CHAPTER I

PRELIMINRIES

1.0 Introduction :

Bayes procedure for testing of hypothesis is a statistical decision problem when parameter space and decision space contains only two points. In the present chapter it is emphasied that hypothesis testing is a decision problem. Also the analogue between MP test in a classical test procedure and Bayes test procedure when null and alternative hypothesis are simple is explained, with various cases arising. The same is explained with the help of examples on normal and binomial distribution in Section 1.3.

As basic requirement . classical test procedures are discribed in section 1.1. In Section 1.2 general decision problems are discussed with illustrations regarding Bayes decision, Bayes risk and construction of Bayes decision function. In Section 1.2 it is shown that as the number of observations increases the Bayes risk decreases. However, if the cost factor due to sampling is considered the Bayes risk may tend to infinity as $n \rightarrow \infty$. In a certain models optimum sample size can be computed, corresponding to which the Bayes risk is minimum.

situations such as this do not arise in practical Problems. The procedure used in practice is to limit the probability of type I error to some preassigned level α (usually 0.01 or 0.05) that is small and to minimize the probability of type II error.

To every $\underset{\sim}{x} \in S$ we assign a number $\phi(\underset{\sim}{x})$, $0 \leq \phi(\underset{\sim}{x}) \leq 1$, which is the probability of rejecting H_1 that $X \sim f_w$, $w \in W_1$, if $\underset{\sim}{x}$ is observed. If H_1 is true ϕ rejects it with probability $\leq \alpha$. We call such a test a randomized test function. If $\phi(x) = I_A(x)$, ϕ will be called non-randomized test. **Consider a problem of finding ϕ for given W_1, W_2 and α**

In the subsequent, this problem will be denoted by (α, W_1, W_2) and let ϕ be a test function for the problem. As a function of w, $\beta_\phi(w)$ defined below is called power function of the test ϕ

$$\beta_\phi(w) = E_w\, \phi(X)$$
$$= P_w\{\text{Reject } H_1\}\ ; \quad w \in \Omega.$$

Now let us formulate the problem of testing of hypothesis **as follows**. Find a test $\phi(x)$ such that $\beta_\phi(w) \leq \alpha$ for $w \in W_1$, and $\beta_\phi(w)$ is maximum for $w \in W_2$. Let ϕ_α be the class of all tests for the problem (α, W_1, W_2) where W_1, W_2 are singleton. A test $\phi^* \in \phi_\alpha$ is said to be a most powerful (MP) test against an alternative $w \in W_2$

for some $k \geq 0$ and $0 \leq \gamma(x) \leq 1$, is most powerful of its size. for testing H1: $w = w_1$ against H2: $w = w_2$

In particular if $k = \infty$ the test

$$\phi(x) = \begin{cases} 1 & \text{if } f_1(x) = 0 \\ 0 & \text{if } f_1(x) > 0 \end{cases} \qquad (1.1.2)$$

is most powerful of size 0 for testing H1 against H2.

Given α, $0 \leq \alpha \leq 1$, there exists a test of form (1.1.1) or (1.1.2) with $\gamma(\underset{\sim}{x}) = \gamma$ (constant) for which $E_{w_1}\phi(\underset{\sim}{X}) = \alpha$. Note that MP test is not unique on the $\{X:\ f_2(\underset{\sim}{x}) = kf_1(\underset{\sim}{x})\}$.

We now consider a problem of testing one-sided hypothesis on a single real valued parameters. Suppose we wish to test H_1: $w \leq w_1$ against the alternative H_2: $w > w_1$ or its dual H_1': $w \geq w_1$ against H_2': $w < w_1$. Here we consider a special class of distributions called family of distributions having MLR property which is large enough to include one parameter exponential family, for which a UMP test of a one-sided hypothesis considered above exists.

