u>

<u>Procedure:</u> The method of product is defined by the equation

$$\frac{n_1 n_4}{n_2 n_3} = \frac{P(2+P)}{(1-P)^2}$$

By substituting the observed values and solving the equation for P we get

$$P = 0.035645$$

and so

$$p = 0.1888$$

The variance of P is given by

$$VP = \frac{2P(1-P)\ (2+P)}{n(1+2P)} \qquad \text{(Appendix I)}$$

$$= \frac{2 \times 0.035645 \times 0.964355 \times 2.035646}{3839 \times 1.071292}$$

$$= 0.00003411$$

Hence the standard error of p is

$$\sqrt{V_p} = \sqrt{\frac{VP}{4P}} = 0.01545 .$$

(d) Weighted mean method:

Procedure: This method is defined by the equation

$$n(4P-1) = n_1 - 3n_2 - 3n_3 + 9n_4$$

ie.

$$4nP = 2n_1 - 2n_2 - 2n_3 + 10n_4$$

By substituting the observed values and solving the equation we get

$$P = 0.045194$$

Hence

$$p = 0.2126$$

The variance of P is given by

$$V_P = \frac{1 + 6P - 4P^2}{4n} \qquad \text{(Appendix I)}$$

$$= \frac{1 + 6 \times 0.045194 - (0.045194)^2}{4 \times 3839}$$

$$= 0.00006413$$

and so the standard error of p is

$$\sqrt{V_p} \doteq \sqrt{\frac{V_P}{4P}} = 0.02133 .$$

(e) Additive method

Procedure: This method is defined by equating $n_1 + n_4$ to its expected value $\frac{n}{4}(2+P) + \frac{n}{4}P = \frac{n}{2}(1+P)$, and so we get the equation

$$nP = n_1 - n_2 - n_3 + n_4$$

By substituting the observed values and solving the equation we get

$$P = 0.057046$$

andso

$$\hat{p} = 0.2388$$

The variance of P is given by

$$VP = \frac{1-P^2}{n} = \frac{1-(0.057046)^2}{3839} \qquad \text{(Appendix I)}$$

$$= 0.000259$$

and so the standard error of $\hat{p}$ is

$$\sqrt{V\hat{p}} = \sqrt{\frac{VP}{4P}} = 0.03373 \ .$$

Now we summarise the results of the five methods by the following table

Method	Recombination $\hat{p}$	Standard error of $\hat{p}$
Maximum likelihood	0.18898	0.01542
Minimum χ^2	0.18917	0.01547
Product formula	0.1888	0.01545
Weighted mean	0.2126	0.02133
Additive method	0.2388	0.03373

The table above shows that the standard error of the maximum likelihood estimate is the smallest, and since the standard error is the square root of the variance, therefore the variance of maximum likelihood estimate is the smallest. That is, the maximum likelihood method is the most efficient.

Example 3.2: De Winton and Haldane have recorded the results of self-pollinating and intercrossing Primula sinensis plants that were heterzygous for the two genes F, f and Ch, ch. These genes are linked and the 4164 individuals observed in the progeny of coupled double heterozygotes showed the following segregation in the table below:

	FCh	Fch	fCh	fch	Total
OBSERVED	2972	171	190	831	4164
EXPECTED	$\frac{n}{4}(2+P)$	$\frac{n}{4}(1-P)$	$\frac{n}{4}(1-P)$	$\frac{n}{4}P$	n

Here $P = (1-p)^2$ is the linkage value and p is the recombination value. We have to estimate the value of the parameter p.

(a) Maximum likelihood method:

Procedure: Let n_1, n_2, n_3, n_4 denote the observed values then the likelihood function is

$$L = \left\{\tfrac{1}{4}(2+P)\right\}^{n_1} \left\{\tfrac{1}{4}(1-P)\right\}^{n_2} \left\{\tfrac{1}{4}(1-P)\right\}^{n_3} \left\{\tfrac{1}{4}P\right\}^{n_4}$$

Then

$$\log L = \text{constand} + n_1 \log(2+P) + (n_2 + n_3)\log(1-P) + n_4 \log P$$

and

$$\frac{\partial \log L}{\partial P} = \frac{n_1}{2+P} - \frac{n_2+n_3}{1-P} + \frac{n_4}{P}$$

Since the estimate P is the solution of the equation

$$\frac{\partial \log L}{\partial P} = 0$$

Then P is the solution of

$$\frac{n_1}{2+P} - \frac{n_2 + n_3}{1-P} + \frac{n_4}{P} = 0$$

By substituting the observed values and solving the equation we get

$$P = 0.824734$$

Hence

$$p = 1-\sqrt{P} = 0.091851$$

The formulas for the variances which are used in the previous example will be used in this example too; therefore, the variance of P will be given by

$$V_P = \frac{2P(1-P)\ (2+P)}{n(1+2P)}$$

$$= 0.7402 \times 10^{-4}$$

Hence

$$\sqrt{V_p} = \sqrt{\frac{V_P}{4P}} = 0.004737$$

(b) <u>Minimum χ^2 Method:</u>

<u>Procedure:</u> The method of minimum χ^2 is defined by making χ^2 minimum in the equation

$$\chi^2 = \frac{4}{n}\left(\frac{n_1^2}{2+P} + \frac{n_2^2 + n_3^2}{1-P} + \frac{n_4^2}{P}\right) - n$$

That is the estimate P will be the solution of

$$\frac{\partial \chi^2}{\partial P} = 0$$

This will lead us to the equation

$$\frac{n_2^2 + n_3^2}{(1-P)} - \frac{n_1^2}{(2+P)} - \frac{n_4^2}{P} = 0$$

Substituting theobserved values and solving the equation we get

$$P = 0.8246$$

The variance of P is given by

$$V_P = \frac{2P(1-P)\ (2+P)}{n(1+2P)}$$

$$= 0.00007407$$

Hence

$$p = 1-\sqrt{P} = 0.09193$$

$$\sqrt{V_p} = \sqrt{\frac{V_P}{4P}} = 0.004739$$

(c) Product Method:

Procedure: This method is defined by the formula

$$\frac{n_1 n_4}{n_2 n_3} = \frac{P(2+P)}{(1-P)^2}$$

Substituting the observed values and solving the equation, we obtain

$$P = 0.8252$$

Hence

$$p = 1-\sqrt{P} = 0.0916$$

The variance of P is given by

$$V_P = \frac{2P(1-P)\ (2+P)}{n(1+2P)}$$

$$= 0.00007427$$

Hence

$$\sqrt{V_p} = \sqrt{\frac{V_P}{4P}} = 0.004743$$

(d) <u>Weighted mean method:</u>

<u>Procedure:</u> This method is defined by the equation

$$n(4P-1) = n_1 - 3n_2 - 3n_3 + 9n_4$$

ie.

$$4nP = 2n_1 - 2n_2 - 2n_3 + 10n_4$$

By substituting theobserved values and solving the equation, we get

$$P = 0.812439$$

and so

$$p = 1 - \sqrt{P} = 0.098652$$

The variance of P is given by

$$V_P = \frac{1+6P-4P^2}{4n}$$

$$= 0.000194189$$

Hence

$$\sqrt{V_p} = \sqrt{\frac{V_P}{4P}} = 0.00773$$

(e) Additive Method:

Procedure: This method is defined by equating $n_1 + n_4$ to its expected value $\frac{n}{4}(2+P) + \frac{n}{4}P = \frac{n}{2}(1+P)$, and so we get the equation

$$nP = n_1 - n_2 - n_3 + n_4$$

By substituting the observed values and solving the equation, we get

$$P = 0.8266$$

and so

$$p = 1 - \sqrt{P} = 0.09083$$

The variance of P is given by

$$V_P = \frac{1-P^2}{n}$$

$$= 0.000076$$

Hence

$$\sqrt{V_p} = \sqrt{\frac{V_P}{4P}} = 0.004798$$

The results of the methods are summarised in the following table

METHOD	Recombination p	Standard error of p
Maximum likelihood	0.091851	0.004737
Minimum χ^2	0.09193	0.004739
Product formula	0.0916	0.004743
Weighted mean	0.098652	0.00773
Additive method	0.09083	0.004798

We see in the column 3 of the table that thestandard error of the maximum likelihood estimate is thesmallest one, so the maximum likelihood method is the most efficient.

<u>Example 3.3</u>: The data of this example is given in the following table:

Frequencies observed in an F_2 segregation for aleurone colour and pale green seedling (BRUNSON'S data).

	CR	Cr+cR+cr	SEEDLING TOTAL
Pg_1	1907	1053	2960
pg_1	300	686	986
Aleurone total	2207	1739	$n = 3946$

In the case involving complementary factors, the probabilities in the four classes will be as in the following table:

	$CRPg_1$	$CRpg_1$	$(Cr+cR+cr)Pg_1$	$(Cr+cR+cr)pg_1$	TOTAL
OBSERVED	1907	300	1053	686	3946
EXPECTED	$\frac{3n}{16}(2+P)$	$\frac{3n}{16}(1-P)$	$\frac{3n}{16}(2-P)$	$\frac{n}{16}(1+3P)$	n

Here $P = p^2$ is the linkage value and p is the recombination value. In this example we will apply one method to estimate p in addition to the maximum likelihood method. This method is called Brunson's formula.

(a) <u>Maximum likelihood method</u>

<u>Procedure</u>: Let n_1, n_2, n_3, n_4 denote the observed values

then the likelihood function is

$$L=\left\{\frac{3}{16}(2+P)\right\}^{n_1}\left\{\frac{3}{16}(1-P)\right\}^{n_2}\left\{\frac{3}{16}(2-P)\right\}^{n_3}\left\{\frac{1}{16}(1+3P)\right\}^{n_4}$$

and

$$\log L = K + n_1\log(2+P) + n_2\log(1-P) + n_3\log(2-P) + n_4\log(1+3P)$$

where K is a constant. The estimate P is the solution of the equation

$$\frac{\partial \log L}{\partial P} = 0$$

ie. the solution of

$$\frac{n_1}{2+P} - \frac{n_2}{1-P} - \frac{n_3}{2-P} - \frac{n_4}{1+3P} = 0$$

By substituting of the observed values and solving the equation, we get

$$P = 0.5902$$

Hence

$$p = \sqrt{P} = 0.7682$$

in equal crossing in male and female, or

$1 - p = 1 - 0.7682 = 0.2318$ crossing over, with coupling.

The variance of P is given by

$$V_P = 1\Big/-E\left(\frac{\partial^2 \log L}{\partial P^2}\right) = \frac{4}{3n}\ \frac{(2+P)(1-P)(2-P)(1+3P)}{5+2P-4P^2}$$

and

$$V_p = \frac{V_P}{4P} = \frac{V_P}{(2p)^2} = \frac{(2+p^2)\ (1-p^2)\ (2-p^2)\ (1+3p^2)}{3n\, p^2(5+2p^2-4p^4)}$$

Substituting for p, we get

$$V_p = 0.000124$$

Hence

$$\sqrt{V_p} = 0.011 \text{ is the standard error of } p$$

(b) BRUNSON'S METHOD

Procedure: The method of Brunson is defined by the formula

$$p^2 = \frac{16}{18n}(n_1 - n_2 - n_3 + 3n_4)$$

where n_1, n_2, n_3 and n_4 are as defined in (a).

Now if we substitute the observed values and solve the equation we get

$$p = 0.767$$

Let T be any function of the frequencies, then the variance of T will be given by the general formula (B) in Appendix (I); ie.

$$\frac{1}{n}V(T) = \sum_i \left\{ \theta_i \left(\frac{dT}{dn_i}\right)^2 \right\} - \left(\frac{dT}{dn}\right)^2$$

where θ_i is the probability corresponding to the ith class.

Here let $T = p^2$ then

$$T = \frac{16}{18n}(n_1 - n_2 - n_3 + 3n_4).$$

Then

$$\sum_1^4 \left\{ \theta_i \left(\frac{dT}{dn_i}\right)^2 \right\} = \frac{64}{81n^2}\left\{ \frac{3}{16}(2+P) + \frac{3}{16}(1-P) + \frac{3}{16}(2-P) + \frac{9}{16}(1+3P) \right\}$$

$$= \frac{32(1+P)}{27n^2}.$$

And

$$\left(\frac{dT}{dn}\right)^2 = \left(-\frac{16(n_1 - n_2 - n_3 + 3n_4)}{18n^2}\right)^2$$

$$= \frac{P^2}{n^2}$$

Then

$$\frac{1}{n} V(T) = \frac{32(1+P)}{27n^2} - \frac{P^2}{n^2}$$

ie. $$V_P = \frac{32+32P - 27P^2}{27n}$$

and since we have

$$V_p = \frac{VP}{4P}$$

then

$$V_p = \frac{32+32P - 27P^2}{108nP}$$

$$= \frac{32+32p^2 - 27p^4}{108n\,p^2}$$

Substituting thevalue of p we obtain

$$V_p = 0.000165$$

The following table shows the comparison of expected with observed frequencies

		CRP_{g_1}	CRp_{g_1}	$(Cr+cR+cr)P_{g_1}$	$(Cr+cR+cr)p_{g_1}$	n
OBSERVED		1907	300	1053	686	3946
EXPECTED	M.L.	1916.42	303.20	1043.08	683.30	3946
	BR.	1915	305	1044	682	3946

We can calculate

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{(\text{Expected})}$$

to show how far the observed values are associated with the expected values. We have

χ^2 for maximum likelihood method $= 0.185$,

χ^2 for Brunson's method $= 0.2165$

In each case the degrees of freedom are 2. From the statistical table we have $\chi^2_{0.05} = 5.99$ for 2 degrees of freedom.. We see in the two methods that the observed frequencies are associated with the expected frequencies but the maximum likelihood method seems better than Brunson's method.

The following table shows the summarised results of the two methods

METHOD	Recombination p	Variance of p
Maximum likelihood	0.7682	0.000124
Brunson's formula	0.767	0.000165

We see from the table that the variance of maximum likelihood estimate is smaller than the variance of the estimate of Brunson's method, therefore the maximum likelihood method is more efficient than Brunson's method. We can also calculate the efficiency of Brunson's method with respect to the maximum likelihood method

$$E = \frac{V_{p\,M.L.}}{V_{p\,BR.}} = \frac{0.000124}{0.000165} = 75\%$$

ie. the efficiency of Brunson's method is 75 per cent.

2. SEVERAL PARAMETERS

Example 3.4: The data in the following table showing the effect of a series of concentrations of rotenone when sprayed on Macrosiphoniella sanborni, the chrysanthemum aphis, in batches of about fifty.

Toxicity of Rotenone to Macrosiphoniella sanborni

Concentration (mg/l)	No. of insects (n)	No. affected (r)	% kill p	Log Concentration (x)	Empirical probit
10.2	50	44	88	1.01	6.18
7.7	49	42	86	0.89	6.08
5.1	46	24	52	0.71	5.05
3.8	48	16	33	0.58	4.56
2.6	50	6	12	0.41	3.82

The last column is obtained from Table I ("Transformation of percentages to probits", - Finney, Probit analysis PP. 22.)

(a) Procedure: If P is the expected proportion of animals killed by the dosage x_0, then P will be in the form

$$P = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx .$$

The estimation of the parameters μ and σ is based upon the probit transformation of the experimental results, ie. to converte the dose x into a probit (euqivalent normal deviate +5), then the probits will be related linearly with the dose x, (or log x). In virtue of the above assumption, P, will be in the following form

$$P = \int_{-\infty}^{Y} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \, dt$$

where

$$Y = 5 + \frac{1}{\sigma}(x - \mu)$$

It is found more convenient to put Y as

$$Y = \alpha + \beta x$$

and estimate the parameters α and β rather than μ and σ , where

$$\mu = \frac{5 - \alpha}{\beta} \quad \text{and} \quad \sigma = \frac{1}{\beta} .$$

Now the probability of r responding is

$$\binom{n}{r} P^r (1 - P)^{n-r}$$

then the likelihood function will be such that

$$L = \prod_i \left[\binom{n}{r} P^r (1-P)^{n-r} \right]$$

and

$$\log L = K + \sum \left[r \log P + (n - r) \log (1 - P) \right]$$

where K is constant. Differentiate with respect to α and β , we get

$$\frac{\partial \log L}{\partial \alpha} = \sum \left[\frac{r}{P} \frac{\partial P}{\partial \alpha} - \frac{n-r}{1-P} \frac{\partial P}{\partial \alpha} \right]$$

$$= \sum \left[\left(\frac{r - nP}{PQ} \right) \frac{\partial P}{\partial \alpha} \right] = \sum \left[\left(\frac{r - nP}{PQ} \right) \frac{\partial P}{\partial Y} \right]$$

where $Q = 1 - P$ and $\frac{\partial P}{\partial \alpha} = \frac{\partial P}{\partial Y}\frac{\partial Y}{\partial \alpha} = \frac{\partial P}{\partial Y}$.

By the same way we obtain

$$\frac{\partial \log L}{\partial \beta} = \sum\left[\left(\frac{r-nP}{PQ}\right)\frac{\partial P}{\partial Y}\,x\right] .$$

Let $\frac{\partial P}{\partial Y}\left(= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Y^2}\right) = Z$, then the maximum likelihood equations will be such that

$$\sum\left[\left(\frac{r-nP}{PQ}\right)Z\right] = 0$$

$$\sum\left[\left(\frac{r-nP}{PQ}\right)Z\,x\right] = 0$$

We can get the values of P and Z correspondings to the values of α , β and x . The variance-covariance matrix of the parameters α and β is given by

$$\underset{\sim}{I}^{-1} = \begin{bmatrix} -E\left(\frac{\partial^2 \log L}{\partial \alpha^2}\right) & -E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) \\ -E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) & -E\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) \end{bmatrix}^{-1}$$

and since we can show that

$$-E\left(\frac{\partial^2 \log L}{\partial \alpha^2}\right) = \sum\left(\frac{nZ^2}{PQ}\right), \qquad -E\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) = \sum\left(\frac{nZ^2x^2}{PQ}\right)$$

and $$-E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) = \sum\left(\frac{nZ^2x}{PQ}\right)$$

then the variance-covariance matrix of the parameters α and β will be such that

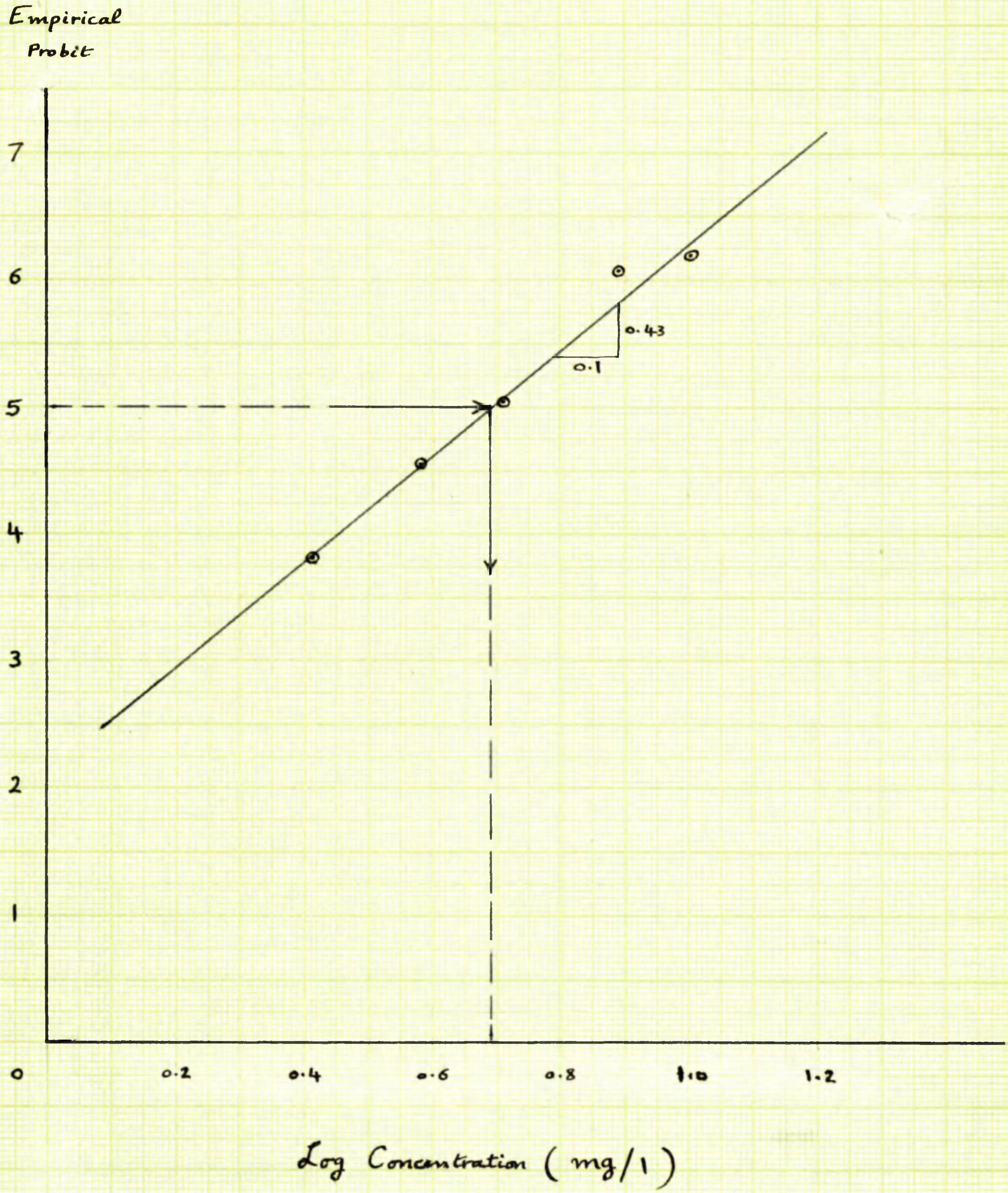
$$\underset{\sim}{I}^{-1} = \begin{bmatrix} \sum\left(\frac{nZ^2}{PQ}\right) & \sum\left(\frac{nZ^2x}{PQ}\right) \\ \sum\left(\frac{nZ^2x}{PQ}\right) & \sum\left(\frac{nZ^2x^2}{PQ}\right) \end{bmatrix}^{-1}$$

(b) <u>The Initial Estimates</u>: Usually the maximum likelihood equations are difficult to solve, therefore we have to get an initial estimate of the maximum likelihood estimates and by successive approximations (section 3, chapter II), we obtain the estimates of maximum likelihood equations. The procedure for getting the initial estimates in this example is to plot the empirical probits in the last column of the table above against the corresponding dosages. Draw a straight line by eye through these points, then by this line we get the value of μ corresponding to the value of $Y = 5$, ie. the value of the dose which kills 50% of the group. Also we get the value of $\frac{1}{\sigma} = \frac{\partial Y}{\partial x}$, which is the rate of increase of the probit value per unit increase in x. After getting the values of μ and σ we calculate the values of α and β from the relations

$$\mu = \frac{5-\alpha}{\beta} \quad and \quad \beta = \frac{1}{\sigma}$$

then by substituting the values of α, β and x in the linear relation

$$Y = \alpha + \beta x$$



Empirical Probit
7
6
5
4
3
2
1
0
0.43
0.1
0.2
0.4
0.6
0.8
1.0
1.2
Log Concentration (mg/1)

we get the value of Y. Corresponding to the values of Yi we find from the tables the values of Pi and Zi and then we calculate the successive approximations.

(c) <u>The Calculations:</u> From the figure we find

$$\beta = \frac{1}{\sigma} = \frac{0.43}{0.1} = 4.3$$

and

$$\mu = 0.69 = \frac{5-\alpha}{\beta} = \frac{5-\alpha}{4.3}$$

$\therefore \quad \alpha = 2.03$

The first approximation is given by

$$\begin{bmatrix} \alpha^{(1)} \\ \beta^{(1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(0)} \\ \beta^{(0)} \end{bmatrix} + \underset{\sim}{I}^{-1}_{\alpha^{(0)},\beta^{(0)}} \begin{bmatrix} \frac{\partial \log L}{\partial \alpha} \\ \frac{\partial \log L}{\partial \beta} \end{bmatrix}_{\alpha^{(0)},\beta^{(0)}}$$

We have $Y_i = \alpha^{(0)} + \beta^{(0)} x_i$, then

$Y_1 = 2.03 + 4.3 \times 1.01 = 6.4$	$P_1 = 0.92, \quad Z_1 = 0.15$
$Y_2 = 2.03 + 4.3 \times 0.89 = 5.9$	$P_2 = 0.82, \quad Z_2 = 0.27$
$Y_3 = 2.03 + 4.3 \times 0.71 = 5.1$	$P_3 = 0.54, \quad Z_3 = 0.40$
$Y_4 = 2.03 + 4.3 \times 0.58 = 4.5$	$P_4 = 0.31, \quad Z_4 = 0.35$
$Y_5 = 2.03 + 4.3 \times 0.41 = 3.8$	$P_5 = 0.12, \quad Z_5 = 0.19$

Then

$$\sum_1^5 \left(\frac{n Z^2}{PQ}\right) = 50 \times 0.3 + 49 \times 0.47 + 46 \times 0.63 + 48 \times 0.58 + 50\times($$

$$= 15 + 23.03 + 28.98 + 27.84 + 16.50$$

$$= 111.35$$

$$\sum_{1}^{5}\left(\frac{nZ^2x}{PQ}\right) = 15.15 + 20.50 + 20.58 + 16.15 + 6.77$$
$$= 79.15$$

$$\sum_{1}^{5}\left(\frac{nZ^2x^2}{PQ}\right) = 15.30 + 18.25 + 14.61 + 9.37 + 2.77$$
$$= 60.30$$

Then the variance-covariance matrix of $\alpha^{(0)}$ and $\beta^{(0)}$ is

$$\underset{\sim}{I}^{-1} = \begin{bmatrix} 111.35 & 79.15 \\ 79.15 & 60.30 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.134 & -0.176 \\ -0.176 & 0.248 \end{bmatrix}$$

Also

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(0)}} = \left(\frac{44-50\times0.92}{0.92\times0.08}\right)0.15 + \left(\frac{42-49\times0.82}{0.82\times0.18}\right)0.27 + \left(\frac{24-46\times0.54}{0.54\times0.46}\right)0.4$$
$$+ \left(\frac{16-48\times0.31}{0.31\times0.69}\right)0.35 + \left(\frac{6-50\times0.12}{0.12\times0.88}\right)0.19$$
$$= -0.29$$

and

$$\left(\frac{\partial \log L}{\partial \beta}\right)_{\beta^{(0)}} = -1.0744$$

Then

$$\begin{bmatrix} \alpha^{(1)} \\ \beta^{(1)} \end{bmatrix} = \begin{bmatrix} 2.03 \\ 4.3 \end{bmatrix} + \begin{bmatrix} 0.134 & -0.176 \\ -0.176 & 0.248 \end{bmatrix} \begin{bmatrix} -0.29 \\ -1.07 \end{bmatrix}$$

$$= \begin{bmatrix} 2.03 \\ 4.3 \end{bmatrix} + \begin{bmatrix} 0.15 \\ -0.25 \end{bmatrix} = \begin{bmatrix} 2.18 \\ 4.05 \end{bmatrix}$$

We repeat the process again because the corrections are not small

$Y_1 = 2.18 + 4.05 \times 1.01 = 6.25$ $\quad P_1 = 0.89, \quad Z_1 = 0.18$

$Y_2 = 2.18 + 4.05 \times 0.89 = 5.78$ $\quad P_2 = 0.78, \quad Z_2 = 0.29$

$Y_3 = 2.18 + 4.05 \times 0.71 = 5.06$ $\quad P_3 = 0.52, \quad Z_3 = 0.40$

$Y_4 = 2.18 + 4.05 \times 0.58 = 4.53$ $\quad P_4 = 0.32, \quad Z_4 = 0.36$

$Y_5 = 2.18 + 4.05 \times 0.41 = 3.84$ $\quad P_5 = 0.12, \quad Z_5 = 0.20$

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(1)}} = -0.81 + 6.4 + 0.01 + 1.06 = 6.66$$

$$\left(\frac{\partial \log L}{\partial \beta}\right)_{\beta^{(1)}} = -0.82 + 5.7 + 0.01 + 0.61 = 5.5$$

Then

$$\begin{bmatrix} \alpha^{(2)} \\ \beta^{(2)} \end{bmatrix} = \begin{bmatrix} 2.18 \\ 4.05 \end{bmatrix} + \begin{bmatrix} 0.134 & -0.176 \\ -0.176 & 0.248 \end{bmatrix} \begin{bmatrix} 6.66 \\ 5.5 \end{bmatrix}$$

$$= \begin{bmatrix} 2.18 \\ 4.05 \end{bmatrix} + \begin{bmatrix} -0.08 \\ 0.19 \end{bmatrix} = \begin{bmatrix} 2.1 \\ 4.24 \end{bmatrix}$$

We repeat the process again because the corrections are not small

$Y_1 = 2.1 + 4.24 \times 1.0. = 6.38$ $\qquad P_1 = 0.92, \quad Z_1 = 0.15$

$Y_2 = 2.1 + 4.24 \times 0.89 = 5.87$ $\qquad P_2 = 0.81, \quad Z_2 = 0.27$

$Y_3 = 2.1 + 4.24 \times 0.71 = 5.11$ $\qquad P_3 = 0.54, \quad Z_3 = 0.40$

$Y_4 = 2.1 + 4.24 \times 0.58 = 4.56$ $\qquad P_4 = 0.33, \quad Z_4 = 0.36$

$Y_5 = 2.1 + 4.24 \times 0.41 = 3.84$ $\qquad P_5 = 0.12, \quad Z_5 = 0.20$

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(2)}} = -4.1 + 4.06 - 1.35 + 0.26 = -1.13$$

$$\left(\frac{\partial \log L}{\partial \beta}\right)_{\beta^{(2)}} = -4.14 + 3.61 - 0.96 + 0.15 = -1.34$$

Then

$$\begin{bmatrix} \alpha^{(3)} \\ \beta^{(3)} \end{bmatrix} = \begin{bmatrix} 2.1 \\ 4.24 \end{bmatrix} + \begin{bmatrix} 0.134 & -0.176 \\ -0.176 & 0.248 \end{bmatrix} \begin{bmatrix} -1.13 \\ -1.34 \end{bmatrix}$$

$$= \begin{bmatrix} 2.1 \\ 4.24 \end{bmatrix} + \begin{bmatrix} 0.03 \\ -0.13 \end{bmatrix} = \begin{bmatrix} 2.13 \\ 4.11 \end{bmatrix}$$

We repeat the process of approximation again because one of the two corrections is still not small.

$Y_1 = 2.13 + 4.11 \times 1.01 = 6.28$ $\qquad P_1 = 0.90, \quad Z_1 = 0.18$

$Y_2 = 2.13 + 4.11 \times 0.89 = 5.80$ $\qquad P_2 = 0.79, \quad Z_2 = 0.29$

$Y_3 = 2.13 + 4.11 \times 0.71 = 5.05$ $\qquad P_3 = 0.52, \quad Z_3 = 0.40$

$Y_4 = 2.13 + 4.11 \times 0.58 = 4.51$ $\qquad P_4 = 0.31, \quad Z_4 = 0.35$

$Y_5 = 2.13 + 4.11 \times 0.41 = 3.82$ $\qquad P_5 = 0.12, \quad Z_5 = 0.20$

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(3)}} = -2 + 5.75 + 0.01 + 1.83 = 5.6$$

$$\left(\frac{\partial \log L}{\partial \beta}\right)_{\beta^{(3)}} = -2.02 + 5.12 + 0.01 + 1.06 = 4.2$$

Then

$$\begin{bmatrix} \alpha^{(4)} \\ \beta^{(4)} \end{bmatrix} = \begin{bmatrix} 2.13 \\ 4.11 \end{bmatrix} + \begin{bmatrix} 0.134 & -0.176 \\ -0.176 & 0.248 \end{bmatrix} \begin{bmatrix} 5.6 \\ 4.2 \end{bmatrix}$$

$$= \begin{bmatrix} 2.13 \\ 4.11 \end{bmatrix} + \begin{bmatrix} 0.01 \\ 0.06 \end{bmatrix} = \begin{bmatrix} 2.14 \\ 4.17 \end{bmatrix}$$

We see here that the corrections are sufficiently small, therefore the estimates of the maximum likelihood equations are

$$\overset{\times}{\alpha} = 2.14 \quad \text{and} \quad \overset{\times}{\beta} = 4.17$$

Now the values of Yi's and $\left(\frac{Z^2}{PQ}\right)_i$'s corresponding to $\overset{\times}{\alpha} = 2.14$, $\overset{\times}{\beta} = 4.17$ and x_i's are

$Y_1 = 2.14 + 4.17 \times 1.01 = 6.35 \qquad (Z^2/PQ)_1 = 0.32, \; P_1 = 0.91$

$Y_2 = 2.14 + 4.17 \times 0.89 = 5.85 \qquad (Z^2/PQ)_2 = 0.48, \; P_2 = 0.80$

$Y_3 = 2.14 + 4.17 \times 0.71 = 5.10 \qquad (Z^2/PQ)_3 = 0.63, \; P_3 = 0.54$

$Y_4 = 2.14 + 4.17 \times 0.58 = 4.56 \qquad (Z^2/PQ)_4 = 0.59, \; P_4 = 0.33$

$Y_5 = 2.14 + 4.17 \times 0.41 = 3.85 \qquad (Z^2/PQ)_5 = 0.38, \; P_5 = 0.125$

$$\sum_1^5 \left(\frac{nZ^2}{PQ}\right) = 50 \times 0.32 + 49 \times 0.48 + 46 \times 0.63 + 48 \times 0.59 + 50 \times 0.38$$

$$= 16.00 + 23.52 + 28.98 + 28.32 + 19.00 = 115.82$$

$$\sum_{1}^{5}\left(\frac{nZ^2x}{PQ}\right) = 16.16 + 20.93 + 20.58 + 16.43 + 7.79 = 81.89$$

$$\sum_{1}^{5}\left(\frac{nZ^2x^2}{PQ}\right) = 16.32 + 18.63 + 14.61 + 9.53 + 3.19 = 62.28$$

Then the variance-covariance matrix of $\overset{x}{\alpha} = 2.14$ and $\overset{x}{\beta} = 4.17$ is

$$\begin{bmatrix} 115.82 & 81.89 \\ 81.89 & 62.28 \end{bmatrix}^{-1} = \frac{1}{507.3}\begin{bmatrix} 62.28 & -81.89 \\ -81.89 & 115.82 \end{bmatrix}$$

ie.

$$\begin{bmatrix} 115.82 & 81.89 \\ 81.89 & 62.28 \end{bmatrix}^{-1} = \begin{bmatrix} 0.123 & -0.161 \\ -0.161 & 0.228 \end{bmatrix}$$

The linear relation between the probit and the log dose is then

$$Y = 2.14 + 4.17x$$

The estimate of the log dose which kills 50% of the group is

$$\overset{x}{\mu} = \frac{5-\overset{x}{\alpha}}{\overset{x}{\beta}} = \frac{5-2.14}{4.17} = 0.686$$

The variance of $\overset{x}{\mu}$ is given by

$$V\overset{x}{\mu} = \frac{1}{\overset{x}{\beta}}\left[\frac{1}{\sum nw} + \frac{(\overset{x}{\mu}-\bar{x})^2}{\sum nw(x-\bar{x})^2}\right]$$

$$\bar{x} = \frac{\sum nx}{\sum n} = \frac{175.11}{243} = 0.721$$

$$(\overset{x}{\mu}-\bar{x})^2 = (0.686 - 0.721)^2 = 0.001225$$

$$\sum nw = \sum n\frac{Z^2}{PQ} = 115.82$$

$$\sum nw(x-\bar{x}) = 10.996$$

Then

$$V_{\hat{\mu}} = \frac{1}{(4.17)^2}\left[\frac{1}{115.82} + \frac{0.001225}{10.996}\right]$$

$$= 0.058083\ (0.0001114 + 0.008634)$$

$$= 0.000508,$$

and so

$$\hat{\mu} = 0.686 \pm 0.023$$

To test the association of theboserved frequencies with the expected frequencies we use χ^2- test.

$$\chi^2 = \sum_1^5 \frac{(np - nP)^2}{nP(1-P)} = \sum_1^5 \frac{n(p-P)^2}{PQ}$$

$$= \frac{50(0.88-0.91)^2}{0.91 \times 0.09} + \frac{49(0.86-0.80)^2}{0.80 \times 0.20} + \frac{46(0.52-0.54)^2}{0.54 \times 0.46} + \frac{48(0.33-0.33)^2}{0.33 \times 0.67}$$

$$+ \frac{50(0.12-0.125)^2}{0.125 \times 0.875}$$

$$= 1.737$$

The degrees of freedom are 3, and $\chi^2_{0.05} = 7.81$ for 3 degrees of freedom from the statistical table. This shows that the observed frequencies are associated sufficiently with the expected frequencies.

<u>Example 3.5</u>: (Data of Example 3.4)

In example 3.4 we used the probit transformation to estimate the parameters μ and σ . In this example we are using the logistic formula

$$P = \frac{1}{1 + e^{\alpha - \beta x}}$$

where P is as defined in example 3.4 and α and β are the parameters to be estimated. The parameter μ will be such that

$$\mu = \frac{\alpha}{\beta}$$

We can show that the maximum likelihood estimates of α and β will be given by the solution of the equations

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^{k} (n_i P_i - m_i) = 0$$

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{k} (m_i - n_i P_i) x_i = 0$$

where K is the number of the groups exposed to the experiment, n_i is the number of theindividuals within the group, m_i is the number which responded and Pi is expected proportion of the individuals killed by x_i, the log dose. Usually in practice the two equations above are difficult to solve, hence in such cases we have to find initial estimates and by successive approximations we obtain the maximum likelihood estimates. The procedure of getting the initial estimates is as follows. Plot $\log_e[(n_i - m_i)/m_i]$ against x_i, then draw by eye a straight line through these points and by this line we get the initial estimates. The following graph shows the initial estimates which are obtained.

Now we start to calculate the values of the points which designate the straight line. Here let

$$l_i = \log_e\left(\frac{n_i - m_i}{m_i}\right)$$

$$= \log_{10}\left(\frac{n_i - m_i}{m_i}\right) \log_e 10 \quad .$$

Then

$$l_1 = \log_e \frac{6}{44} = (0.4771 - 1.3424)\ 2.3 = -1.89$$

$$l_2 = \log_e \frac{7}{42} = (0.0000 - 0.7782)\ 2.3 = -1.79$$

$$l_3 = \log_e \frac{22}{24} = (1.3424 - 1.3802)\ 2.3 = -.0.087$$

$$l_4 = \log_e \frac{32}{16} = (0.3010 - 0.0000)\ 2.3 = 0.69$$

$$l_5 = \log_e \frac{44}{6} = (1.3424 - 0.4771)\ 2.3 = 1.89$$

When $x = 0$, we get from the graph that $l = 4.4 = \alpha$,

and when $l = 0$ we get from the graph also that $x = 0.7$.