Let $\underset{\sim}{X} \sim f_w$, $w \in \Omega$, $\Omega \subseteq R$, where $\{f_w\}$ has an MLR in $T(\underset{\sim}{x})$ (Refer defination 0.1). For testing H_1: $w \leq w_1$ against H_2: $w > w_1$, $w_1 \in \Omega$, any test of the form

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } T(\underset{\sim}{x}) > t_0 \\ \gamma & \text{if } T(\underset{\sim}{x}) = t_0 \\ 0 & \text{if } T(\underset{\sim}{x}) < t_0 \end{cases}$$

has a non-decreasing power function and is UMP of its size $E_{w_1}\phi(\underset{\sim}{X})$.

Moreover, for every $0 \leq \alpha \leq 1$ and every $w_1 \in \Omega$ there exists a t_o, $-\infty \leq t_o \leq \infty$, and $0 \leq \gamma \leq 1$ such that the test described above is a UMP size α test of H_1 against H_2.

By interchanging inequlaities throughout we see that it provides a solution of the dual problem H_1': $w \geq w_1$ aginst H_2': $w < w_1$.

By restricting the class ϕ_α of all tests of size α, there do not exists UMP tests for many important null hypothesis. For example the UMP test dose not exists(i) for testing H_1: $w_1 \leq w \leq w_2$ and (ii) H_1': $w = w_1$ in case of one parameter exponential family. In this case one has to look for a UMP test in a restricted class of test.

Defination :

A size α test ϕ of H_1: $w \in W_1$ against the alternative H_2: $w \in W_2$ is said to be unbiased if

$$E_w \phi(\underset{\sim}{X}) \geq \alpha \text{ for all } w \in W_2$$

Let U_α be the class of all unbiased size α tests of H_1, if there exists a test $\phi \in U_\alpha$ that has maximum power at each $w \in W_2$, we call ϕ a UMP unbiased size α test.

In case of exponential family with density defined in (0.2) The UMP unbiased test for testing H_1: $w = w_1$ against H_2: $w \neq w_1$ the UMP unbiased test is given by

$$\phi(x) = \begin{cases} 1 & \text{if } T(x) < C_1 \text{ or } T(x) > C_2 \\ \gamma_1 & \text{if } T(x) = C_1 \\ \gamma_2 & \text{if } T(x) = C_2 \\ 0 & \text{if } C_1 < T(x) < C_2 \end{cases}$$

and $E_{W_1} [\ \phi(X)\ ] = \alpha.$

$$\frac{d}{dW} E_W[\phi(X)\ ] \ / \ W = W_1 \ = 0$$

<u>1.1.1 Example</u> :

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed (i.i.d) $b(1,p)$ random variables and let H_1: $p = p_1$, H_2: $p = P_2$, $P_2 > P_1$. Then MP size α test of H_1 against H_2 is of the form

$$\phi(x_1, \ldots x_n) = \begin{cases} 1, & \lambda(x) = k \\ \gamma, & \lambda(x) = k \\ 0\ , & \lambda(x) < k \end{cases}$$

where $\lambda(x) = \dfrac{P_2^{\sum_1^n x_i} (1-P_2)^{n-\sum_1^n x_i}}{P_1^{\sum_1^n x_i} (1-P_1)^{n-\sum_1^n x_i}}$

In this case γ and k are determined by

$$E_{P_1}[\phi(X)\ ] = \alpha.$$

$$\lambda(x) = \left(\frac{P_2}{P_1}\right)^{\sum_1^n x_i} \left(\frac{1-P_2}{1-P_1}\right)^{n-\sum_1^n x_i}$$

And since $P_2 > P_1$, $\lambda(x)$ is an increasing function of Σx_i, it follows that $\lambda(x) > k$ if and only if $\Sigma x_i > k_1$, where k_1 is some constant. Thus MP test reduces to

$$
\phi(x) = \begin{cases} 1 & \text{if } \sum_{1}^{n} x_i > K_1 \\ \gamma & \text{if } \sum_{1}^{n} x_i = K_1 \\ 0 & \text{otherwise} \end{cases}
$$