Since $l = \alpha - \beta x$ then

$$\alpha - 0.7\beta = 0$$

ie.

$$\beta = \frac{\alpha}{0.7} = \frac{4.4}{0.7} = 6.3$$

Hence the initial estimates of α and β are $\alpha^{(0)} = 4.4$ and $\beta^{(0)} = 6.3$

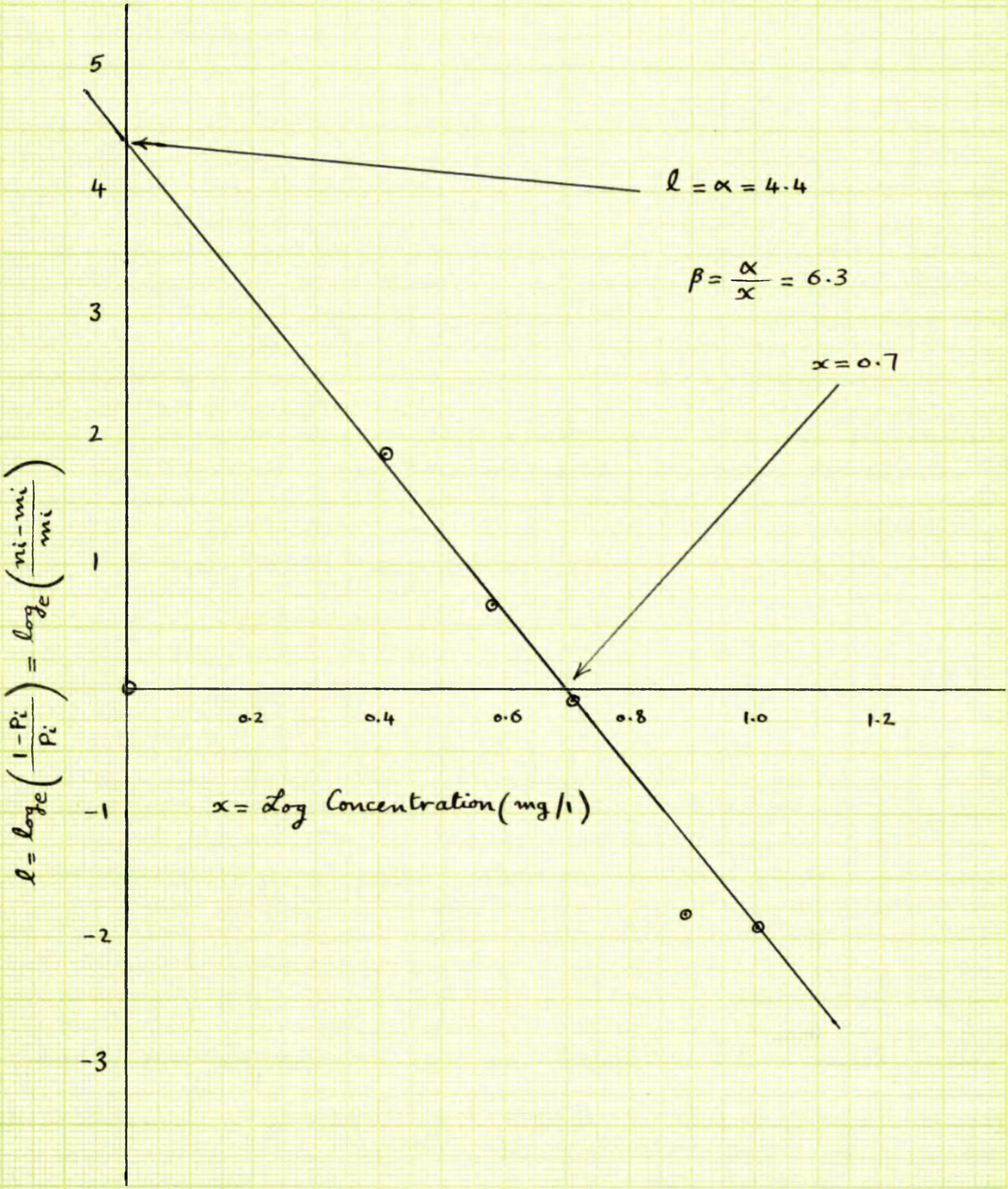
Now we calculate P_i's according to the values of x_i's , $\alpha^{(0)}$ and $\beta^{(0)}$ Here we have

$$P_i = \frac{1}{1 + e^{\alpha - \beta x_i}} \ ,$$

ie.

$$\log_{10}\left(\frac{1}{P_i} - 1\right) = \frac{\alpha - \beta x_i}{2.3}$$

Then



$\ell = \log_e\left(\frac{1-P_i}{P_i}\right) = \log_e\left(\frac{n_i-m_i}{m_i}\right)$
$\ell = \alpha = 4.4$
$\beta = \frac{\alpha}{x} = 6.3$
$x = 0.7$
x = Log Concentration (mg/l)
5
4
3
2
1
-1
-2
-3
0.2
0.4
0.6
0.8
1.0
1.2

$$\log_{10}\left(\frac{1}{P_1} - 1\right) = \frac{4.4 - 6.3 \times 1.01}{2.3} = \bar{1}.1465 ,$$

$$\frac{1}{P_1} - 1 = 0.14 \qquad\qquad P_1 = 0.88 ,$$

$$\log_{10}\left(\frac{1}{P_2} - 1\right) = \frac{4.4 - 6.3 \times 0.89}{2.3} = \bar{1}.4752,$$

$$\frac{1}{P_2} - 1 = 0.299 \qquad\qquad P_2 = 0.77,$$

$$\log_{10}\left(\frac{1}{P_3} - 1\right) = \frac{4.4 - 6.3 \times 0.71}{2.3} = \bar{1}.9683,$$

$$\frac{1}{P_3} - 1 = 0.93 \qquad\qquad P_3 = 0.52 ,$$

$$\log_{10}\left(\frac{1}{P_4} - 1\right) = \frac{4.4 - 6.3 \times 0.58}{2.3} = 0.2374,$$

$$\frac{1}{P_4} - 1 = 1.73 \qquad\qquad P_4 = 0.37 ,$$

$$\log_{10}\left(\frac{1}{P_5} - 1\right) = \frac{4.4 - 6.3 \times 0.41}{2.3} = 0.7883 ,$$

$$\frac{1}{P_5} - 1 = 6.14 \qquad\qquad P_5 = 0.14$$

Hence

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(0)}} = \sum_{1}^{5}\left(n_i P_i - m_i\right) = -1.59 ,$$

and

$$\left(\frac{\partial \log L}{\partial \beta}\right)_{\beta^{(0)}} = \sum_{1}^{5}\left(m_i - n_i P_i\right) x_i = 2.43 .$$

The information matrix is given by

$$\frac{1}{n}\underset{\sim}{I} = \begin{bmatrix} -\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial \alpha^2}\right) & -\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) \\ -\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) & -\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) \end{bmatrix}$$

Then the variance-covariance matrix of and is

$$\underset{\sim}{I}^{-1} = \begin{bmatrix} -E\left(\frac{\partial^2 \log L}{\partial \alpha^2}\right) & -E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) \\ -E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) & -E\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) \end{bmatrix}^{-1}_{\alpha^{(0)}, \beta^{(0)}}$$

Now

$$-E\left(\frac{\partial^2 \log L}{\partial \alpha^2}\right) = -\sum_1^5 n_i \frac{\partial P_i}{\partial \alpha} = \sum_1^5 n_i P_i (1-P_i)$$

$$-E\left(\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right) = -\sum_1^5 -n_i x_i \frac{\partial P_i}{\partial \alpha} = -\sum_1^5 n_i x_i P_i (1-P_i)$$

$$-E\left(\frac{\partial^2 \log L}{\partial \beta^2}\right) = -\sum_1^5 - n x_i \frac{\partial P_i}{\partial \beta} = \sum_1^5 n_i x_i^2 P_i (1-P_i)$$

then by substituting the values of x_i's, n_i's and P_i's we get

$$\sum_1^5 n_i P_i (1-P_i) = 42.6483$$

$$-\sum_1^5 n_i x_i P_i (1-P_i) = -30.1767$$

$$\sum_1^5 n_i x_i^2 P_i (1-P_i) = 22.7244$$

Hence

$$\underset{\sim}{I}^{-1} = \begin{bmatrix} 42.65 & -30.18 \\ -30.18 & 22.72 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.39 & 0.52 \\ 0.52 & 0.73 \end{bmatrix}$$

The first approximation is then

$$\begin{bmatrix} \alpha^{(1)} \\ \beta^{(1)} \end{bmatrix} = \begin{bmatrix} 4.4 \\ 6.3 \end{bmatrix} + \begin{bmatrix} 0.39 & 0.52 \\ 0.52 & 0.73 \end{bmatrix} \begin{bmatrix} -1.59 \\ 2.43 \end{bmatrix}$$

$$= \begin{bmatrix} 4.4 \\ 6.3 \end{bmatrix} + \begin{bmatrix} 0.64 \\ 0.95 \end{bmatrix} = \begin{bmatrix} 5.04 \\ 7.25 \end{bmatrix}$$

Repeat the process again for the second approximation

$$\log_{10}\left(\frac{1}{P_1} - 1\right) = \frac{5.04 - 7.25 \times 1.01}{2.3} = \bar{1}.0076$$

$$\frac{1}{P_1} - 1 = 0.102 \qquad P_1 = 0.91$$

$$\log_{10}\left(\frac{1}{P_2} - 1\right) = \frac{5.04 - 7.25 \times 0.89}{2.3} = \bar{1}.3859$$

$$\frac{1}{P_2} - 1 = 0.243 \qquad P_2 = 0.80$$

$$\log_{10}\left(\frac{1}{P_3} - 1\right) = \frac{5.04 - 7.25 \times 0.71}{2.3} = \bar{1}.9533$$

$$\frac{1}{P_3} - 1 = 0.898 \qquad P_3 = 0.52$$

$$\log_{10}\left(\frac{1}{P_4} - 1\right) = \frac{5.04 - 7.25 \times 0.58}{2.3} = 0.3630$$

$$\frac{1}{P_4} - 1 = 2.31 \qquad P_4 = 0.30$$

$$\log_{10}\left(\frac{1}{P_5} - 1\right) = \frac{5.04 - 7.25 \times 0.41}{2.3} = 0.8989$$

$$\frac{1}{P_5} - 1 = 7.92 \qquad P_5 = 0.11$$

Then

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(1)}} = -3.48$$

$$\left(\frac{\partial \log L}{\partial \beta}\right)_{\beta^{(1)}} = 2.17$$

Hence

$$\begin{bmatrix} \alpha^{(2)} \\ \beta^{(2)} \end{bmatrix} = \begin{bmatrix} 5.04 \\ 7.25 \end{bmatrix} + \begin{bmatrix} 0.39 & 0.52 \\ 0.52 & 0.73 \end{bmatrix} \begin{bmatrix} -3.48 \\ 2.17 \end{bmatrix}$$

$$= \begin{bmatrix} 5.04 \\ 7.25 \end{bmatrix} + \begin{bmatrix} -0.23 \\ -0.23 \end{bmatrix} = \begin{bmatrix} 4.81 \\ 7.02 \end{bmatrix}$$

We repeat the process again to get the third approximation

$$\log_{10}\left(\frac{1}{P_1} - 1\right) = \frac{4.81 - 7.02 \times 1.01}{2.3} = \bar{1}.0086$$

$$\frac{1}{P_1} - 1 = 0.102 \qquad P_1 = 0.91$$

$$\log_{10}\left(\frac{1}{P_2} - 1\right) = \frac{4.81 - 7.02 \times 0.89}{2.3} = \bar{1}.3749$$

$$\frac{1}{P_2} - 1 = 0.237 \qquad P_2 = 0.81$$

$$\log_{10}\left(\frac{1}{P_3} - 1\right) = \frac{4.81 - 7.02 \times 0.71}{2.3} = \bar{1}.9243$$

$$\frac{1}{P_3} - 1 = 0.84 \qquad P_3 = 0.54$$

$$\log_{10}\left(\frac{1}{P_4} - 1\right) = \frac{4.81 - 7.02 \times 0.58}{2.3} = 0.3167$$

$$\frac{1}{P_4} - 1 = 2.07 \qquad P_4 = 0.33$$

$$\log_{10}\left(\frac{1}{P_5} - 1\right) = \frac{4.81 - 7.02 \times 0.41}{2.3} = 0.8356$$

$$\frac{1}{P_5} - 1 = 6.85 \qquad P_5 = 0.13$$

Then

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(2)}} = 0.37$$

$$\left(\frac{\partial \log L}{\partial \beta}\right)_{\beta^{(2)}} = -0.16$$

Hence

$$\begin{bmatrix} \alpha^{(3)} \\ \beta^{(3)} \end{bmatrix} = \begin{bmatrix} 4.81 \\ 7.02 \end{bmatrix} + \begin{bmatrix} 0.39 & 0.52 \\ 0.52 & 0.73 \end{bmatrix} \begin{bmatrix} 0.37 \\ -0.16 \end{bmatrix}$$

$$= \begin{bmatrix} 4.81 \\ 7.02 \end{bmatrix} + \begin{bmatrix} 0.06 \\ 0.07 \end{bmatrix} = \begin{bmatrix} 4.87 \\ 7.09 \end{bmatrix}$$

We repeat the process again to obtain the fourth approximation

$$\log_{10}\left(\frac{1}{P_1} - 1\right) = \frac{4.87 - 7.09 \times 1.01}{2.3} = \bar{1}.0040$$

$$\frac{1}{P_1} - 1 = 0.101 \qquad P_1 = 0.91$$

$$\log_{10}\left(\frac{1}{P_2} - 1\right) = \frac{4.87 - 7.09 \times 0.89}{2.3} = \bar{1}.3739$$

$$\frac{1}{P_2} - 1 = 0.237 \qquad P_2 = 0.81$$

$$\log_{10}\left(\frac{1}{P_3} - 1\right) = \frac{4.87 - 7.09 \times 0.71}{2.3} = \bar{1}.9287$$

$$\frac{1}{P_3} - 1 = 0.849 \qquad P_3 = 0.54$$

$$\log_{10}\left(\frac{1}{P_4} - 1\right) = \frac{4.87 - 7.09 \times 0.58}{2.3} = 0.3295$$

$$\frac{1}{P_4} - 1 = 2.13 \qquad\qquad P_4 = 0.32$$

$$\log_{10}\left(\frac{1}{P_5} - 1\right) = \frac{4.87 - 7.09 \times 0.41}{2.3} = 0.8535$$

$$\frac{1}{P_5} - 1 = 7.14 \qquad\qquad P_5 = 12$$ [illegible]

Then

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(3)}} = -0.61$$

$$\left(\frac{\partial \log L}{\partial \beta}\right)_{\beta^{(3)}} = 0.32$$

Hence

$$\begin{bmatrix} \alpha^{(4)} \\ \beta^{(4)} \end{bmatrix} = \begin{bmatrix} 4.87 \\ 7.09 \end{bmatrix} + \begin{bmatrix} 0.39 & 0.52 \\ 0.52 & 0.73 \end{bmatrix} \begin{bmatrix} -0.61 \\ 0.32 \end{bmatrix}$$

$$= \begin{bmatrix} 4.87 \\ 7.09 \end{bmatrix} + \begin{bmatrix} -0.07 \\ -0.08 \end{bmatrix} = \begin{bmatrix} 4.80 \\ 7.01 \end{bmatrix}$$

We repeat the process again to get another approximation.

$$\log_{10}\left(\frac{1}{P_1} - 1\right) = \frac{4.80 - 7.01 \times 1.01}{2.3} = \bar{1}.0087$$

$$\frac{1}{P_1} - 1 = 0.102 \qquad\qquad P_1 = 0.91$$

$$\log_{10}\left(\frac{1}{P_2} - 1\right) = \frac{4.80 - 7.01 \times 0.89}{2.3} = \bar{1}.3744$$

$$\frac{1}{P_2} - 1 = 0.237 \qquad\qquad P_2 = 0.81$$

$$\log_{10}\left(\frac{1}{P_3} - 1\right) = \frac{4.80 - 7.01 \times 0.71}{2.3} = \bar{1}.9230$$

$$\frac{1}{P_3} - 1 = 0.838 \qquad P_3 = 0.54$$

$$\log_{10}\left(\frac{1}{P_4} - 1\right) = \frac{4.80 - 7.01 \times 0.58}{2.3} = 0.3192$$

$$\frac{1}{P_4} - 1 = 2.09 \qquad P_4 = 0.32$$

$$\log_{10}\left(\frac{1}{P_5} - 1\right) = \frac{4.80 - 7.01 \times 0.41}{2.3} = 0.8373$$

$$\frac{1}{P_5} - 1 = 6.88 \qquad P_5 = 0.13$$

Then

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\alpha^{(4)}} = -0.11$$

$$\left(\frac{\partial \log L}{\partial \alpha}\right)_{\beta^{(4)}} = 0.11$$

Hence

$$\begin{bmatrix} \alpha^{(5)} \\ \beta^{(5)} \end{bmatrix} = \begin{bmatrix} 4.80 \\ 7.01 \end{bmatrix} + \begin{bmatrix} 0.39 & 0.52 \\ 0.52 & 0.73 \end{bmatrix} \begin{bmatrix} -0.11 \\ 0.11 \end{bmatrix} = \begin{bmatrix} 4.81 \\ 7.03 \end{bmatrix}$$

We notice here that the estimates $\begin{bmatrix} 4.87 \\ 7.09 \end{bmatrix}$, $\begin{bmatrix} 4.80 \\ 7.01 \end{bmatrix}$, $\begin{bmatrix} 4.81 \\ 7.03 \end{bmatrix}$ are around the estimate $\begin{bmatrix} 4.81 \\ 7.02 \end{bmatrix}$, hence $\alpha^{(2)} = 4.81$ and $\beta^{(2)} = 7.02$ will be the maximum likelihood estimates of α and β, ie. $\overset{x}{\alpha} = 4.81$ and $\overset{x}{\beta} = 7.02$. The value of $\overset{x}{\mu}$ is then

$$\overset{x}{\mu} = \frac{\overset{x}{\alpha}}{\overset{x}{\beta}} = \frac{4.81}{7.02} = 0.685$$

ie. the value of the log dose which kills 50% of the group exposed to the experiment.

Now to get the variance-covariance matrix of maximum likelihood estimates $\overset{x}{\alpha}$ and $\overset{x}{\beta}$, we have to find the P_i's corresponding to x_i's, $\overset{x}{\alpha}$ and $\overset{x}{\beta}$. The values of these P_i's are calculated in the third process of approximation and these are

$P_1 = 0.91$, $P_2 = 0.81$, $P_3 = 0.54$, $P_4 = 0.33$, $P_5 = 0.13$,

then

$$\sum_1^5 n_i P_i (1-P_i) = 39.02$$

$$-\sum_1^5 n_i x_i P_i (1-P_i) = -27.24$$

$$\sum_1^5 n_i x_i^2 P_i (1-P_i) = 20.30$$

Then

$$\underset{\sim}{I}^{-1} = \begin{bmatrix} 39.02 & -27.24 \\ -27.24 & 20.30 \end{bmatrix}^{-1} = \begin{bmatrix} 0.4044 & 0.5426 \\ 0.5426 & 0.7773 \end{bmatrix}$$

Now we use χ^2 to show the association of the observed frequencies with the expecte.

$$\chi^2 = \sum_1^5 \frac{n_i (p_i - P_i)^2}{P_i (1-P_i)}$$

$$= \frac{50(0.88-0.91)}{0.91 \times 0.09} + \frac{49(0.86-0.81)}{0.81 \times 0.19} + \frac{46(0.52-0.54)}{0.54 \times 0.46}$$

$$+ \frac{48(0.33-0.33)}{0.33 \times 0.67} + \frac{50(0.12-0.13)}{0.13 \times 0.87}$$

$$\chi^2 = 0.549 + 0.796 + 0.074 + 0.000 + 0.044$$
$$= 1.463$$

The degrees of freedom of $\chi^2 = 1.463$ are 3, and since $\chi^2_{0.05} = 7.81$ for 3 degrees of freedom, hence the observed frequencies are sufficiently associated with the expected. The variance of $\overset{x}{\mu}$ is given by

$$V_{\overset{x}{\mu}} = \frac{1}{\overset{x}{\beta}{}^2}\left[\frac{1}{\sum n\omega} + \frac{(\overset{x}{\mu}-\bar{x})^2}{\sum n\omega(x-\bar{x})^2}\right],$$

where $\omega = pq$ is obtainable from Table III P.571 in (4)

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = \frac{175.11}{243} = 0.721, \qquad \overset{x}{\beta}{}^2 = 49.28$$

$$(\overset{x}{\mu} - \bar{x})^2 = (0.685 - 0.721)^2 = 0.001296$$

$$\sum_1^5 n_i w_i = 39.33, \qquad \sum_1^5 n_i w_i (x-\bar{x})^2 = 1.315$$

Hence

$$V_{\overset{x}{\mu}} = \frac{1}{49.28}\left[\frac{1}{39.33} + \frac{0.001296}{1.315}\right]$$

$$= 0.00053572$$

and so

$$\overset{x}{\mu} = 0.685 \pm 0.023$$

<u>Example 3.6</u>: This example is on the blood groups where there are three parameters r, p and q which represent the gene frequencies of O, A and B. The expected probabilities of the

six genotypes (four phenotypes) in random mating are found as follows

Phenotype	Genotype	Probability	
O	OO	r^2	
A	AA AO	p^2 $2pr$	p^2+2pr
B	BB BO	q^2 $2qr$	q^2+2qr
AB	AB	$2pq$	

The data is in the following table

Phenotype	O	A	B	AB	TOTAL
Observed	176	182	60	17	435
Expected	nr^2	$n(p^2+2pr)$	$n(q^2+2qr)$	$2pqn$	n

(a) <u>Bernstein's Method:</u>

We can consider the estimates of Bernstein's method as an initial estimates to the maximum likelihood estimates. The estimates of this method are given by

$$r = \left(1+\tfrac{1}{2}D\right)\left(r'+\tfrac{1}{2}D\right)$$

$$p = \left(1+\tfrac{1}{2}D\right)p'$$

$$q = \left(1+\tfrac{1}{2}D\right)q'$$

where

$$-D = r' + p' + q' - 1$$

and

$$r' = \sqrt{\frac{\bar{O}}{n}}, \qquad p' = 1 - \sqrt{\frac{\bar{O}+\bar{B}}{n}}, \qquad q' = 1 - \sqrt{\frac{\bar{O}+\bar{A}}{n}}$$

where $\bar{O}$, $\bar{A}$ and $\bar{B}$ are the observed frequencies. By substituting the observed values we obtain

$r = 0.64234$, $p = 0.26449$, $q = 0.09317$

(b) <u>Maximum Likelihood Method:</u>

The likelihood function is

$$L = (r^2)^{\bar{O}} \, (p^2+2pr)^{\bar{A}} \, (q^2+2qr)^{\bar{B}} \, (2pq)^{\overline{AB}} \times C$$

where C is constant. Now we can put the probabilities as follows

$$\theta_1 = r^2$$
$$\theta_2 = (1-q)^2 - r^2$$
$$\theta_3 = q^2 + 2q(1-p-q)$$
$$\theta_4 = 2pq$$

for the partial derivative of θ_i with respect to p which is desired to be put in the form $\frac{\partial \theta_i}{\partial p} = \frac{\partial \theta_i}{\partial r} \frac{\partial r}{\partial p}$,

and

$$\theta_1 = r^2$$
$$\theta_2 = p^2 + 2p(1-p-q)$$
$$\theta_3 = (1-p)^2 - r^2$$
$$\theta_4 = 2pq$$

for the partial derivative of θ_i with respect to q which is desired to be put in the form $\frac{\partial \theta_i}{\partial q} = \frac{\partial \theta_i}{\partial r} \frac{\partial r}{\partial q}$.