In particular $n = 5$, $P_1 = 1/2$, $P_2 = 3/4$, $\alpha = 0.05$ MP test is given by

$$
\phi(x) = \begin{cases} 1 & , \text{ if } \sum_{1}^{n} x_i > 4 \\ 0.122 & , \text{ if } \sum_{1}^{n} x_i = 4 \\ 0 & , \text{ if } \sum_{1}^{n} x_i < 4 \end{cases}
$$

where k and γ are determined by,

$$
0.05 = \alpha = \sum_{k+1}^{5} \binom{5}{r} (1/2)^5 + \gamma \binom{5}{k} (1/2)^5 .
$$

Thus the MP size $\alpha = 0.05$ test is to reject $p = 1/2$ in favour of $p = 3/4$ if $\sum_{1}^{5} x_i = 5$ and reject $p = 1/2$ with probability 0.122 if $\sum_{1}^{n} x_i = 4$.

Let X1,X2,..........Xn be i.i.d. b(1,p) random variables, the UMP test for testing H1: p= P1 against H2: P P2 of size follows.

The p.d.f. of X1 is given by,

we have $f_p(x) = \binom{1}{x} p^x (1-p)^{1-x}$, $x = 0,1$.

$$= \binom{1}{x} e^{x \log \frac{p}{1-p} + \log(1-p)}$$

The corresponding likelihood is:

$$f_p(x) = \prod_{1}^{n} \binom{1}{x_i} e^{\log \frac{p}{1-p} \sum_{1} x_i + n\log(1-p)}$$

This is in the form of one parameter exponential density (ref. 0.2) with $T(x) = \sum_{1} x_i$

Therefore UMP test is of the form.

$$\phi(x) = \begin{cases} 1, & \text{if } T(x) > t_o \\ \gamma, & \text{if } T(x) = t_o \\ 0, & \text{if } T(x) < t_o \end{cases}$$

t_o and γ are determined by

$$E_{P_1}[\phi(x)] = \alpha.$$

for $P_1 = 1/2$, $n=5$ $\alpha = 0.05$ it follows that $t_o = 4$ and $\gamma = 0.122$.

Now let use consider the case where alternative hypothesis is two sided.

H_1: $p = P_1$ against H_2: $p \neq P_1$ the UMPU test is

$$\phi(x) = \begin{cases} 1 & \text{if } \sum_{1}^{n} x_i < C_1 \text{ or } \sum_{1}^{n} X_i > C_2. \\ \gamma_1 & \text{if } \sum_{1}^{n} x_i = C_1 \\ \gamma_2 & \text{if } \sum_{1}^{n} x_i = C_2 \\ 0 & \text{if } C_1 < \sum_{1}^{n} X_i < C_2. \end{cases}$$

where C_1 and C_2 are determined

for $X_i \sim Bi(10;p)$, $n=12$, $\alpha = 0.05$ as 49 and 71 respectively.

1.2 General Decision Problem :

1.2.a Bayes Risk And Bayes Decisions :

Consider a decision problem defined by a parameter space Ω, a decision space D and a loss function L. For any distribution P of the parameter W, let the risk be $\rho(P,d)$ for $d \in D$ where $\rho(P,d) = \int_{\Omega} L(w,d)\, d\, P(w)$ and it is assumed that it is finite for every $d \in D$. Then the Bayes risk $\rho^*(P) = \inf_{d \in D} \rho(P,d)$. Any decision d* whose risk is equal to Bayes risk is called a Bayes decision against the distribution P if exists.

i.e. $\rho(P,d^*) = \inf_{d \in D} \rho(P,d)$

1.2.a Example : $\Omega : \{0, \ 1\}$

$D : \{\text{All the numbers } d, \ 0 \leq d \leq 1\}$

$$L(w,d) = |w-d|$$

$$Pr(W = 0) = 3/4; \ Pr(W = 1) = 1/4$$

$$\begin{aligned} \rho(P,d) &= L(0,d)\, Pr(W = 0) + L(1,d)\, Pr(W = 1) \\ &= 3/4.d + 1/4 \,.\, (1-d) \\ &= d/2 + 1/4 \end{aligned}$$

For d = 0, $\rho(P,d)$ is minimum.

$\therefore$ d = 0 is the unique Bayes decision with Bayes risk $\rho^*(P) = 1/4$.

In this case, of the two possible values of the parameter, 0 is a logical estimate of the parameter which of course is the Bayes decision. If $P[W=0] < P[W=1]$ then Bayes decision would be W = 1.

If $D' = \{0, 1\}$ then

$$\rho^*(P) = \inf_{d \in D} \rho(P,d)$$

$$= 1/4$$

Note that there does not exists a d* in D' = (0,1) for which $\inf_{d \in D'} \rho(P,d)$ is attained i.e. Bayes rule in D' does not exists.

1.2.b. Concavity of Bayes Risk :

Theorem : 1.2

For any distributions P_1 and P_2 of W and for any number α $(0 \leq \alpha \leq 1)$;

$$\rho^*[\alpha P_1+(1-\alpha)\ P_2] \geq \alpha \rho^*(P_1)+(1-\alpha)\ \rho^*(P_2)$$

Proof :

We have $\rho(P,d) = \int_{\Omega} L(w,d)\ d\ P(w)$ for any $d \in D$

$$\rho[\alpha P_1+(1-\alpha)\ P_2,d] = \alpha\ \rho(P_1,d) + (1-\alpha)\rho(P_2,d)$$

Now,

$$\rho^*[\alpha P_1+(1-\alpha)\ P_2\ ] = \inf_{d \in D} \rho[\alpha P_1+(1-\alpha)\ P_2,d\ ]$$

$$= \inf_{d \in D} [\alpha\ \rho(P_1,d)+(1-\alpha)\rho(P_2,d)]$$

This gives

$$\rho^*[\alpha P_1+(1-\alpha)\ P_2] \geq \alpha \inf_{d \in D} \rho(P_1,d)+ (1-\alpha) \inf_{d \in D} \rho(P_2,d)$$

$$= \alpha\ \rho^*(P_1) + (1-\alpha)\ \rho^*(P_2).$$

1.2.c. Randomization ~~or~~ and mixed decisions :

Let $d_1, d_2, \ldots.$ be a sequence of decisions (i.e. the number of decisions are countable). Let us assign the probabilities $P_1, P_2, \ldots$ to the sequence of decisions. $d_i \in D$. The process of selecting one of the decision d_i on the basis of these probabilities is called mixed or randomized decision. Thus a randomized decision is nothing but a probability distribution defined on the decision space D.

$$L(w,d) = \sum_{i=1}^{\infty} P_i \, L(w,d_i) \qquad (1.2.1)$$

is the loss associated with the mixed decision d for $w \in \Omega$.

Let M denote the set of all mixed decisions in a given problem; where D is the class of pure decisions. Trivially we can regard each pure decision (non-randomised decision) d as a mixed decision in which pure decision must be selected with probability 1. Hence $D \subset M$.

The loss function given above (1.2.1) for mixed decision is weighted average of the loss functions defined for pure decisions. Therefore, whenever the risk $\rho(P,d)$ for mixed decision exists its value must be the weighted average of the risks $\rho(P,d_i)$ of pure decisions d_i.

Hence $\inf_{d \in M} \rho(P,d) = \inf_{d \in D} \rho(P,d) = \rho^*(p)$

It follows that if the Bayes risk $\rho^*(p)$ is finite and is attained for a mixed decision in M then this risk must also be attained for some pure decision in D. Hence when we come across two or more pure decisions each yields a Bayes risk. It is advisible to perform an auxillary randomization to select one of these Bayes decisions. Randomization in this situation is irrelevant because any method of selecting one of the Bayes decisions is acceptable.

1.2.d. Decision Problem with Observations :

It may be possible to observe the value of the random variable or a random vector which gives informations about the value of W and helps in taking good decision. Let S be the sample sapce of all possible values of the observations X. The decision chosen depends on observations; so for each possible value $x \in S$ a decision $\delta(x) \in D$. The class of all decision functions δ will be denoted by Δ .