Then by taking the log of the likelihood function and differentiating with respect to p and q as independent parameters we get

$$\frac{\partial \log L}{\partial p} = \frac{\bar{O}}{\theta_1}(-2r) + \frac{\bar{A}}{\theta_2}(2r) + \frac{\bar{B}}{\theta_3}(-2q) + \frac{\overline{AB}}{\theta_4}(2q)$$

and

$$\frac{\partial \log L}{\partial q} = \frac{\bar{O}}{\theta_1}(-2r) + \frac{\bar{A}}{\theta_2}(-2p) + \frac{\bar{B}}{\theta_3}(2r) + \frac{\overline{AB}}{\theta_4}(2p)$$

By substituting the known values we get

$$\left(\frac{\partial \log L}{\partial p}\right)_{p^{(0)}} = (-3.11362)176 + (3.13543)182 + (-1.45217)60 + (3.75086)17$$
$$= -0.20444$$

$$\left(\frac{\partial \log L}{\partial q}\right)_{q^{(0)}} = (-3.11362)176 + (-1.27104)182 + (10.00685)60 + (10.73307)1$$
$$= -0.09321$$

where $p^{(0)}$ and $q^{(0)}$ are the Brunstein's Method estimates.

To get the information matrix we have to find

$$\frac{\partial^2 \log L}{\partial p^2} = -\left[\frac{\bar{O}}{\theta_1^2}(-2r)^2 + \frac{\bar{A}}{\theta_2^2}(2r)^2 + \frac{\bar{B}}{\theta_3^2}(-2q)^2 + \frac{\overline{AB}}{\theta_4^2}(2q)^2\right]$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial p^2}\right) = \frac{1}{\theta_1}(-2r)^2 + \frac{1}{\theta_2}(2r)^2 + \frac{1}{\theta_3}(-2q)^2 + \frac{1}{\theta_4}(2q)^2$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial q^2}\right) = \frac{1}{\theta_1}(-2r)^2 + \frac{1}{\theta_2}(-2p)^2 + \frac{1}{\theta_3}(2r)^2 + \frac{1}{\theta_4}(2p)^2$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial p\,\partial q}\right) = \frac{1}{\theta_1}(-2r)^2 + \frac{1}{\theta_2}(2r)(-2p) + \frac{1}{\theta_3}(-2q)(2r) + \frac{1}{\theta_4}(2q)(2p)$$

By substituting the values of $r^{(0)}$, $p^{(0)}$, $q^{(0)}$ and n we get

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial p^2}\right) = 435 \times 9.00315 \times \frac{1}{435} = 9.00315$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial q^2}\right) = 435 \times 23.21612 \times \frac{1}{435} = 23.21612$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial p \partial q}\right) = 435 \times 2.47676 \times \frac{1}{435} = 2.47676$$

The information matrix is then

$$\frac{1}{n}\underset{\sim}{I} = \begin{bmatrix} 9.00315 & 2.47676 \\ 2.47676 & 23.21612 \end{bmatrix}$$

and so

$$\underset{\sim}{I}^{-1} = \frac{1}{435\,\Delta}\begin{bmatrix} 23.21612 & -2.47676 \\ -2.47676 & 9.00315 \end{bmatrix}$$

$$= \begin{bmatrix} 0.00026305 & -0.00002806 \\ -0.00002806 & 0.00010202 \end{bmatrix}$$

where Δ is the determinant of $\frac{1}{435}\underset{\sim}{I}$. Then the first approximation is given by

$$\begin{bmatrix} p^{(1)} \\ q^{(1)} \end{bmatrix} = \begin{bmatrix} p^{(0)} \\ q^{(0)} \end{bmatrix} + \underset{\sim}{I}^{-1}_{(0)} \begin{bmatrix} \frac{\partial \log L}{\partial p} \\ \frac{\partial \log L}{\partial q} \end{bmatrix}_{(0)}$$

$$= \begin{bmatrix} 0.26449 \\ 0.09317 \end{bmatrix} + \begin{bmatrix} 0.00026305 & -0.00002806 \\ -0.00002806 & 0.00010202 \end{bmatrix} \begin{bmatrix} -0.20444 \\ -0.09321 \end{bmatrix}$$

$$\begin{bmatrix} p^{(1)} \\ q^{(1)} \end{bmatrix} = \begin{bmatrix} 0.26449 \\ 0.09317 \end{bmatrix} + \begin{bmatrix} -0.00005116 \\ -0.00000377 \end{bmatrix}$$

$$= \begin{bmatrix} 0.26444 \\ 0.09317 \end{bmatrix}$$

Since the corrections are very small then

$p = 0.26444$,

$q = 0.09317$,

and $r = 1-(p+q) = 0.64239$

are the maximum likelihood estimates.

Here the variance-covariance matrix of $\hat{p}$ and $\hat{q}$ is

$$\begin{bmatrix} 0.00026305 & -0.00002806 \\ -0.00002806 & 0.00010202 \end{bmatrix}$$

and the variance of is given by

$$V_{\hat{r}} = 10^{-8}\ (26305 + 10202) + 2 \times 10^{-8}\ (-2806)$$

$$= 0.00030893$$

The following table shows the results obtained

Parameter	Estimate	Variance
$\hat{p}$	0.26444	0.00026305
$\hat{q}$	0.09317	0.00010202
$\hat{r}$	0.64239	0.00030893

(c) Wald Technique:

Let $\theta_1 = r^2$, $\theta_2 = p^2 + 2pr$, $\theta_3 = q^2 + 2qr$ and $\theta_4 = 2pq$ and let n_i, $i = 1, \dots, 4$ be the observed frequency for θ_i. Then the likelihood function is

$$L = \left(\frac{\theta_1}{\sum \theta_i}\right)^{n_1} \left(\frac{\theta_2}{\sum \theta_i}\right)^{n_2} \left(\frac{\theta_3}{\sum \theta_i}\right)^{n_3} \left(\frac{\theta_4}{\sum \theta_i}\right)^{n_4}$$

where $\sum \theta_i = 1$ is the imposed restriction for identifiability of the four parameters. Then

$$\log L = \sum_1^4 n_i \log \theta_i - n \log \sum_1^4 \theta_i$$

$$\frac{\partial \log L}{\partial \theta_i} = \frac{n_i}{\theta_i} - \frac{n}{\sum_1^4 \theta_i}.$$

By equating the last eqution to zero we get

$$\overset{\times}{\theta}_i = \frac{n_i}{n}$$

Therefore

$$\overset{\times}{\theta}_1 = \frac{n_1}{n} = \frac{176}{435} = 0.40459$$

$$\overset{\times}{\theta}_2 = \frac{n_2}{n} = \frac{182}{435} = 0.41840$$

$$\overset{\times}{\theta}_3 = \frac{n_3}{n} = \frac{60}{435} = 0.13793$$

$$\overset{\times}{\theta}_4 = \frac{n_4}{n} = \frac{17}{435} = 0.03908$$

Now we have

$$\theta_1 = r^2, \quad \theta_2 = p^2 + 2pr, \quad \theta_3 = q^2 + 2qr, \quad \theta_4 = 2pq$$

then

$$\sqrt{\theta_1} = r, \quad \sqrt{\theta_1 + \theta_2} = p + r, \quad \sqrt{\theta_1 + \theta_3} = q + r$$

and since

$$p + q + r = 1$$

then

$$\sqrt{\theta_1 + \theta_2} - \sqrt{\theta_1} + \sqrt{\theta_1 + \theta_3} - \sqrt{\theta_1} + \sqrt{\theta_1} = 1$$

ie.

$$\sqrt{\theta_1 + \theta_2} + \sqrt{\theta_1 + \theta_3} - \sqrt{\theta_1} - 1 = 0 = h(\theta)$$

that is we have one restriction for the unrestricted parameters. Now we have to test the null hypothesis by asking whether the estimates of the unrestricted parameters $\theta_i, i = 1, \ldots, 4$ satisfy the restriction above. The statistic of Wald test is

$$n\, h'(\check{\theta}) \left[H'_{\check{\theta}} \left(\frac{1}{n} \underset{\sim}{I}_{\check{\theta}} + H_{1\check{\theta}} H'_{1\check{\theta}} \right)^{-1} H_{\check{\theta}} \right]^{-1} h(\check{\theta})$$

The restrictions including the identifiable are

$$h_1(\check{\theta}) = \sum_1^4 \check{\theta}_i - 1 = 0$$

$$h_2(\check{\theta}) = \sqrt{\check{\theta}_1 + \check{\theta}_2} + \sqrt{\check{\theta}_1 + \check{\theta}_3} - \sqrt{\check{\theta}_1} - 1 = 0.008$$

then

$$h(\check{\theta}) = \begin{bmatrix} 0 \\ 0.008 \end{bmatrix}, \qquad h'(\check{\theta}) = \begin{bmatrix} 0 & 0.008 \end{bmatrix}$$

and

$$H'_{\check{\theta}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0.44 & 0.55 & 0.68 & 0 \end{bmatrix}$$

Also

$$\frac{1}{n}\underset{\sim}{I}_{\overset{x}{\theta}} = \begin{bmatrix} \frac{1}{\overset{x}{\theta}_1}-1 & -1 & -1 & -1 \\ -1 & \frac{1}{\overset{x}{\theta}_2}-1 & -1 & -1 \\ -1 & -1 & \frac{1}{\overset{x}{\theta}_3}-1 & -1 \\ -1 & -1 & -1 & \frac{1}{\overset{x}{\theta}_4}-1 \end{bmatrix}$$

then

$$\left(\frac{1}{n}\underset{\sim}{I}_{\overset{x}{\theta}} + H_{1\overset{x}{\theta}} H'_{1\overset{x}{\theta}}\right)^{-1} = \begin{bmatrix} 0.40459 & 0 & 0 & 0 \\ 0 & 0.41840 & 0 & 0 \\ 0 & 0 & 0.13793 & 0 \\ 0 & 0 & 0 & 0.03908 \end{bmatrix}$$

Hence

$$H'_{\overset{x}{\theta}}\left(\frac{1}{n}\underset{\sim}{I}_{\overset{x}{\theta}} + H_{1\overset{x}{\theta}} H'_{1\overset{x}{\theta}}\right)^{-1} H_{\overset{x}{\theta}}$$

$$= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0.44 & 0.55 & 0.68 & 0 \end{bmatrix} \begin{bmatrix} 0.40459 & 0 & 0 & 0 \\ 0 & 0.41840 & 0 & 0 \\ 0 & 0 & 0.13793 & 0 \\ 0 & 0 & 0 & 0.03908 \end{bmatrix} \begin{bmatrix} 1 & 0.44 \\ 1 & 0.55 \\ 1 & 0.68 \\ 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.40459 & 0.41840 & 0.13793 & 0.03908 \\ 0.178 & 0.23 & 0.0938 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0.44 \\ 1 & 0.55 \\ 1 & 0.68 \\ 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.27 \end{bmatrix}.$$

Therefore

$$n\,h'(\overset{*}{\theta})\left[H'_{\overset{*}{\theta}}\left(\tfrac{1}{n}\underset{\sim}{I}_{\overset{*}{\theta}}+H_{1\overset{*}{\theta}}H'_{1\overset{*}{\theta}}\right)^{-1}H_{\overset{*}{\theta}}\right]^{-1}h(\overset{*}{\theta})$$

$$= 435\begin{bmatrix}0 & 0.008\end{bmatrix}\begin{bmatrix}1 & 0.5\\ 0.5 & 0.27\end{bmatrix}^{-1}\begin{bmatrix}0\\ 0.008\end{bmatrix}$$

$$= 435 \times (0.02)^{-1} \times 64 \times 10^{-6} = 1.392$$

We have from Statistical Tables that $\chi^2_{0.05} = 3.84$ for one degree of freedom. Since

$$(\chi^2 = 1.392) < (\chi^2_{0.05} = 3.84),$$

we accept the null hypothesis on 5% level of significance.

(d) Lagrange Multiplier Technique

To apply Lagrange multiplier technique we consider the probabilities

$$\frac{r^2}{(p+q+r)^2},\quad \frac{p^2+2pr}{(p+q+r)^2},\quad \frac{q^2+2qr}{(p+q+r)^2}\quad \text{and}\quad \frac{2pq}{(p+q+r)^2}.$$

The likelihood function is then

$$L=\left(\frac{r^2}{(p+q+r)^2}\right)^{176}\left(\frac{p^2+2pr}{(p+q+r)^2}\right)^{182}\left(\frac{q^2+2qr}{(p+q+r)^2}\right)^{60}\left(\frac{2pq}{(p+q+r)^2}\right)^{17}$$

and

$$\log L = 2 \times 176 \log r + 182 \log(p^2+2pr) + 60 \log(q^2+2qr)$$
$$+ 17 \log 2pq - 2 \times 435 \log(p+q+r)$$

Differentiating with respect to r, p and q we get

$$\frac{\partial \log L}{\partial r} = 2\left\{\frac{176}{r} + \frac{182}{p+2r} + \frac{60}{q+2r} - \frac{435}{p+q+r}\right\}$$

$$\frac{\partial \log L}{\partial p} = 2\left\{\frac{182(p+r)}{p^2+2pr} + \frac{17}{2p} - \frac{435}{p+q+r}\right\}$$

$$\frac{\partial \log L}{\partial q} = 2\left\{\frac{60(q+r)}{q^2+2qr} + \frac{17}{2q} - \frac{435}{p+q+r}\right\}$$

Differentiating again with respect to r, p and q and taking theexpected value we get

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial r^2}\right) = 2\left\{1 + \frac{2p}{p+2r} + \frac{2q}{q+2r} - \frac{1}{(p+q+r)^2}\right\}$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial p^2}\right) = 2\left\{\frac{(p+r)^2}{p^2+2pr} + \frac{q}{p} - \frac{1}{(p+q+r)^2}\right\}$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial q^2}\right) = 2\left\{\frac{(q+r)^2}{q^2+2qr} + \frac{p}{q} - \frac{1}{(p+q+r)^2}\right\}$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial r\,\partial p}\right) = 2\left\{\frac{p}{p+2r} - \frac{1}{(p+q+r)^2}\right\}$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial r\,\partial q}\right) = 2\left\{\frac{q}{q+2r} - \frac{1}{(p+q+r)^2}\right\}$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial p\,\partial q}\right) = 2\left\{-\frac{1}{(p+q+r)^2}\right\}.$$

Consider the Bernstein method estimates as initial estimates of the maximum likelihood estimates. Then substituting these estimates which are

$r = 0.64234$, $p = 0.26449$, $q = 0.09317$

in the equations above we get

$$\frac{1}{n}\frac{\partial \log L}{\partial r} = \frac{0.053}{435}$$

$$\frac{1}{n}\frac{\partial \log L}{\partial p} = \frac{-0.126}{435}$$

$$\frac{1}{n}\frac{\partial \log L}{\partial q} = \frac{-0.007}{435}$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial r^2}\right) = 0.952$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial p^2}\right) = 2.72$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial q^2}\right) = 12.104$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial r\,\partial p}\right) = -1.658$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial r\,\partial q}\right) = -1.8648$$

$$-\frac{1}{n}E\left(\frac{\partial^2 \log L}{\partial p\,\partial q}\right) = -2$$

We have here only the identifiable restriction

$$h(\theta) = r + p + q - 1 = 0 \quad ,$$

therefore

$$H(\theta) = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \quad ,$$

and

$$\frac{1}{n}\underset{\sim}{I}_{\theta^{(0)}} + H_{1\theta^{(0)}} H'_{1\theta^{(0)}} = \begin{bmatrix} 0.952 & -1.658 & -1.865 \\ -1.658 & 2.72 & -2 \\ -1.865 & -2 & 12.104 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Then

$$\left(\frac{1}{n}\underset{\sim}{I}_{\theta^{(0)}} + H_{1\theta^{(0)}} H'_{1\theta^{(0)}}\right)^{-1} = \begin{bmatrix} 1.952 & -0.658 & -0.865 \\ -0.658 & 3.72 & 1 \\ -0.865 & 1 & 13.104 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.57 & 0.114 & 0.047 \\ 0.114 & 0.3 & 0.03 \\ 0.047 & 0.03 & 0.082 \end{bmatrix}$$

Then the first approximation will be given by

$$\begin{bmatrix} r^{(1)} \\ p^{(1)} \\ q^{(1)} \end{bmatrix} = \begin{bmatrix} 0.64234 \\ 0.26449 \\ 0.09317 \end{bmatrix} + \begin{bmatrix} 0.57 & 0.114 & 0.047 \\ 0.114 & 0.3 & 0.03 \\ 0.047 & 0.03 & 0.082 \end{bmatrix} \begin{bmatrix} 0.053 \\ -0.126 \\ -0.007 \end{bmatrix} \frac{1}{435}$$

then

$$\begin{bmatrix} r^{(1)} \\ p^{(1)} \\ q^{(1)} \end{bmatrix} = \begin{bmatrix} 0.64234 \\ 0.26449 \\ 0.09317 \end{bmatrix} + \begin{bmatrix} 0.0000356 \\ -0.0000734 \\ -0.0000042 \end{bmatrix}$$

$$= \begin{bmatrix} 0.64238 \\ 0.26442 \\ 0.09317 \end{bmatrix}$$

We repeat the process again to get the second approximation, using the new estimates (first approximation estimates). Then we find that

$$\left(\frac{\partial \log L}{\partial r}\right)_{r^{(1)}} = -0.0136$$

$$\left(\frac{\partial \log L}{\partial p}\right)_{p^{(1)}} = 0.046$$

$$\left(\frac{\partial \log L}{\partial q}\right)_{q^{(1)}} = -0.036$$

and so the second approximation will be given by

$$\begin{bmatrix} r^{(2)} \\ p^{(2)} \\ q^{(2)} \end{bmatrix} = \begin{bmatrix} 0.64238 \\ 0.26442 \\ 0.09317 \end{bmatrix} + \begin{bmatrix} 0.57 & 0.114 & 0.047 \\ 0.114 & 0.3 & 0.03 \\ 0.047 & 0.03 & 0.082 \end{bmatrix} \begin{bmatrix} -0.014 \\ 0.046 \\ -0.036 \end{bmatrix} \frac{1}{435}$$

$$= \begin{bmatrix} 0.64238 \\ 0.26442 \\ 0.09317 \end{bmatrix} + \begin{bmatrix} -0.00001 \\ 0.00003 \\ -0.00001 \end{bmatrix} = \begin{bmatrix} 0.64237 \\ 0.26445 \\ 0.09316 \end{bmatrix}$$

By repeating the process again using the new estimates we get

$$\left(\frac{\partial \log L}{\partial r}\right)_{r^{(2)}} = 0.0032$$

$$\left(\frac{\partial \log L}{\partial p}\right)_{p^{(2)}} = -0.032$$

$$\left(\frac{\partial \log L}{\partial q}\right)_{q^{(2)}} = 0.057$$

and so the third approximation will be given by

$$\begin{bmatrix} r^{(3)} \\ p^{(3)} \\ q^{(3)} \end{bmatrix} = \begin{bmatrix} 0.64237 \\ 0.26445 \\ 0.09316 \end{bmatrix} + \begin{bmatrix} 0.57 & 0.114 & 0.047 \\ 0.114 & 0.3 & 0.03 \\ 0.047 & 0.03 & 0.082 \end{bmatrix} \begin{bmatrix} 0.003 \\ -0.032 \\ 0.057 \end{bmatrix} \frac{1}{435}$$

$$= \begin{bmatrix} 0.64237 \\ 0.26445 \\ 0.09316 \end{bmatrix} + \begin{bmatrix} 0.000002 \\ -0.000017 \\ 0.000008 \end{bmatrix} = \begin{bmatrix} 0.64237 \\ 0.26443 \\ 0.09317 \end{bmatrix}$$