For any g.p.d.f. ξ of the parameter W and any decision function $\delta \in \Delta$ the risk

$$\rho(\xi, \delta) = E\ L[W, \delta(X)]$$
$$= \int_{\Omega}\int_{S} L(w, \delta(x)]\ f(x/w)\ \xi(w)\ dx\, dw \qquad \ldots (1.2.2)$$

Assume for each value $w \in \Omega$ the function $L[w, \delta(.)]$ is measurable and integrable over the set S. For any particular value of $w \in \Omega$, $\rho(w, \delta)$ denotes the risk of the decision function δ when $W = w$ and is given by

$$R(w,\partial) = \int_S L[w,\partial(x)]\ f(x/w)\,dx \qquad (1.2.3)$$

$$\rho(\xi,\partial) = \int_\Omega \rho(w,\partial)\ \xi(w)\ dw \qquad (1.2.4)$$

Let $\partial^* \in D$ s.t.

$$\rho(\xi,\partial^*) = \inf_{\partial \in \Delta} \rho(\xi,\partial) = \rho^*(\xi) \qquad (1.2.5)$$

Then ∂^* is Bayes decision against ξ and $\rho^*(\xi)$ is Bayes risk.

1.2.e. Construction of Bayes Decision Function :

We have by (1.2.2)

$$\rho(\xi,\partial) = \int_\Omega \int_S L[w,\partial(x)]\ f(x/w)\ \xi(w)\ dx\,dw$$

Since the loss function is non-negative or a bounded function the order of integration in the above integral can be interchanged

$$\rho(\xi,\partial) = \int_S \int_\Omega \left\{ L[w,\partial(x)]\ f(x/w)\ \xi(w)\ dw \right\} dx \qquad \ldots\ (1.2.6)$$

For each value $x \in S$ let $\partial^*(x) = d^*$ where d^* is any decision in D which minimizes the integral.

$$\int_\Omega L(w,d)\ f(x/w)\ \xi(w)\ dw \qquad (1.2.7)$$

Let f_1 is the marginal g.p.d.f. of X the value of $f_1(x)$ can be 0 only on set of points x which has probability 0.

$$f_1(x) = \int_\Omega f(x/w)\ \xi(w)\ dw \qquad (1.2.8)$$

Now, instead of finding a decision d^* which minimizes (1.2.7) we can find equivalently a decision d^* which minimizes

$$\int_{\Omega} L(w,d) \left[\frac{f(x/w)\, \xi(w)}{f_1(x)} \right] dw \qquad (1.2.9)$$

which is the conditional expectation and can be written as $E[L(w,d)/x)$. Therefore, any minimizing decision d* is simply a decision which yields the smallest expected loss under the conditional distribution of W when the observed value X is x. In statistical decision problem the marginal distribution of W is called the <u>prior distribution</u> of W. Because it is distribution of W before X has been observed. And the conditional distribution of W when the value of X is known is called the <u>posterior distribution</u> of W because it is the distribution of W after X has been observed.

<u>Example 1.2.e</u> **As an illustration consider the following example.**

Let $X_1, X_2, \ldots, X_n$ be i.i.d. normal variates with mean θ variance σ_0^2 and the prior distribution of θ be normal with mean μ and variance τ^2. Let the loss function be given

$$L(\theta, a) = \begin{cases} 0, & \text{if } |\theta - a| \leq C \\ 1, & \text{if } |\theta - a| > C \end{cases}$$

where C is positive and known.

Let $x_1, x_2, \ldots, x_n$ be observations denoted by $x=(x_1, x_2, \ldots, x_n)$. The posterior distribution of θ is normal with mean $\hat{\theta}(x)$ and variance σ_1^2 where

$$\tilde{\theta}(x) = \left(\frac{\mu}{\tau^2} + \frac{n\bar{x}}{\sigma_o^2} \right) \sigma_1^2$$