Repeating the process again using the new estimates we get

$$\left(\frac{\partial \log L}{\partial r}\right)_{r^{(3)}} = -0.0023$$

$$\left(\frac{\partial \log L}{\partial p}\right)_{p^{(3)}} = 0.018$$

$$\left(\frac{\partial \log L}{\partial q}\right)_{q^{(3)}} = -0.035$$

Then

$$\begin{bmatrix} r^{(4)} \\ p^{(4)} \\ q^{(4)} \end{bmatrix} = \begin{bmatrix} 0.64237 \\ 0.26443 \\ 0.09317 \end{bmatrix} + \begin{bmatrix} 0.57 & 0.114 & 0.047 \\ 0.114 & 0.3 & 0.03 \\ 0.047 & 0.03 & 0.082 \end{bmatrix} \begin{bmatrix} -0.002 \\ 0.018 \\ -0.035 \end{bmatrix} \frac{1}{435}$$

$$= \begin{bmatrix} 0.64237 \\ 0.26443 \\ 0.09317 \end{bmatrix} + \begin{bmatrix} -0.000002 \\ 0.000010 \\ -0.000010 \end{bmatrix}$$

$$\begin{bmatrix} r^{(4)} \\ p^{(4)} \\ q^{(4)} \end{bmatrix} = \begin{bmatrix} 0.64237 \\ 0.26444 \\ 0.09316 \end{bmatrix}$$

We see here that the sets of estimates of r, p and q obtained by the four successive approximations are slightly different from each other and they are close to the estimates obtained by the technique of ~~2.2~~ (b) above. In fact in this case, the obtaining of the accurate estimates to five decimal points is unlikely and so it is unlikely that the estimates of ~~2.2~~ (b) can be arrived at in which the two parameters p and q are considered to be independent and r is kept as dependent since $p + q + r = 1$. Hence if we approximate the estimates of ~~2.2~~(b) and each set of estimates of ~~2.2~~ (d) to four decimal points, we will find the estimates of each set are equal to the corresponding estimates of the others, except the estimate of p in the set of the second approximation in which $p^{(2)} = 0.26445$. Therefore, we will consider that the maximum likelihood estimates of the restricted parameters r, p and q are

$$\begin{bmatrix} r^{\otimes} \\ p^{\otimes} \\ q^{\otimes} \end{bmatrix} = \begin{bmatrix} 0.6424 \\ 0.2644 \\ 0.0932 \end{bmatrix}$$

Now we test the hypothesis by asking whether these restricted estimates are sufficiently near to the maximum likelihood estimates. Since the Bernstein estimates are very close to the

Lagrange multiplier estimates, therefore we will use the variance-covariance matrix of Bernstein's estimates as the variance-covariance matrix of Lagrange multiplier estimates, the statistic of Lagrange multiplier test is

$$\frac{1}{n}\begin{bmatrix} \frac{\partial \log L}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L}{\partial \theta_m} \end{bmatrix}'_{\overset{\otimes}{\theta}} \left[\frac{1}{n}\underset{\sim}{I}_\theta + H_{1\theta}H_{1\theta}'\right]^{-1}_{\overset{(B)}{\theta}} \begin{bmatrix} \frac{\partial \log L}{\partial \theta_1} \\ \vdots \\ \frac{\partial \log L}{\partial \theta_m} \end{bmatrix}_{\overset{\otimes}{\theta}}$$

where $\overset{(B)}{\theta}$ denotes the Bernstein's estimates. Here we have

$$\left(\frac{\partial \log L}{\partial r}\right)_{\overset{\otimes}{r}} = -0.012$$

$$\left(\frac{\partial \log L}{\partial p}\right)_{\overset{\otimes}{p}} = 0.127$$

$$\left(\frac{\partial \log L}{\partial q}\right)_{\overset{\otimes}{q}} = -0.278$$

Since the Lagrange multiplier statistic is, in our example, distributed as χ^2- distribution with one degree of freedom then,

$$\chi^2_{[1]} = \frac{10^{-6}}{435}\begin{bmatrix} -12 & 127 & -278 \end{bmatrix}\begin{bmatrix} 0.57 & 0.114 & 0.047 \\ 0.114 & 0.3 & 0.03 \\ 0.047 & 0.03 & 0.082 \end{bmatrix}\begin{bmatrix} -12 \\ 127 \\ -278 \end{bmatrix}$$

$$= 21 \times 10^{-6}$$

We have that $\chi^2_{0.05} = 3.84$ for one degree of freedom, and since

$$\left(\chi^2_{[1]} = 21 \times 10^{-6}\right) < \left(\chi^2_{0.05} = 3.84\right)$$

we accept the hypothesis on 5% level of significance. Furthermore, the hypothesis is accepted on 99.5% level of significance.

The variance-covariance matrix of $\overset{\otimes}{r}$, $\overset{\otimes}{p}$ and $\overset{\otimes}{q}$ will be given by $\frac{1}{n}\bar{A}_{\theta}^{(8)}$, where $\theta^{(8)} = \overset{\otimes}{\theta}$, as it denoted in 9. Ch. II. The Procedure of getting $\bar{A}_{\theta}^{(8)}$ is discussed in 8. Ch. II.

Here

$$\frac{1}{n}\bar{A}_{\theta}^{(8)} = \begin{bmatrix} 0.00039 & -0.00029 & -0.000092 \\ -0.00029 & 0.00035 & -0.000053 \\ -0.000092 & -0.000053 & 0.00014 \end{bmatrix}$$

If we look back at the variances of $\overset{\times}{r}$, $\overset{\times}{p}$, $\overset{\times}{q}$ which obtained in ~~2.2~~ (b) we will see that the variances of $\overset{\otimes}{r}$, $\overset{\otimes}{p}$, $\overset{\otimes}{q}$ are slightly larger than of $\overset{\times}{r}$, $\overset{\times}{p}$, $\hat{q}$ by the fifth decimal points. The reason for these differences is of course due to the operation of the approximations to the numbers used for the whole work of this technique.

CHAPTER IV

LIKELIHOOD RATIO TEST

1. Introduction:

The following important definitions are worth mentioning.

Definition 1. If CR is the critical region of the test (the critical region of rejection of the null hypothesis H_0 against the alternative hypothesis H_1), then P(CR:H_0), the probability of rejecting H_0 against H_1 (no matter which one is true) is called the power function. The value of P(CR:H_0) at the parameter point is called the power function of the test at that value of the parameter.

Definition 2. Let α be the probability of rejecting H_0 against H_1 when H_0 is true. Then α is called the significance level of the test, or the size of the test.

Definition 3. A test is said to be unbiased if

$$P(CR:H_0)\ (H_1 \text{ is true}) > P(CR:H_0)\ (H_0 \text{ is true}) .$$

Definition 4. If there are two tests with the same size, and if

$$P_1(CR:H_0) > P_2(CR:H_0), \qquad H_1 \text{ is true}$$

then the first test is said to be uniformly more powerful than the second. Hence if there are n tests with thesame size, then the one which is uniformly more powerful than each one of the n tests is called the uniformly most powerful test.

The likelihood ratio test is related to the maximum likelihood method of estimation and it is modified by the Neyman-Pearson theory of testing the statistical hypothesis.

It has been shown that likelihood ratio test is the uniformly most powerful test if such exists. In (7) and (20) it has been discussed that the likelihood ratio test has the property of unbiasedness. It is worth while showing that this test is based on a sufficient statistic if such exists.

Let $x_1, ----, x_n$, be a random sample drawn from a population has a distribution defined by $f(x,\theta)$, and let $t(x_1,--,x_n)$ be a sufficient statistic for θ. Then the likelihood function will be factorised such that

$$L(x,\theta) = L_1(t,\theta)\, L_2(x_1,--,x_n)$$

if we denote by $L(x,\check{\theta})$ the maximum of $L(x,\theta)$ specified by the null hypothesis H_0, and $L(x,\hat{\theta})$ the maximum of $L(x,\theta)$ specified by the whole space of the parameters, then the likelihood ratio test as we will show later λ is given by

$$\lambda = \frac{L(x,\check{\theta})}{L(x,\hat{\theta})} .$$

Then

$$\lambda = \frac{L_1(t,\check{\theta})\, L_2(x_1,---,x_n)}{L_1(t,\hat{\theta})\, L_2(x_1,--,x_n)} = \frac{L_1(t,\check{\theta})}{L_1(t,\hat{\theta})}$$

since the numerator and the denominator are both functions of a sufficient statistic, then λ will be a function of a sufficient statistic, and so the likelihood ratio test is based on a sufficient statistic.

Let $x_1, ----, x_n$ be a random sample drawn from a population with probability density function defined by $f(x;\theta_1,---,\theta_m)$ and let Ω denotes the whole space of the m parameters and ω the subspace specified by the null hypothesis H_0. Then the alternative hypothesis H_1 will be specified by the subspace $\Omega - \omega$. Let $L(\Omega)$ denotes the likelihood function designated by the whole space of the m parameters and $L(\omega)$ denotes the likelihood function designated by ω. Then the likelihood ratio test is defined by the statistic

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})},$$

where $L(\hat{\omega})$ and $L(\hat{\Omega})$ are the maximum of $L(\omega)$ and $L(\Omega)$ respectively. Since each of $L(\hat{\omega})$ and $L(\hat{\Omega})$ is positive and $L(\hat{\omega})$ is a subset of $L(\hat{\Omega})$, then $0 \leq \lambda \leq 1$ and the critical region for the test will be defined by $0 \leq \lambda \leq \lambda_\alpha$ where λ_α is a proper fraction accordingly to the desirable probability α which is as defined in definition 2. Therefore, we reject the null hypothesis H_0, if, and only if,

$$\lambda \leq \lambda_\alpha$$

It has been shown by S. S. Wilks (22), that for large samples and under some conditions, $-2 \log \lambda$ is distributed as χ^2 - distribution with $m-r$ degrees of freedom, where r is the numbe of the parameters after the restrictions; ie. if K is the number of the parameters which specify the null hypothesis H_0, then $k+r=m$, (Appendix II). We will show in the following sections that λ or the function of λ is distributed as t - distribution and F - distribution, also we will show that $-2 \log \lambda$ has χ^2 - distribution.

2. A test of the Significance of the Population Mean:

(a) H_0 Simple and H_1 Composite:

Let $x_1, ----, x_n$ be a random sample drawn from a population distributed normally with unknown mean μ and known variance σ_0^2.

Here $H_0 : \mu = \mu_0$ will be tested against $H_1 : \mu \neq \mu_0$. The space Ω and the subspace ω are then

$$\Omega = \left\{ (\mu, \sigma_0^2) : -\infty < \mu < \infty , \quad 0 < \sigma_0^2 < \infty \right\}$$

and

$$\omega = \left\{ (\mu, \sigma_0^2) : \mu = \mu_0 , \quad 0 < \sigma_0^2 < \infty \right\}$$

Then

$$L(\Omega) = \frac{1}{(2\pi\sigma_0^2)^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2\sigma_0^2} \sum (x-\mu)^2 \right\}$$

$$L(\omega) = \frac{1}{(2\pi\sigma_0^2)^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2\sigma_0^2} \sum (x-\mu_0)^2 \right\}$$

and so

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \exp\left\{ \frac{1}{2\sigma_0^2} \sum (x-\bar{x})^2 - \frac{1}{2\sigma_0^2} \sum (x-\mu_0)^2 \right\}$$

$$= \exp\left\{ -\frac{n}{2\sigma_0^2} (\bar{x}-\mu_0)^2 \right\} \Rightarrow -2 \log \lambda = \frac{n}{\sigma_0^2} (\bar{x}-\mu_0)^2$$

Since $\left\{ \frac{\sqrt{n}}{\sigma_0} (\bar{x}-\mu_0) \right\}^2$ is distributed as χ^2-distribution with 1 degrees of freedom, therefore we reject the hypothesis H_0 if, and only if , $\chi^2 \geq \chi^2_\alpha$.

(b) <u>H_0 and H_1 are both composite:</u>

Let $x_1 , \text{----} , x_n$ be a random sample drawn from a population distributed normally with unknown mean μ and unknown variance σ^2 . Here the null hypothesis $H_0 : \mu = \mu_0$ will be tested against the alternative hypothesis $H_1 : \mu \neq \mu_0$, and so Ω and ω will be such that

$$\Omega = \left\{ (\mu, \sigma^2) : -\infty < \mu < \infty, \quad 0 < \sigma^2 < \infty \right\}$$

$$\omega = \left\{ (\mu, \sigma^2) : \mu = \mu_0, \quad 0 < \sigma^2 < \infty \right\}$$

Then by getting the maximum likelihood estimates for the required parameters, we obtain

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left[\frac{\Sigma (x - \bar{x})^2}{\Sigma (x - \mu_0)^2} \right]^{\frac{n}{2}}$$

$$= \frac{1}{\left[1 + \frac{n(\bar{x} - \mu_0)^2}{\Sigma (x - \bar{x})^2} \right]^{\frac{n}{2}}} = \frac{1}{\left[1 + \frac{t^2}{n-1} \right]^{\frac{n}{2}}}$$

and so the likelihood ratio test will be based on the statistic t, therefore we reject H_0 if and only if,

$$|t| \geqslant t_\alpha$$

Since $F = t^2$, then we can say that H_0 will be rejected if and only if,

$$(t^2 = F) \geqslant (t^2_\alpha = F_\alpha)$$

Example 4.1

If $x_1, ----, x_{25}$ is a random sample from a population having a distribution defined by $N(x; \theta, 4)$ and the sample mean $\bar{x} = 1$. Test the null simple hypothesis $H_0 : \theta = 0$, against the alternative composite hypothesis $H_1 : \theta > 0$. Use the significance level of the test $\alpha = 0.05$.

We have from (a) section 2 that

$$\chi^2 = \left\{ \frac{\sqrt{n}\,(\bar{x} - \mu_0)}{\sigma_0} \right\}^2$$

is distributed as χ^2- distribution with 1 degrees of freedom and H_0 will be rejected if, and only if,

$$\chi^2 \geqslant \chi^2_\alpha$$

Here

$$\chi^2 = \left\{ \frac{\sqrt{25}\,(1-0)}{2} \right\}^2 = \left\{ \frac{5}{2} \right\}^2 = (2.5)^2$$

From the statistical table $\chi^2_{0.05} = 3.84$ for 1 degrees of freedom. We find here that

$$\left(\chi^2 = (2.5)^2\right) > \left(\chi^2_{0.05} = 3.84\right)$$

therefore we reject the hypothesis H_0 in favour of the alternative hypothesis H_1 on 5% level of significance.

Example 4.2

Let $x_1, ---, x_{10}$ be a random sample from a population which has a distribution defined by $N(\theta, \sigma^2)$, and let the sample mean $\bar{x} = 0.6$ and $\sum_1^{10} (x_i - \bar{x})^2 = 3.6$. Test the null composite hypothesis $H_0 : \theta = 0$ against the alternative composite hypothesis $H_1 : \theta \neq 0$ at the 5% significance level.

We have from (b) section 2 that

$$t = \frac{\sqrt{n}\,(\bar{x} - \mu_0)}{\sqrt{\left(\frac{\sum_1^n (x_i - \bar{x})^2}{n-1}\right)}}$$

is distributed as t- distribution with $n-1$ degrees of freedom, By substituting the observed values we get

$$t = \frac{\sqrt{10}\,(0.6 - 0)}{\sqrt{\left(\frac{3.6}{9}\right)}} = 3$$

We have $t_{0.05} = 2.26$ for 9 degrees of freedom (Statistical Table). Since

$$(t = 3) > (t_{0.05} = 2.26)$$

we reject the null composite hypothesis H_0 in favour of the alternative composite hypothesis H_1 on 5% significance level.

3. <u>The test of the Equality of two populations means:</u>

Let $x_1, -----, x_n$ and $y_1, ----, y_m$ are two random samples drawn from two populations with probability density functions $f_1(x; \mu_1, \sigma^2)$ and $f_2(y; \mu_2, \sigma^2)$. We have to test the null composite hypothesis $H_0 : \mu_1 = \mu_2 = \mu$ against the alternative composite hypothesis $H_1 : \mu_1 \neq \mu_2$. Here

$$\Omega = \left\{(\mu_1, \mu_2, \sigma^2) : -\infty < \mu_1 < \infty\ ,\ -\infty < \mu_2 < \infty,\ 0 < \sigma^2 < \infty\right\}$$

$$\omega = \left\{(\mu_1, \mu_2, \sigma^2) : -\infty < \mu_1 = \mu_2 = \mu < \infty,\quad 0 < \sigma^2 < \infty\right\}$$

and $x_1, \ldots, x_n, y_1, \ldots, y_m$ are $n+m$ random variables then

$$L(\omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m+n}{2}} \exp\left\{-\frac{\sum_1^n (x-\mu)^2 + \sum_1^m (y-\mu)^2}{2\sigma^2}\right\}$$

and

$$L(\Omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m+n}{2}} \exp\left\{-\frac{\sum_1^n (x-\mu_1)^2 + \sum_1^m (y-\mu_2)^2}{2\sigma^2}\right\}$$

By solving the equations

$$\frac{\partial \log L(\omega)}{\partial \mu} = 0, \qquad \frac{\partial \log L(\omega)}{\partial \sigma^2} = 0,$$

$$\frac{\partial \log L(\Omega)}{\partial \mu_i} = 0, \quad (i = 1, 2), \quad \frac{\partial \log L(\Omega)}{\partial \sigma^2} = 0$$

we get the maximum likelihood estimates of these parameters, and then substituting these estimates in $L(\omega)$ and $L(\Omega)$ we obtain $L(\hat{\omega})$ and $L(\hat{\Omega})$. Finally we can show that

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left\{\frac{1}{1 + \frac{[nm/(n+m)](\bar{x}-\bar{y})^2}{\sum_1^n (x-\bar{x})^2 + \sum_1^m (y-\bar{y})^2}}\right\}^{\frac{m+n}{2}}$$

or

$$\lambda^{\frac{2}{n+m}} = \frac{1}{1 + \frac{[nm/(n+m)](\bar{x}-\bar{y})^2}{\sum_1^n (x-\bar{x})^2 + \sum_1^m (y-\bar{y})^2}}.$$

Since

$$T = \frac{\sqrt{\left(\frac{nm}{n+m}\right)}\ (\bar{x}-\bar{y})}{\sqrt{\frac{\Sigma(x-\bar{x})^2+\Sigma(y-\bar{y})^2}{n+m-2}}}$$

is distributed as t- distribution, then

$$\lambda^{\frac{2}{n+m}} = \frac{n+m-2}{n+m-2+T^2}$$

and the test will be based on the statistic T with $n+m-2$ degrees of freedom. We reject the null composite hypothesis H_0 if $\lambda \leqslant \lambda_\alpha$, ie. if $|T| \geqslant t_\alpha$, where t_α is obtainable from the statistical tables with corresponding probability α and $n+m-2$ degrees of freedom and we accept it otherwise. The probability α, the significance level of the test will be put in the form

$$\alpha = Pr\left\{\lambda \leqslant \lambda_\alpha : H_0\right\}$$

or

$$\alpha = Pr\left\{|T| \geqslant t_\alpha : H_0\right\}$$

In virtue of the two forms above we can say that the hypothesis H_0 will be rejected if and only if

$$Pr\left\{\lambda \leqslant \lambda_\alpha : H_0\right\} \leqslant \alpha$$

or

$$Pr\left\{|T| \geqslant t_\alpha : H_0\right\} \leqslant \alpha$$

and we accept it otherwise.

4. <u>The Test of the Equality of Several Means:</u>

Let $x_{1j}, x_{2j}, ---, x_{kj}$ be a random sample of size k drawn from the jth population whose distribution is normal with unknown mean

μ_j and variance σ^2, where $j=1,\ldots,\ell$ say. Here we have to test the null composite hypothesis $H_0: \mu_1=\mu_2=\cdots=\mu_\ell=\mu$ against all the alternative composite hypotheses. Now the whole space of the parameters and the subspace which specified by the hypothesis H_0 will be as follows

$$\Omega = \left\{ (\mu_1,\ldots,\mu_\ell,\sigma^2) : -\infty < \mu_j < \infty \,, \quad 0 < \sigma^2 < \infty \right\}$$

and

$$\omega = \left\{ (\mu_1,\ldots,\mu_\ell,\sigma^2) : -\infty < \mu_1 = \cdots = \mu_\ell = \mu < \infty \,, \quad 0 < \sigma^2 < \infty \right\}$$

Then

$$L(\Omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{\ell k}{2}} \exp\left\{ - \frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\mu_j)^2}{2\sigma^2} \right\}$$

$$L(\omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{\ell k}{2}} \exp\left\{ - \frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\mu)^2}{2\sigma^2} \right\}$$

By solving the equations

$$\frac{\partial \log L(\omega)}{\partial \mu} = 0, \quad \frac{\partial \log L(\omega)}{\partial \sigma^2} = 0, \quad \frac{\partial \log L(\Omega)}{\partial \sigma^2} = 0, \quad \frac{\partial \log L(\Omega)}{\partial \mu_j} = 0, \quad j=1,\ldots,\ell$$

we get the maximum likelihood estimates of these parameters. Substituting these estimates in $L(\omega)$ and $L(\Omega)$ we obtain $L(\hat{\omega})$ and $L(\hat{\Omega})$. Then we can show that

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left\{ \frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{.j})^2}{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{\bar{x}})^2} \right\}^{\frac{\ell k}{2}},$$

ie.

$$\lambda^{\frac{2}{\ell k}} = \frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j})^2}{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x})^2}$$

$$= \frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j})^2}{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j}+\bar{x}_{\cdot j}-\bar{x})^2}$$

$$= \frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j})^2}{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j})^2+\sum_{j=1}^{\ell}\sum_{i=1}^{k}(\bar{x}_{\cdot j}-\bar{x})^2}$$

$$= \frac{1}{1+\dfrac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(\bar{x}_{\cdot j}-\bar{x})^2}{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j})^2}}$$

Since

$$\frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(\bar{x}_{\cdot j}-\bar{x})^2/\sigma^2}{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j})^2/\sigma^2} = \frac{\chi^2_{[\ell-1]}}{\chi^2_{[\ell(k-1)]}}$$

then $F=\left\{\chi^2_{[\ell-1]}/(\ell-1)\right\}\Big/\left\{\chi^2_{[\ell(k-1)]}/\ell(k-1)\right\}$ is distributed as F- distribution with $\ell-1$ and $\ell(k-1)$ degrees of freedom, and so the test will be based on the statistic F. Here as λ decreased F increased therefore we reject the hypothesis H_0 if, and only if,

$$F \geqslant F_\alpha$$

where α is the significance level of the test. That is

$$\alpha = Pr\left\{\lambda \leqslant \lambda_\alpha : H_0\right\}$$

or

$$\alpha = Pr\left\{ F \geqslant F_\alpha : H_0 \right\}$$

and so we reject thehypothesis H_0 if, and only if,

$$Pr\left\{ \lambda \leqslant \lambda_\alpha : H_0 \right\} \leqslant \alpha$$

ie.

$$Pr\left\{ F \geqslant F_\alpha : H_0 \right\} \leqslant \alpha$$

We can find F_α from the statistical tables if α and the corresponding degrees of freedom of F are known.

4.1 <u>The case of the Effects of two factors on an outcome</u>

Let x_{ij} , $i = 1, ---, k$; $j = 1, ----, \ell$ be ℓk stochastically independent random variables having normal distribution with mean μ_{ij} and variance σ^2. If we put μ_{ij} in the form

$$\mu_{ij} = \mu + a_i + b_j$$

where

$$\sum_{i=1}^{k} a_i = 0 \quad \text{and} \quad \sum_{j=1}^{\ell} b_j = 0 ,$$

then

$$\mu_{i1} = \mu_{i2} = ---- = \mu_{i\ell} , \qquad i = 1, 2, ---, k$$

it means that

$$b_1 = b_2 = ---- = b_\ell = 0 \qquad \text{since} \quad \sum_{j=1}^{\ell} b_j = 0$$

and

$$\mu_{j1} = \mu_{j2} = ---- = \mu_{jk} , \qquad j = 1, 2, ---, \ell$$

it means $a_1 = a_2 = ---- = a_k = 0$ since $\sum_{i=1}^{k} a_i = 0$.

Therefore we can replace the null composite hypothesis $H_0 : \mu_{i1} = --- = \mu_{i\ell}$ by $H_0 : b_1 = b_2 = --- = b_\ell = 0$ in order to test it against all the alternative composite hypotheses. Here Ω and ω will be such that

$$\Omega = \left\{ (\mu, a_1, \ldots, a_k, b_1, \ldots, b_\ell, \sigma^2) : \begin{array}{ll} -\infty < \mu < \infty & \\ -\infty < a_i < \infty & \sum_{i=1}^{k} a_i = 0 \\ & \\ -\infty < b_j < \infty & \sum_{j=1}^{\ell} b_j = 0 \\ 0 < \sigma^2 < \infty & \end{array} \right\}$$

$$\omega = \left\{ (\mu, a_1, \ldots, a_k, b_1, \ldots, b_\ell, \sigma^2) : \begin{array}{ll} -\infty < \mu < \infty & \\ -\infty < a_i < \infty & \sum_{i=1}^{k} a_i = 0 \\ & \\ b_1 = b_2 = \cdots = b_\ell = 0 & \\ 0 < \sigma^2 < \infty & \end{array} \right\}$$

and so

$$L(\omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{\ell k}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \mu - a_i)^2 \right\}$$

$$L(\Omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{\ell k}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \mu - a_i - b_j)^2 \right\}$$

Now

$$\frac{\partial \log L(\omega)}{\partial \sigma^2} = -\frac{\ell k}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \mu - a_i)^2 ,$$

then $\dfrac{\partial \log L(\omega)}{\partial \sigma^2} = 0$ gives $\hat{\sigma}_\omega^2 = \dfrac{1}{\ell k} \sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \hat{\mu} - \hat{a}_i)^2$,

where $\hat{\mu}$ and $\hat{a}_i$ as will be shown are the maximum likelihood estimates of μ and a_i. Here

$$\frac{\partial \log L(\omega)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \mu) \quad \text{since} \quad \sum_{i=1}^{k} a_i = 0$$

then $\dfrac{\partial \log L(\omega)}{\partial \mu} = 0$ gives $\hat{\mu} = \dfrac{1}{\ell k} \sum_{j=1}^{\ell} \sum_{i=1}^{k} x_{ij}$

Now any a_i, $i = 1, ----, k$ can be written as $a_i = -(a_1 + -- + a_{i-1} + a_{i+1} + --- + a_k)$ and so we can therefore take the partial derivative of $\log L(\omega)$ with respect to a_i for $i = 1, ----, k-1$. Here we have

$$\log L(\omega) = \frac{-1}{2\sigma^2} \left\{ \sum_{j=1}^{\ell} \sum_{i=1}^{k-1} (x_{ij} - \mu - a_i)^2 + \sum_{j=1}^{\ell} (x_{kj} - \mu - a_k)^2 \right\} - \frac{\ell k}{2} \log 2\pi\sigma^2$$

then

$$\frac{\partial \log L(\omega)}{\partial a_i} = \frac{2}{2\sigma^2} \sum_{j=1}^{\ell} \sum_{i=1}^{k-1} (x_{ij} - \mu - a_i), \qquad i = 1, --, k-1$$

and so $\dfrac{\partial \log L(\omega)}{\partial a_i} = 0$ gives $\sum_{j=1}^{\ell} \sum_{i=1}^{k-1} (x_{ij} - \hat{\mu} - a_i) = 0$,

ie. $\sum_{j=1}^{\ell} \sum_{i=1}^{k-1} x_{ij} - \ell(k-1)\hat{\mu} - \ell \sum_{i=1}^{k-1} a_i = 0$,

ie. $-\sum_{j=1}^{\ell} x_{kj} + \ell\hat{\mu} + \ell a_k = 0$ then

$$\hat{a}_k = \frac{1}{\ell} \sum_{j=1}^{\ell} x_{kj} - \hat{\mu} \quad \text{and so} \quad \hat{a}_i = \frac{1}{\ell} \sum_{j=1}^{\ell} x_{ij} - \hat{\mu}$$

In the same way we can show that the maximum likelihood estimates of the parameters in Ω are

$$\hat{\mu} = \frac{1}{\ell k} \sum_{j=1}^{\ell} \sum_{i=1}^{k} x_{ij} , \qquad \hat{a}_i = \frac{1}{\ell} \sum_{j=1}^{\ell} x_{ij} - \hat{\mu} , \qquad \hat{b}_j = \frac{1}{k} \sum_{i=1}^{k} x_{ij} - \hat{\mu}$$

and $$\hat{\sigma}_{\Omega}^{2} = \frac{1}{\ell k} \sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \hat{\mu} - \hat{a}_i - \hat{b}_j)^2 .$$

The likelihood ratio is then

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left[\frac{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \hat{\mu} - \hat{a}_i - \hat{b}_j)^2}{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \hat{\mu} - \hat{a}_i)^2} \right]^{\frac{\ell k}{2}}$$

ie.

$$\lambda^{\frac{2}{\ell k}} = \frac{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \hat{\mu} - \hat{a}_i - \hat{b}_j)^2}{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \hat{\mu} - \hat{a}_i)^2} = \frac{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} + \bar{x} - \bar{x}_{i.} - \bar{x}_{.j})^2}{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} - \bar{x} - \bar{x}_{i.} + \bar{x})^2}$$

where $\hat{\mu} = \bar{x}$, $\hat{a}_i = \bar{x}_{i.} - \bar{x}$, $\hat{b}_j = \bar{x}_{.j} - \bar{x}$. Then by using the analysis of variance $\lambda^{\frac{2}{\ell k}}$ could be written such that

$$\lambda^{\frac{2}{\ell k}} = \frac{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} + \bar{x} - \bar{x}_{i.} - \bar{x}_{.j})^2}{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} + \bar{x} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{.j} - \bar{x})^2}$$

$$= \frac{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} + \bar{x} - \bar{x}_{i.} - \bar{x}_{.j})^2}{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} + \bar{x} - \bar{x}_{i.} - \bar{x}_{.j})^2 + \sum_{j=1}^{\ell} \sum_{i=1}^{k} (\bar{x}_{.j} - \bar{x})^2}$$

$$= \frac{1}{1 + \frac{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (\bar{x}_{.j} - \bar{x})^2}{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} + \bar{x} - \bar{x}_{i.} - \bar{x}_{.j})^2}} .$$

Since

$$\frac{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (\bar{x}_{.j} - \bar{x})^2 / \sigma^2}{\sum_{j=1}^{\ell} \sum_{i=1}^{k} (x_{ij} + \bar{x} - \bar{x}_{i.} - \bar{x}_{.j})^2 / \sigma^2} = \frac{\chi^2_{[\ell-1]}}{\chi^2_{[(\ell-1)(k-1)]}} ,$$

then

$$F = \frac{\chi^2_{[\ell-1]}/(\ell-1)}{\chi^2_{[(\ell-1)(k-1)]}/(\ell-1)(k-1)}$$ is distributed as

F- distribution with $\ell-1$ and $(\ell-1)(k-1)$ degrees of freedom, and so the test will be based on the statistic F. Since F increased as λ decreases, we reject the hypothesis H_0 if, and only if,

$$F \geqslant F_\alpha$$

where α is the significance level of the test such that

$$\alpha = Pr\{F \geqslant F_\alpha : H_0\}$$

or

$$Pr\{F \geqslant F_\alpha : H_0\} \leqslant \alpha$$

4.2 <u>In Case When the Variance is Known:</u>

Let $x_{1j}, x_{2j}, \dots\dots, x_{kj}$ be a random sample drawn from jth population whose distribution is normal with unknown mean μ_j and known variance σ_0^2, where $j=1,2,\dots,\ell$, say. Here we have to test the null composite hypothesis $H_0: \mu_1=\mu_2=\dots=\mu_\ell=\mu$ against all the alternative hypotheses. Then the whole space of the parameters and the subspace which is specified by the hypothesis H_0 are as follows

$$\Omega = \left\{(\mu_1, \dots, \mu_\ell, \sigma_0^2) : -\infty < \mu_j < \infty, \quad 0 < \sigma_0^2 < \infty\right\}$$

and

$$\omega = \left\{(\mu_1, \dots, \mu_\ell, \sigma_0^2) : -\infty < \mu_1 = \dots = \mu_\ell = \mu < \infty, \quad 0 < \sigma_0^2 < \infty\right\}$$

and so $L(\Omega) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{\ell k}{2}} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_1^{\ell}\sum_1^{k}(x_{ij}-\mu_j)^2\right\}$

and

$$L(\omega) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{\ell k}{2}} \exp\left\{-\frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\mu)^2}{2\sigma_0^2}\right\}$$

By solving the equations

$$\frac{\partial \log L(\Omega)}{\partial \mu_j} = 0 \quad \text{and} \quad \frac{\partial \log L(\omega)}{\partial \mu} = 0 \ , \qquad j = 1, \ldots, \ell$$

we get the maximum likelihood estimates of the required parameters. Substituting these estimates in $L(\Omega)$ and $L(\omega)$ give us

$$L(\hat{\Omega}) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{\ell k}{2}} \exp\left\{-\frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j})^2}{2\sigma_0^2}\right\}$$

$$L(\hat{\omega}) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{\ell k}{2}} \exp\left\{-\frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x})^2}{2\sigma_0^2}\right\}$$

The likelihood ratio test is then

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \exp\left\{\frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x}_{\cdot j})^2 - \sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij}-\bar{x})^2}{2\sigma_0^2}\right\}$$

$$= \exp\left\{-\frac{k\sum_{j=1}^{\ell}(\bar{x}_{\cdot j}-\bar{x})^2}{2\sigma_0^2}\right\}$$

$$= \exp\left\{-\frac{\sum_{j=1}^{\ell}(\bar{x}_{\cdot j}-\bar{x})^2}{2\sigma_0^2/k}\right\}$$

then

$$-2\log\lambda = \sum_{j=1}^{\ell}(\bar{x}_{\cdot j}-\bar{x})^2 \Big/ \frac{\sigma_0^2}{k}$$

Since $\sum_{j=1}^{\ell}(\bar{x}_j - \bar{x})^2 \Big/ \frac{\sigma_0^2}{k}$ is distributed as χ^2- distribution with $\ell-1$ degrees of freedom, then $-2\log\lambda$ is distributed as $\chi^2_{[\ell-1]}$. And so we reject the hypothesis H_0 if, and only if,

$$\chi^2 \geqslant \chi^2_\alpha$$

where α is the significance level of the test.

Example 4.3

Let $x_{1j}, x_{2j}, \ldots, x_{8j}$, $j=1,2$ be a random sample from the jth population has a distribution defined by $N(x; \theta_j, \sigma^2)$ and let $\bar{x}_{\cdot 1} = 75.2$, $\bar{x}_{\cdot 2} = 78.6$, $\sum_1^8 (x_{i1} - \bar{x}_{\cdot 1})^2 = 71.2$ and $\sum_1^8 (x_{i2} - \bar{x}_{\cdot 2})^2 = 54.8$. Test the null composite hypothesis $H_0: \theta_1 = \theta_2$ against the alternative composite hypothesis $H_1: \theta_1 \neq \theta_2$ at 5% level of significance.

We have from section 3 that

$$T = \frac{\sqrt{\frac{nm}{n+m}}\,(\bar{x}_{\cdot 1} - \bar{x}_{\cdot 2})}{\sqrt{\left(\frac{\sum_1^n (x_{i1} - \bar{x}_{\cdot 1})^2 + \sum_1^n (x_{i2} - \bar{x}_{\cdot 2})^2}{n+m-2}\right)}}$$

is distributed as t-distribution with $n+m-2$ degrees of freedom. By substituting the observed values we get

$$T = \frac{\sqrt{\frac{64}{16}}\,(75.2 - 78.6)}{\sqrt{\left(\frac{71.2 + 54.8}{14}\right)}} = -2.27$$

We have $t_{0.05} = 2.14$ for 14 degrees of freedom (Statistical Table). Since

$$(|T| = 2.27) > (t_{0.05} = 2.14)$$

we reject the hypothesis H_0 on 5% level of significance.

<u>Example 4.4</u>

Let μ_1, μ_2, μ_3 be respectively the means of three independent normal distributions having common but unknown variance σ^2. Test the null composite hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \mu$ against all possible alternative hypotheses at the 5% level of significance. The following table shows the observed values of three samples of size 5 obtained from three populations.

Sample					
(1)	3	0	-1	0	2
(2)	2	5	1	3	5
(3)	4	3	6	8	5

$$\bar{x}_{.1} = \frac{1}{5}\sum_{1}^{5} x_{i1} = \frac{4}{3} = 0.8$$

$$\bar{x}_{.2} = \frac{1}{5}\sum_{1}^{5} x_{i2} = \frac{16}{5} = 3.2$$

$$\bar{x}_{.3} = \frac{1}{5}\sum_{1}^{5} x_{i3} = \frac{26}{5} = 5.2$$

$$\bar{x} = \frac{1}{15}\sum_{1}^{3}\sum_{1}^{5} x_{ij} = \frac{46}{15} = 3.07$$

We have from section 4 that

$$F = \frac{\chi^2_{[\ell-1]} / \ell-1}{\chi^2_{[\ell(k-1)]} / \ell(k-1)} = \frac{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(\bar{x}_{.j} - \bar{x})^2 / \ell-1}{\sum_{j=1}^{\ell}\sum_{i=1}^{k}(x_{ij} - \bar{x}_{.j})^2 / \ell(k-1)}$$

is distributed as F-distribution with $\ell-1, \ell(k-1)$ degrees of freedom. By substituting the observed values we get

$$\frac{\sum_1^3 \sum_1^5 (\bar{x}_{.j} - \bar{x})^2}{\ell - 1} = \frac{5}{2} \sum_1^3 (\bar{x}_{.j} - \bar{x})^2 = \frac{48.5335}{2} = 24.26675$$

$$\frac{\sum_1^3 \sum_1^5 (x_{ij} - \bar{x}_{.j})^2}{\ell(k-1)} = \frac{1}{12} \sum_1^3 \sum_1^5 (x_{ij} - \bar{x}_{.j})^2 = \frac{192}{12 \times 5} = \frac{192}{60}$$

Then

$$F = \frac{24.26675 \times 60}{192} = 7.6$$

We have $F_{0.05} = 3.89$ for 2 and 12 degrees of freedom (Statistical Table). Since

$$(F = 7.6) > (F_{0.05} = 3.89)$$

we reject the hypothesis H_0 on 5% level of significance.

<u>Example 4.5</u>

If S_1, S_2, S_3 are three samples of size 4 from three populations having normal distribution with mean $\mu_{ij} = \mu + a_i + b_j$, $\sum_1^4 a_i = 0$, $\sum_1^3 b_j = 0$ and common but unknown variance σ^2. Test the null composite hypothesis $H_0: b_1 = b_2 = b_3 = b$ against all possible alternative hypotheses. The following table shows the observed values

Sample				
(1)	3	-1	0	6
(2)	5	2	2	6
(3)	7	2	5	10

$$\bar{x}_{.1} = \frac{1}{4}\sum_{1}^{4} x_{i1} = \frac{8}{4} = 2$$

$$\bar{x}_{.2} = \frac{1}{4}\sum_{1}^{4} x_{i2} = \frac{15}{4} = 3.75$$

$$\bar{x}_{.3} = \frac{1}{4}\sum_{1}^{4} x_{i3} = \frac{24}{4} = 6$$

$$\bar{x} = \frac{1}{12}\sum_{1}^{3}\sum_{1}^{4} x_{ij} = \frac{47}{12} = 3.9$$

$$\bar{x}_{1.} = \frac{1}{3}\sum_{1}^{3} x_{1j} = \frac{15}{3} = 5$$

$$\bar{x}_{2.} = \frac{1}{3}\sum_{1}^{3} x_{2j} = \frac{3}{3} = 1$$

$$\bar{x}_{3.} = \frac{1}{3}\sum_{1}^{3} x_{3j} = \frac{7}{3} = 2.3$$

$$\bar{x}_{4.} = \frac{1}{3}\sum_{1}^{3} x_{4j} = \frac{22}{3} = 7.3$$

We have from 41 that

$$\frac{(4-1)\sum_{1}^{3}\sum_{1}^{4}(\bar{x}_{.j}-\bar{x})^2}{\sum_{1}^{3}\sum_{1}^{4}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x})^2}$$

is distributed as F-distribution with (3-1) and (3-1)(4-1) degrees of freedom. Then by substituting the observed values we get

$$\sum_{1}^{3}\sum_{1}^{4}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x})^2 = \sum_{1}^{3}\sum_{1}^{4} x_{ij}^2 + 12\bar{x}^2 - 3\sum_{1}^{4}\bar{x}_{i.}^2 - 4\sum_{1}^{3}\bar{x}_{.j}^2$$

$$= 293 + 182.52 - (253.74 + 216.25) = 5.53$$

$$(4-1)\sum_{1}^{3}\sum_{1}^{4}(\bar{x}_{.j}-\bar{x})^2 = 3(216.25 - 182.52) = 101.19$$

Therefore

$$F = \frac{101.19}{5.53} = 18.3$$

We have $F_{0.05} = 5.14$ for 2 and 6 degrees of freedom (Statistical Table). Since

$$(F = 18.3) > (F_{0.05} = 5.14)$$

we reject the hypothesis H_0 on 5% level of significance.

5. A Test of Significance of the Correlation Coefficient:

If x and y have a divariate normal distribution with means μ_1 and μ_2, variances σ_1^2 and σ_2^2 and correlation coefficient ρ. Here the null composite hypothesis will be $H_0: \rho = 0$, ie. x and y are independent, and the alternative composite hypothesis will be $H_1: \rho \neq 0$, ie. x and y are dependent. The space of the whole parameters and the subspace specified by H_0 are then

$$\Omega = \left\{ (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) : -\infty < \mu_1, \mu_2 < \infty, \quad 0 < \sigma_1^2, \sigma_2^2 < \infty, \ -1 < \rho < 1 \right\}$$

$$\omega = \left\{ (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) : -\infty < \mu_1, \mu_2 < \infty, \quad 0 < \sigma_1^2, \sigma_2^2 < \infty, \ \rho = 0 \right\}$$

This problem has been discussed in details in (15), therefore it is worth while to put this discussion in Appendix III and mention here the statistic on which the likelihood ratio test is based. The author has determined the probability density function of the statistic R, the correlation coefficient of the random sample (x_i, y_i), $i = 1, 2, ---, n$ when $\rho = 0$ and $n > 2$. The form of this probability density function is

$$g(r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} (1 - r^2)^{\frac{n-4}{2}} \quad ,$$

where $-1 < r < 1$ is the observed value of R. If the significance level of the test is α, then

$$\frac{\alpha}{2} = \int_c^1 g(r)\, dr \qquad 0 < c < 1$$

If α and n are known, then c will be determined and so we reject the hypothesis $H_0: \rho = 0$, if, and only if,

$$|r| \geq c$$

and we accept it otherwise.

<u>Example 4.6</u>

A random sample of size $n = 6$ from a beveriate normal distribution yields the value of the correlation coefficient to be 0.89. Would we accept or reject, at the 5% significance level, the null hypothesis that $\rho = 0$?

We have

$$\frac{\alpha}{2} = \int_c^1 g(r)\,dr$$

and

$$g(r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} (1-r^2)^{\frac{n-4}{2}}$$

Since $n = 6$ [illegible]

$$\frac{0.05}{2} = \int_c^1 \frac{\Gamma\left(\frac{6-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{6-2}{2}\right)} (1-r^2)^{\frac{6-4}{2}}\,dr$$

$$= \frac{3}{4}\int_c^1 (1-r^2)\,dr = \frac{3}{4}\left\{\frac{2}{3} - c + \frac{1}{3}c^3\right\}$$

i.e. $\quad c^3 - 3c + 1.9 = 0$

By solving this equation we obtain $C = 0.8$[illegible]. Here $|r| = 0.89 > 0.81$, thus we reject the null hypothesis $H_0: \rho = 0$ on 5% level of significance.

6. A Test of Equality of Variances of Two Populations:

Let $x_{1j}, x_{2j}, \text{----}, x_{kj}$, $j=1,2$ be a random sample drawn from population whose distribution is normal with mean μ_j and variance σ_j^2. We have to test the null composite hypothesis $H_0: \sigma_1^2=\sigma_2^2=\sigma^2$ against the alternative composite hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$. Here the space Ω of the whole parameters and the subspace ω which is specified by H_0 are as follows

$$\Omega = \left\{ (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : -\infty < \mu_1, \mu_2 < \infty, \quad 0 < \sigma_1^2, \sigma_2^2 < \infty \right\}$$

$$\omega = \left\{ (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : -\infty < \mu_1, \mu_2 < \infty, \quad 0 < \sigma_1^2 = \sigma_2^2 < \infty \right\}$$

Then

$$L(\Omega) = \frac{1}{(2\pi\sigma_1^2)^{\frac{k}{2}} (2\pi\sigma_2^2)^{\frac{k}{2}}} \exp\left\{ -\sum_1^k \sum_1^2 \frac{(x_{ij}-\mu_j)^2}{2\sigma_j^2} \right\},$$

and

$$L(\omega) = \frac{1}{(2\pi\sigma^2)^k} \exp\left\{ -\frac{1}{2\sigma^2} \sum_1^k \sum_1^2 (x_{ij}-\mu_j)^2 \right\}$$

Solving the equations

$$\frac{\partial \log L(\Omega)}{\partial \mu_j} = 0, \quad \frac{\partial \log L(\Omega)}{\partial \sigma_j^2} = 0, \quad \frac{\partial \log L(\omega)}{\partial \sigma^2} = 0, \quad \frac{\partial \log L(\omega)}{\partial \mu_j} = 0,$$

$j=1,2$, we get the maximum likelihood estimates of these parameters. Then substituting these estimates in $L(\Omega)$ and $L(\omega)$ we obtain $L(\hat{\Omega})$ and $L(\hat{\omega})$. The statistic λ is then

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{(\hat{\sigma}_1^2 \hat{\sigma}_2^2)^{\frac{k}{2}}}{(\hat{\sigma}^2)^k} = \frac{\left[\frac{1}{k}\sum_1^k (x_{i1}-\bar{x}_{.1})^2 \ \frac{1}{k}\sum_1^k (x_{i2}-\bar{x}_{.2})^2 \right]^{\frac{k}{2}}}{\left[\frac{\sum_1^k (x_{i1}-\bar{x}_{.1})^2 + \sum_1^k (x_{i2}-\bar{x}_{.2})^2}{2k} \right]^k}$$

ie.

$$\lambda = 2^k \left[\frac{\sum_1^k (x_{i2} - \bar{x}_{.2})^2}{\sum_1^k (x_{i1} - \bar{x}_{.1})^2} \right]^{\frac{k}{2}} \Big/ \left[1 + \frac{\sum_1^k (x_{i2} - \bar{x}_{.2})^2}{\sum_1^k (x_{i1} - \bar{x}_{.1})^2} \right]^k$$

ie.

$$\lambda = 2^k \left[\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \right]^{\frac{k}{2}} \Big/ \left[1 + \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \right]^k$$

Consider

$$f(z) = \frac{z^{\frac{k}{2}}}{(1+z)^k}, \qquad \text{either } z > 1 \text{ or } 0 < z < 1$$

then

$$\frac{\partial \log f(z)}{\partial z} = \frac{k}{2z(1+z)}(1-z)$$

$$\begin{aligned} &= 0 \quad &&\text{if } z = 1 \\ &< 0 \quad &&\text{if } z > 1 \\ &> 0 \quad &&\text{if } 0 < z < 1 \end{aligned}$$

That is means that λ decreases when z increases and λ decreases when z decreases. Since $\left[\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} \right]$ is distributed as F-distribution with $[k-1, k-1]$ degrees of freedom, then the likelihood ratio test may be based on the statistic F. If α is the significance level of the test, then we accept the null composite hypothesis $H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$, if, and only if,

$$F_{1-\frac{\alpha}{2}} < F < F_{\frac{\alpha}{2}}$$

and we reject it otherwise. The alternative hypothesis in this case is called "two-sided". In the case when the alternative hypothesis is "one-sided" the critical region will be as follows:

When $\hat{\sigma}_2^2 > \hat{\sigma}_1^2$ then $\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} > 1$ and so we reject the null composite hypothesis $H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$ if, and only if,

$$F \geqslant F_\alpha \, .$$

When $\hat{\sigma}_2^2 < \hat{\sigma}_1^2$ then $0 < \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2} < 1$ and so we reject the hypothesis H_0 if, and only if,

$$F \leqslant F_{1-\alpha}$$

The test will be applied as well, when the sizes of the two samples are different. If m and n are the sizes of the two samples then F will be distributed as F-distribution with $[m-1, n-1]$ degrees of freedom.

Here $F_{1-\alpha}$ represents the lower percentage point. We can find this point from the statistical Tables by interchanging the degreed of freedom $(m-1)$ and $(n-1)$ and taking the reciprocal of the tabulated value.

7. <u>A Test of the Equality of the Variances of K Populations:</u>

Let $x_{i1}, x_{i2}, -----, x_{in_i}$ be a random sample drawn from the ith population whose distribution is normal with unknown mean μ_i and variance σ_i^2, $i = 1, ----, k$. Now we have to test the null hypothesis $H_0: \sigma_1^2 = \sigma_2^2 = --- = \sigma_k^2 = \sigma^2$ against all the possible alternative hypotheses. Here

$$\Omega = \left\{ (\mu_1, ---, \mu_k, \sigma_1^2, -----, \sigma_k^2) : -\infty < \mu_i < \infty , \quad 0 < \sigma_i^2 < \infty \right\}$$

and

$$\omega = \left\{ (\mu_1, ---, \mu_k, \sigma_1^2, ----, \sigma_k^2) : -\infty < \mu_i < \infty , \quad 0 < \sigma_i^2 = \sigma^2 < \infty \right\}$$

Then

$$L(\Omega) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1^{n_1} \cdots \sigma_k^{n_k}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(x_{ij} - \mu_i)^2}{\sigma_i^2} \right\}$$

and

$$L(\omega) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2 \right\}$$

where $n = \sum_{1}^{k} n_i$. By solving the equations

$$\frac{\partial \log L(\Omega)}{\partial \mu_i} = 0, \quad \frac{\partial \log L(\Omega)}{\partial \sigma_i^2} = 0, \quad \frac{\partial \log L(\omega)}{\partial \mu_i} = 0 \quad \text{and} \quad \frac{\partial \log L(\omega)}{\partial \sigma^2} = 0,$$

we get the maximum likelihood estimates of the required parameters. Then by substituting these estimates in $L(\omega)$ and $L(\Omega)$ we obtain

$$L(\hat{\omega}) = \left(\frac{1}{2\pi\hat{\sigma}^2} \right)^{\frac{n}{2}} \exp\left\{ -\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2 \right\}$$

and

$$L(\hat{\Omega}) = \frac{1}{(2\pi)^{\frac{n}{2}} \hat{\sigma}_1^{n_1} \cdots \hat{\sigma}_k^{n_k}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{(x_{ij} - \bar{x}_{i.})^2}{\hat{\sigma}_i^2} \right\},$$

where $\hat{\sigma}^2, \bar{x}_{i.}, \hat{\sigma}_i^2$ are the maximum likelihood estimates of σ^2, μ_i and σ_i^2 respectively, then the likelihood ratio is

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{\hat{\sigma}_1^{n_1} \hat{\sigma}_2^{n_2} \cdots \hat{\sigma}_k^{n_k}}{\hat{\sigma}^n}$$

then

$$\log \lambda = \sum_{i=1}^{k} n_i \log \hat{\sigma}_i - n \log \hat{\sigma}$$

$$= \frac{1}{2} \left\{ \sum_{i=1}^{k} n_i \log \hat{\sigma}_i^2 - n \log \hat{\sigma}^2 \right\}$$

$$\therefore \quad -2 \log \lambda = n \log \hat{\sigma}^2 - \sum_{i=1}^{k} n_i \log \hat{\sigma}_i^2 .$$

Now we use the modified test by Bartlett which is defined by the statistic

$$T = \frac{(n-k)\log \sigma'^2 - \sum_1^k (n_i-1)\log \sigma_i'^2}{1+\frac{1}{3(k-1)}\left[\sum_1^k \left(\frac{1}{n_i-1}\right) - \frac{1}{n-k}\right]},$$

where

$$\sigma'^2 = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2, \qquad \sigma_i'^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{i.})^2$$

are unbiased estimates of σ^2 and σ_i^2 respectively. Here the statistic T is distributed as χ^2-distribution with $k-1$ degrees of freedom. Therefore we reject the hypothesis H_0 if, and only if,

$$(T = \chi^2) \geqslant \chi^2_\alpha .$$

Example 4.7

In sampling from two normal distributions the following observed values were obtained from samples of size 25: $S_1^2 = 1.25$, $S_2^2 = 1.97$. Test at the 5% level for equality of variances.

Here the null hypothesis will be such that $H_0: \sigma_1^2 = \sigma_2^2$ and the alternative hypothesis $H_1: \sigma_1^2 \neq \sigma_2^2$. We have from section 6 that

$$\frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}$$

is distributed as F-distribution with (K-1) and (K-1) degrees of freedom, where K is the sample size. Then

$$F = \frac{S_2^2}{S_1^2} = \frac{1.97}{1.25} = 1.576$$

is distributed as F-distribution with 24 and 24 degrees of freedom. We have from Statistical Tables that $F_{\frac{0.05}{2}} = 2.27$

with 24 and 24 degrees of freedom and $F_{1-\frac{0.05}{2}} = \frac{1}{2.27} = 0.44$ with 24 and 24 degrees of freedom. Since

$$\left(F_{1-\frac{0.05}{2}} = 0.44\right) < \left(F = 1.576\right) < \left(F_{\frac{0.05}{2}} = 2.27\right)$$

therefore we accept the hypothesis H_0.

Example 4.8

Given the following 5 sample variances based on 10 observations each, test the hypothesis that the 5 population variances are equal. The sample variances are 22, 40, 30, 32, 12.

Here the null hypothesis will be such that $H_0: \sigma_1^2 = --- = \sigma_5^2 = \sigma^2$. We have from section 7, that

$$T = \left[(n-k)\log \sigma'^2 - \sum_1^k (n_i-1)\log \sigma_i'^2\right] \Big/ 1 + \frac{1}{3(k-1)}\left[\sum_1^k\left(\frac{1}{n_i-1}\right) - \frac{1}{n-k}\right]$$

is distributed as χ^2 - distribution with $k-1$ degrees of freedom, where

$$\sigma'^2 = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2, \qquad \sigma_i'^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i.})^2.$$

Here

$$\sigma'^2 = \frac{n_1\hat{\sigma}_1^2 + --- + n_5\hat{\sigma}_5^2}{n-k} = \frac{10(22+40+30+32+12)}{50-5} = \frac{272}{9}$$

$$\sigma_i'^2 = \frac{n_i}{n_i-1}\hat{\sigma}_i^2 \quad , \quad \text{then}$$

$$\sum_1^5 (n_i-1)\log \sigma_i'^2 = 9\left\{\log\frac{10}{9}\times 22 + \log\frac{10}{9}\times 40 + ----+ \log\frac{10}{9}\times 12\right\}$$

$$= 9\left\{5(\log 10 - \log 9) + \log 22 + \log 40 + -- + \log 12\right\}$$

$$= 65.1141$$

$$(n-k)\log \hat{\sigma}^2 = 45 \log \frac{272}{9} = 45(\log 272 - \log 9)$$

$$= 45(2.4346 - 0.9542) = 66.6180$$

Therefore

$$T = \frac{66.6180 - 65.1141}{1 + \frac{1}{12}\left[\frac{5}{9} - \frac{1}{45}\right]} = 1.44$$

We have from the Statistical Tables that $\chi^2_{0.05} = 9.49$ with 4 degrees of freedom. Since

$$(T = 1.44) < (\chi^2_{0.05} = 9.49)$$

therefore we accept the hypothesis H_0.

APPENDIX I

The Sampling Variance of Statistics

If $x_1, x_2, ---, x_n$ are n random variables, then the mean value and the variance of x will be defined by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

respectively.

Now if x is expressed in a linear function such that

$$x = n_1 l_1 + l_2 n_2 + ------ + l_r n_r \ ,$$

where $n_i, i = 1, 2, ---, r$ denotes the observed frequency in the ith class, then the mean value of x will become

$$n\sum_{i=1}^{r}(\theta_i l_i)$$

where θ_i (a linear function of x) is the probability corresponding to the class i. Hence the variance of x will be given by

$$V(x) = n\left\{\sum_i(\theta_i l_i^2) - \left[\sum_i(\theta_i l_i)\right]^2\right\} \quad ---------(A)$$

For any function of the observed frequencies by which the statistic is defined, there is a general formula which affords a variance very near to the sampling variance of the statistic. The formula is

$$V(x) = n\sum_i\left\{\theta_i\left(\frac{\partial x}{\partial n_i}\right)^2\right\} - n\left(\frac{\partial x}{\partial n}\right)^2 \quad --------(B)$$

where θ_i and n_i are as defined above. Now we are interested in three forms of functions by which x is defined.

(a) Let x be defined by

$$nx = n_1 - n_2 - n_3 + n_4$$

then the variance of x will be given by either formula (A) or formula (B), but we are going to apply formula (B) which is the general one. We have

$$\sum_{i=1}^{4}\left\{\theta_i\left(\frac{\partial x}{\partial n_i}\right)^2\right\} = \frac{1}{4n^2}\left(2+x+1-x+1-x+x\right) = \frac{1}{n^2},$$

and

$$\left(\frac{\partial x}{\partial n}\right)^2 = \left(-\frac{n_1-n_2-n_3+n_4}{n^2}\right)^2 = \frac{x^2}{n^2},$$

where

$$\theta_1 = \frac{1}{4}(2+x), \qquad \theta_2=\theta_3=\frac{1}{4}(1-x), \qquad \theta_4=\frac{1}{4}x .$$

Then

$$V(x) = n\sum_{i=1}^{4}\left\{\theta_i\left(\frac{\partial x}{\partial n_i}\right)^2\right\} - n\left(\frac{\partial x}{\partial n}\right)^2 = \frac{1-x^2}{n}$$

(b) Let x be defined by

$$4nx = 2n_1 - 2n_2 - 2n_3 + 10n_4$$

then

$$\sum_{i=1}^{4}\left\{\theta_i\left(\frac{\partial x}{\partial n_i}\right)^2\right\} = \frac{1}{16n^2}\left(2+x+1-x+1-x+25x\right) = \frac{1+6x}{4n^2}$$

and

$$\left(\frac{\partial x}{\partial n}\right)^2 = \left[\frac{1}{2n^2}\left(n_1-n_2-n_3+5n_4\right)\right]^2 = \frac{x^2}{n^2}$$

Then

$$V(x) = n\sum_{i=1}^{4}\left\{\theta_i\left(\frac{\partial x}{\partial n_i}\right)^2\right\} - n\left(\frac{\partial x}{\partial n}\right)^2 = \frac{1+6x-4x^2}{4n} .$$

(c) Let x be in the form

$$\frac{n_1 n_4}{n_2 n_3} = \frac{x(2+x)}{(1-x)^2} ,$$

then

$$\log n_1 + \log n_4 - \log n_2 - \log 3 = \log x + \log(2+x) - 2\log(1-x)$$

Differentiating with respect to n_i, $i = 1, ---, 4$ we get

$$\frac{\partial x}{\partial n_1} = \frac{1}{n_1} \frac{x(1-x)(2+x)}{2(1+2x)} ,$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial x}{\partial n_4} = \frac{1}{n_4} \frac{x(1-x)(2+x)}{2(1+2x)}$$

then

$$\sum_{i=1}^{4} \left\{ \theta_i \left(\frac{\partial x}{\partial n_i} \right)^2 \right\} = \left\{ \frac{x(1-x)(2+x)}{2(1+2x)} \right\}^2 \frac{1}{4} \left\{ \frac{2+x}{n_1^2} + \frac{1-x}{n_2^2} + \frac{1-x}{n_3^2} + \frac{x}{n_4^2} \right\}$$

$$= \left\{ \frac{x(1-x)(2+x)}{2(1+2x)} \right\}^2 \frac{1}{4n^2} \left\{ \frac{2+x}{\left(\frac{n_1}{n}\right)^2} + \frac{1-x}{\left(\frac{n_2}{n}\right)^2} + \frac{1-x}{\left(\frac{n_3}{n}\right)^2} + \frac{x}{\left(\frac{n_4}{n}\right)^2} \right\}$$

By replacing $\frac{n_i}{n}$ by θ_i we get

$$\sum_{i=1}^{4} \left\{ \theta_i \left(\frac{\partial x}{\partial n_i} \right)^2 \right\} = \left\{ \frac{x(1-x)(2+x)}{2(1+2x)} \right\}^2 \frac{16}{4n^2} \left\{ \frac{1}{2+x} + \frac{2}{1-x} + \frac{1}{x} \right\}$$

Hence

$$n \sum_{i=1}^{4} \left\{ \theta_i \left(\frac{\partial x}{\partial n_i} \right)^2 \right\} = \frac{2x(1-x)(2+x)}{n(1+2x)} .$$

Since the formula

$$\frac{n_1 n_4}{n_2 n_3} = \frac{x(2+x)}{(1-x)^2}$$

does not involve the number n, therefore

$$\frac{\partial x}{\partial n} = 0$$

and so

$$V(x) = n \sum_{i=1}^{4} \left\{ \theta_i \left(\frac{\partial x}{\partial n_i} \right)^2 \right\} - n \left(\frac{\partial x}{\partial n} \right)^2$$

$$= \frac{2x(1-x)(2+x)}{n(1+2x)} .$$

If y is such that

$$x = y^2 \quad \text{or} \quad x = 1 - y^2$$

then the variance of y will be as follows; the variance of the statistic which satisfies the maximum likelihood will be given by

$$V(x) = 1 \Big/ n \sum_i \left\{ \frac{1}{\theta_i} \left(\frac{\partial \theta_i}{\partial x} \right)^2 \right\}$$

where θ_i is as defined before. Then the variance of y will be given by

$$V(y) = 1 \Big/ n \sum_i \left\{ \frac{1}{\theta_i} \left(\frac{\partial \theta_i}{\partial y} \right)^2 \right\} .$$

Now

$$\frac{\partial \theta_i}{\partial y} = \frac{\partial \theta_i}{\partial x} \frac{\partial x}{\partial y} = 2y \frac{\partial \theta_i}{\partial x}$$

and

$$\left(\frac{\partial \theta_i}{\partial y} \right)^2 = \left(2y \frac{\partial \theta_i}{\partial x} \right)^2 = 4x \left(\frac{\partial \theta_i}{\partial x} \right)^2$$

Hence

$$V(y) = 1 \Big/ 4xn \sum_i \left\{ \frac{1}{\theta_i} \left(\frac{\partial \theta_i}{\partial x} \right)^2 \right\}$$

i.e.

$$V(y) = \frac{1}{4x} V(x) .$$

APPENDIX II

THEOREM: (For Large Samples)

If $f(x;\theta_1,--,\theta_m)$ is the probability density function of a population, and the maximum likelihood estimates of θ_i exist with a known distribution function, then the distribution of $-2\log\lambda$ is, except the terms of order $\frac{1}{\sqrt{n}}$, distributed as χ^2 with $m-r$ degrees of freedom, where λ is the likelihood retio and $m-r$ is the number of the parameters which specify the null hypothesis.

Proof:

Let $x_1,-----,x_n$ be a random sample drawn from a population which has a distribution function $f(x;\theta_1,--,\theta_m)$.

Then the likelihood function is

$$L(x;\theta) = \prod_1^n f(x;\theta_1,---,\theta_m).$$

Let the null composite hypothesis $H_0: \theta_i=\theta_{0i}, i=r+1,--,m$ be tested against all the possible alternative composite hypotheses and let Ω be the whole space of the m parameters and ω be the subspace specified by H_0. Then $L(\Omega)$ and $L(\omega)$ will be the likelihood functions designated by Ω and ω respectively. The likelihood ratio test will be defined by

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

where $L(\hat{\omega})$ and $L(\hat{\Omega})$ are the maximum of $L(\omega)$ and $L(\Omega)$ respectively. To find the approximation to the distribution of λ we have to assume that the maximum likelihood estimates of θ_i, $\hat{\theta}_i$, say, exist, and so their distribution will be such that

$$L(\Omega) = \frac{|a_{ij}|^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}} \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{m} a_{ij}y_iy_j\right\}(1+\psi)$$

where $\|a_{ij}\|$ is positive definite, $a_{ij} = -E\left(\frac{\partial^2 \log f}{\partial\theta_i \partial\theta_j}\right)$, $y_i = (\check{\theta}_i - \theta_i)\sqrt{n}$ and ψ is of order $\frac{1}{\sqrt{n}}$. By taking the logarithm of $L(\Omega)$ and differentiating with respect to θ_ℓ, $\ell = 1, 2, \text{------}, m$, we get

$$\frac{\partial \log L(\Omega)}{\partial\theta_\ell} = \frac{1}{2}\left[\frac{1}{|a_{ij}|}\frac{\partial|a_{ij}|}{\partial\theta_\ell} - \sum_{i,j}\frac{\partial a_{ij}}{\partial\theta_\ell} y_i y_j + \frac{1}{2}\sqrt{n}\sum_j a_{\ell j} y_j + \frac{1}{2}\sqrt{n}\sum_i a_{i\ell} y_i\right]$$

Since $\|a_{ij}\|$ is symmetric and $i, j = 1, 2, \text{-----}, m$, then

$$\sum_j a_{\ell j} y_j = \sum_i a_{i\ell} y_i$$

then

$$\frac{\partial \log L(\Omega)}{\partial\theta_\ell} = \frac{1}{2}\left[\frac{1}{|a_{ij}|}\frac{\partial|a_{ij}|}{\partial\theta_\ell} - \sum_{i,j}\frac{\partial a_{ij}}{\partial\theta_\ell} y_i y_j + \sqrt{n}\sum_j a_{\ell j} y_j\right]$$

Solving the equations

$$\frac{\partial \log L(\Omega)}{\partial\theta_\ell} = 0 \qquad \ell = 1, 2, \text{---}, m$$

we obtain $\hat{\theta}_1, \text{----}, \hat{\theta}_m$. We can show that $|\check{\theta}_\ell - \hat{\theta}_\ell|$ is of order $\frac{1}{\sqrt{n}}$, since a_{ij} of order 1 and $\|a_{ij}\|$ is non-singular. Then

$$L(\hat{\Omega}) = \frac{|a_{ij}|^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}}(1 + \psi')$$

where ψ' is of order $\frac{1}{\sqrt{n}}$.

Now, we can write $L(\omega)$ as

$$L(\omega) = \frac{|a_{0ij}|^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}}\exp\left\{-\frac{1}{2}\sum_{i,j=1}^{r} a_{ij} y_i' y_j' - \frac{1}{2}\chi_0^2\right\}(1 + \psi_0'),$$

where ψ_0' is of order $\frac{1}{\sqrt{n}}$, $\chi_0^2 = \sum_{i,j=r+1}^{m} a'_{ij} y_i y_j$, $a_{0ij} = -E\left(\frac{\partial^2 \log f}{\partial\theta_i \partial\theta_j}\right)_{\theta_0}$ $i, j = r+1, \text{-------}, m$; $y_i' = y_i - G_i$, where G_i is a linear function of θ_{0i}, $i = r+1, \text{--}, m$, and $\|a'_{ij}\| = \|a''_{ij}\|^{-1}$ where $\|a''_{ij}\|$ is defined by

$$\|a_{ij}\|^{-1} = \begin{bmatrix} A & B \\ B' & \|a''_{ij}\| \end{bmatrix}, \qquad \text{A is an } r \times r \text{ matrix.}$$

By solving the equations

$$\frac{\partial \log L(\omega)}{\partial \theta_\ell} = 0 \qquad \ell = 1, 2, \ldots, m$$

and substituting the estimates obtained in $L(\omega)$ we get

$$L(\hat{\omega}) = \frac{|a_{0ij}|^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}} \exp\left(-\tfrac{1}{2}\chi_0^2\right)\left(1 + \psi_0''\right),$$

where ψ_0'' is of order $\frac{1}{\sqrt{n}}$. Then λ is

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \exp\left(-\tfrac{1}{2}\chi_0^2\right)\left(1 + \psi_1\right)$$

where ψ_1 is of order $\frac{1}{\sqrt{n}}$, and so

$$-2\log\lambda = \chi_0^2 + \psi_2 \qquad \psi_2 = O\left(\tfrac{1}{\sqrt{n}}\right)$$

Here if we neglect ψ_2, then

$$-2\log\lambda = \chi_0^2$$

ie. $-2\log\lambda$ is distributed as χ^2-distribution. Now we have to show that the degrees of freedom of $-2\log\lambda$ are $m-r$. The characteristic function of $-2\log\lambda$ is

$$\Phi(t) = E\left[e^{it(-2\log\lambda)}\right] = \int \cdots \int L(\omega)\, e^{it(\chi_0^2 + \psi_2)}\, dy_1 \cdots dy_m$$

$$= \frac{|a_{0ij}|^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}} \int \cdots \int \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{r} a_{ij}\, y_i' y_j' + \chi_0^2\left(it - \tfrac{1}{2}\right)\right\}\left(1 + \psi_3\right)$$

where $\psi_3 = O\left(\frac{1}{\sqrt{n}}\right)$. Then

$$\Phi(t) = \left(\tfrac{1}{2}\right)^{\frac{m-r}{2}} \left(\tfrac{1}{2} - it\right)^{-\frac{m-r}{2}} \qquad \text{as } n \to \infty$$

on any finite interval $|t| < c$. And since this form is the characteristic function of any quantity distributed as χ^2-distribution with $m-r$ degrees of freedom, then $-2\log\lambda$ is distributed as χ^2-distribution with $m-r$ degrees of freedom.

APPENDIX III

The Distribution of theSample Correlation Coefficient when $\rho=0$

Let $(x_1, y_1), \text{------}, (x_n, y_n)$ be a random sample from a population having a bivariate normal distribution with means, variances and correlation coefficient $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and ρ respectively. Let r be the sample correlation coefficient. We can show that if the null hypothesis $H_0: \rho=0$ is true, then the likelihood ratio test will be such that

$$\lambda = \left\{ 1 - \left[\frac{\sum(x_i-\bar{x})(y_i-y)}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}} \right]^2 \right\}^{\frac{n}{2}}$$

ie, $$1-\lambda^{\frac{2}{n}} = r^2$$

Hence the test may be based on r, thus we must know the distribution of r.

Let $c = \sum(x_i-\bar{x})(y_i-\bar{y})$, $v_1 = \sum(x_i-\bar{x})^2$ and $v_2 = \sum(y_i-\bar{y})^2$, then r will be such that

$$r = \frac{c}{\sqrt{v_1 v_2}}.$$

Now we need to show that r is independent of $\bar{x}, \bar{y}, v_1$ and v_2. When $\rho = 0$, the moment generating function of r will be given by

$$M_r(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi\sigma_1\sigma_2} \right)^n exp \left\{ t \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i-\bar{y})^2}} - \frac{1}{2}\sum D_i \right\} dx_1 dy_1 \cdots dx_n dy_n$$

where $D_i = \left(\frac{x_i-\mu_1}{\sigma_1} \right)^2 + \left(\frac{y_i-\mu_2}{\sigma_2} \right)^2$. Let $l_i = \frac{x_i-\mu_1}{\sigma_1}$ and

$k_i = \frac{y_i-\mu_2}{\sigma_2}$, then

$$M_r(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^n \exp\left\{ t \frac{\sum(l_i - \bar{l})(k_i - \bar{k})}{\sqrt{\sum(l_i - \bar{l})^2 \sum(k_i - \bar{k})^2}} - \frac{1}{2}\sum(l_i^2 + k_i^2) \right\} dl_1 dk_1 \cdots dl_n dk_n$$

We see here that the moment generating function of r is independent of μ_1, μ_2, σ_1 and σ_2. In virtue of the generality of the theorem on page 123 in (15), will be independent of $\bar{x}, \bar{y}, v_1$ and v_2. Hence we can write

$$E(r^2)E(v_1 v_2) = E(c^2)$$

or

$$E(r^2) = \frac{E(c^2)}{E(v_1 v_2)}.$$

Now we are showing that the moment generating function of

$$\sum(x_i - \mu_1)(y_i - \mu_2)$$

is $(1-t^2)^{-\frac{n}{2}}$, where $-1 < t < 1$. Let $A_i = x_i - \mu_1$ and $B_i = y_i - \mu_2$; here A_i and B_i are two random variables distributed normally with means zero and variances one. Then the moment generating function of AB is

$$M_{AB}(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{tAB - \frac{1}{2}(A^2+B^2)} \, dA\, dB,$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(A-tB)^2 - \frac{1}{2}B^2(1-t^2)} \, dA\, dB.$$

Let $u = A - tB$ and $v = B$ then

$$\begin{vmatrix} \frac{\partial u}{\partial A} & \frac{\partial u}{\partial B} \\ \frac{\partial v}{\partial A} & \frac{\partial v}{\partial B} \end{vmatrix} = \begin{vmatrix} 1 & -t \\ 0 & 1 \end{vmatrix} = 1 \quad ; \quad \text{i.e. } J = 1.$$

Hence

$$M_{AB}(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi}\, e^{-\frac{1}{2}u^2 - \frac{1}{2}v^2(1-t^2)}\, du\, dv \ ,$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi}\, e^{-\frac{1}{2}u^2 - \frac{1}{2}Z^2}\, (1-t^2)^{-\frac{1}{2}}\, du\, dz \ ,$$

where $Z^2 = v^2(1-t^2)$. Then

$$M_{AB}(t) = (1-t^2)^{-\frac{1}{2}}$$

and so

$$M_{\sum_1^n A_iB_i}(t) = (1-t^2)^{-\frac{n}{2}} \ .$$

Now we can analyse $\sum(x_i - \mu_1)(y_i - \mu_2)$ such that

$$\sum(x_i - \mu_1)(y_i - \mu_2) = \sum(x_i - \bar{x})(y_i - \bar{y}) + n(\bar{x} - \mu_1)(\bar{y} - \mu_2) \ .$$

Since it can be shown that $\sum(x_i - \bar{x})(y_i - \bar{y})$ is independent of $\bar{x}$ and $\bar{y}$, then the two terms in the left hand side will be independent. Since the moment generating function of $n(\bar{x} - \mu_1)(\bar{y} - \mu_2)$ is $(1-t^2)^{-\frac{1}{2}}$ then the moment generating function of $\sum(x_i - \bar{x})(y_i - \bar{y})$ will be $(1-t^2)^{-\frac{n-1}{2}}$. Since

$$M_c(t) = E(e^{tc}) \ ,$$

then it is easy to show that

$$E(c^m) = M_c^m(0)$$

where $M_c^m(0)$ is the mth derivative of the moment generating function at $t=0$ under the integral sign. From this we find that $M_c^m(0)$ is an odd function when m is odd, and hence its integration over $(-\infty, \infty)$ equal to zero. But when m is even then $M_c^m(0)$ becomes an even function. In our problem m is even, equal to 2. Hence

$$E(c^2) = M_c^2(0) = \left\{ \frac{\partial^2}{\partial t^2} (1-t^2)^{-\frac{n-1}{2}} \right\}_{t=0}$$
$$= n-1 .$$

Now, since each of v_1 and v_2 having a χ^2- distribution with $n-1$ degrees of freedom and since it can be shown that the moment generating function of χ^2 with γ degrees of freedom is

$$(1-2t)^{-\frac{\gamma}{2}} , \qquad \left(|t| < \tfrac{1}{2}\right)$$

then the moment generating function of each of v_1 and v_2 is

$$(1-2t)^{-\frac{n-1}{2}} .$$

Hence

$$E(v_1) = E(v_2) = M'_{v_1}(0) = M'_{v_2}(0) = \left\{ \frac{\partial}{\partial t} (1-2t)^{-\frac{n-1}{2}} \right\}_{t=0}$$
$$= n-1 .$$

Then

$$E(r^2) = \frac{E(c^2)}{E(v_1)\,E(v_2)} = \frac{n-1}{(n-1)^2} = \frac{1}{n-1} .$$

We can show, that if x_1 and x_2 are stochastically independent random variables each having gamma distribution, and their joint probability density function is

$$f(x_1, x_2) = \frac{1}{\Gamma(\alpha)\Gamma(\beta)}\, x_1^{\alpha-1}\, x_2^{\beta-1}\, e^{-x_1-x_2} , \qquad \begin{matrix} 0 < x_1 < \infty & \alpha > 0 \\ 0 < x_2 < \infty & \beta > 0 \end{matrix}$$

then the marginal probability density function of $Z = \frac{x_1}{x_1+x_2}$ will be given by

$$g_1(z) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, z^{\alpha-1} (1-z)^{\beta-1} . \qquad 0 < z < 1$$

Then

$$E(z) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 z^{1+\alpha-1} (1-z)^{\beta-1} \, dz$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}$$

$$= \frac{\Gamma(\alpha+\beta)\,\Gamma(\alpha+1)}{\Gamma(\alpha)\,\Gamma(\alpha+\beta+1)} = \frac{\alpha}{\alpha+\beta} .$$

We see here that if $\alpha = \frac{1}{2}$ and $\beta = \frac{n-2}{2}$ then,

$$E(z) = E(r^2)$$

Hence at $\alpha = \frac{1}{2}$ and $\beta = \frac{n-2}{2}$ we get

$$g_1(z) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} z^{\frac{1}{2}-1} (1-z)^{\frac{n-4}{2}} .$$

Since we are interested in the distribution of r, we let $P = \sqrt{z}$

then

$$g_2(P) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} (P^2)^{\frac{1}{2}-1} (1-P^2)^{\frac{n-4}{2}} \, 2P \qquad 0<P<1$$

$$= 2 \, \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} (1-P^2)^{\frac{n-4}{2}} .$$

Then

$$g(r) = \frac{1}{2} g_2(r) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} (1-r^2)^{\frac{n-4}{2}} \qquad - < r < 1$$

REFERENCES

(1) J. AITCHISON & S. D. SILVEY, Maximum-Likelihood Estimation Procedures and Associated Tests of Significance, The J. of the R. Stat. Soc., Series B (Methodolgical), Vol.22, No.1,1960

(2) J. AITCHISON & J. A. BROWN, The Lognormal Distribution, Cambridge University Press (1956).

(3) Z. W. BIRNBAUM, Introduction to Probability & Mathematical Statistics, Harper & Brothers, Publishers, New York, 1962.

(4) J. BERKSON, A statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response Based on the Logistic Function, J. Amer. Stat. Associ, Vol.48, (1953

(5) J. BERKSON, Application of the Logistic Function to Bioassay, J. Amer. Stat. Assoc. Vol. 39, (357-365), 1944.

(6) H. CRAMER, Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey, 1946.

(7) J. F. DALY, On the Unbiased Character of Likelihood Ratio Tests for Independence in Normal Systems, Ann. Math. Stat. Vol. 11, 1940.

(8) D. A. S. FRASER, Statistics: An Introduction, John Wiley & Sons, New York.

(9) R. A. FISHER, Contributions to Mathematical Statistics, John Wiley & Sons, New York, 1950.

(10) R. A. FISHER & F. YATES, Statistical Tables for Biological, Agricultural and Medical Research, Oliver and Boyd, London.

(11) R. A. FISHER & B. BALMUKAND, The Estimation of Linkage from the Offspring of Selfed Heterozygotes, Journal Genetics, Vol. 20, 79-92, 1928.

(12) R. A. FISHER, Statistical Methods for Research Workers, Oliver and Boyd.

(13) D. J. FINNEY, Probit Analysis, Cambridge University Press, 1947.

(14) J. B. HUTCHINSON, The Application of the "Method of Maximum Likelihood" to the Estimation of Linkage, Genetics, Vol. 14, 519-537, 1929.

(15) R. V. HOGG & A. T. CRAIG, Introduction to Mathematical Statistics, University of Iowan, New York, The Macmillan Company.

(16) J. O. IRWIN & E. A. CHEESEMAN, On the Maximum Likelihood Method of Determining Dosage-Response Curve & Approximations to the Median-Effective Dose, in Cases of A Quantal Response, J. R. Stat. Soc. Suppl. 6, 1939.

(17) M. G. KENDALL. The Advanced Theory of Statistics. Vol. II, Charles Griffin & Company Ltd., London, 1946.

(18) J. F. KENNEY & E. S. KEEPING, Mathematical of Statistics, Vol. II, D. VAN NOSTRAND COMPANY, Princeton, New Jersey.

(19) D. V. LINDLY & J. C. P. MILLER, Cambridge Elementary Statistical Tables, Cambridge University Press, 1958.

(20) K. MATHER & R. A. FISHER, Statistical Analysis in Biology

(21) E. PAULSON, On Certain Likelihood-Ratio Tests Associated with the Experimental Distribution, Ann. Math. Stat. Vol. 12, 1941.

(22) C. R. RAO, Advanced Statistical Methods in Biometric Research, New York, John Wiley & Sons, London.

(23) S. S. WILKS, The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses, Ann. Math. Stat. Vol. 9, 1938.

SUMMARY

This thesis is the comprehensive study of the method of maximum likelihood and its relative merit over other methods of estimation. This method of estimation, developed by R. A. Fisher in 1921 is the oldest method. Since that time, Fisher and some others have introduced wide successive developments which led the maximum likelihood method to be used in most practical applications.

In chapters I and II (where single and several parameters are considered) it has been shown that the method of maximum likelihood has all the properties of the best method of estimation; that is, the estimators of the maximum likelihood method have the property of consistency, and they are asymptotically most efficient, having normal distribution and also they are unbiased estimators. Also it has been shown that if a sufficient estimator exists, then the method of maximum likelihood affords it. The inequality of Fisher has been discussed which supplies the maximum attainable variance when the equality holds. There has also been discussed the process of the successive approximations by which the maximum likelihood estimates can be obtained in cases when the maximum likelihood equations are difficult to be solved. The Wald technique and Lagrange multiplier technique are explained for estimating the unrestricted and the restricted parameters with their tests respectively.

In chapter III there has been shown the practical applications of the method of maximum likelihood. In the field of genetics we applied some other methods in addition to the

maximum likelihood method and we saw that the estimates of this method are the most efficient. In thefield of bioassa we have shown the applications of the method of maximum likelihood for estimating the two parameters using the probi transformation and the logistic formula. In the field of blood groups, the application of the maximum likelihood meth has been shown for estimating the three parameters. We hav mentioned the Bernstien method and applied both the Wald and the Lagrange multiplier techniques for estimating the unrestricted and the restricted parameters.

In chapter IV we discussed the likelihood ratio test which is frequently unbiased and based on a sufficient stati and also it is the uniformly most powerful test. In virtue of the desirable properties mentioned above, this test becomes more accurate for testing the statistical hypothesis than the others.