and

$$\frac{1}{\sigma_1^2} = \frac{1}{\tau^2} + \frac{n}{\sigma_o^2}$$

For each given x we are interested to find that $\partial(x)$ for which

$$\int_{\Omega} L(\theta,\partial(x))\, P(\theta/x)d\theta \text{ is minimum.}$$

That is $\int_{|\theta-\partial(x)|>C} P(\theta/x)d\theta$ should be minimum

$$\int_{|\theta - \partial(x)|>C} P(\theta/x)\, d\theta = \int_{-\infty}^{\partial(x)-C} P(\theta/x)\, d\theta + \int_{\partial(x)+C}^{\infty} P(\theta/x)d\theta.$$

$$= \int_{-\infty}^{\partial(x)-C} \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp - \frac{1}{2\sigma_1^2}(\theta-\tilde{\theta}(x))^2\, d\theta$$

$$+ \int_{\partial(x)+C}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_1} \exp - \frac{1}{2\sigma_1^2}(\theta-\tilde{\theta}(x))^2\, d\theta.$$

Differenciating with respect to (w.r.t.) $\partial(x)$ and equating to O we get

$$\frac{1}{\sqrt{2\pi}\sigma_1} \exp - \frac{1}{2\sigma_1^2}[\partial(x)-C-\tilde{\theta}(x)]^2 - \frac{1}{\sqrt{2\pi}\sigma_1} \exp - \frac{1}{2\sigma_1^2}[\partial(x)+C-\tilde{\theta}(x)]^2 = 0$$

...(1.2.10)

That is $[\partial(x) - C - \tilde{\theta}(x)]^2 = [\partial(x) + C - \tilde{\theta}(x)]^2$

That is $\partial(x) - C - \tilde{\theta}(x) = \pm [\partial(x) + C - \tilde{\theta}(x)]$

For $C > 0$, $\partial^*(x) = \tilde{\theta}(x)$ = posterior mean.

Differenciating one more time L.H.S. of (1.2.10) and putting

$\partial(x) = \hat{\theta}(x)$ we get value of the expression positive.

Hence $\int_{\Omega} L(\theta, \partial(x)))\, P(\theta/x)\, d\theta$ is minimum at $\hat{\theta}(x)$.

Consider $L(\theta,a) = (\theta - a)^2$

For that $\partial^*(x) = \hat{\theta}(x)$.

Now the Bayes risk corresponding to the square error loss function denote d by $\rho^*_{\partial(x)}(n) = \sigma_1^2$ where $\sigma_1^2 = \dfrac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma_o^2}}$

If we do not take any observation

$$\partial^*(x) = \mu$$

The correppnding Bayes risk denoted by $\rho^*\partial(0)$ is

$$E(\theta - \mu)^2 = \tau^2 = \frac{1}{1/\tau^2} = \rho^*\partial(0)$$

gives

$$\rho^*_{\partial(x)}(n+1) < \rho^*_{\partial(x)}(n) < \rho^*\ \partial(0)$$

Hence Bayes risk is decreasing function of n. But in practice observations will add the cost so we have to consider cost of sampling also.

1.2.f. Cost Function :

The cost of observing the value of X may depend on x and the population from which it is drawn, that is, population parameter W. Let $C(W, x_1, x_2, \ldots, x_n)$ be the cost function when $x_1, x_2, \ldots, x_n$ observed and W is true value.

The total cost $= E_{W,X}[L(W, \partial(X))] + E_{W,X}[C(w, X_1, \ldots, X_n)]$

where the expectation is taken w.r.t. X as well as W.

Let 'c' be the cost of an observation.

Total risk $= \rho^*_{\partial(x)} + nc$.

Let us study the cost with reference to example (1.2.e).

$$\text{Total cost} = \frac{1}{\frac{1}{\varphi^2} + \frac{n}{\sigma_o^2}} + n\,c = C \text{ (say)}$$

Assuming C is defined for n positive real numbers and by differenciating and equating to O, we get,

$$\frac{\partial C}{\partial n} = \frac{-1/\sigma_o^2}{\left(\frac{1}{\varphi^2} + \frac{n}{\sigma_o^2}\right)^2} + c = 0.$$

That is

$$\left(\frac{1}{\varphi^2} + \frac{n}{\sigma_o^2}\right)^2 = \frac{1}{c\sigma_o^2}$$

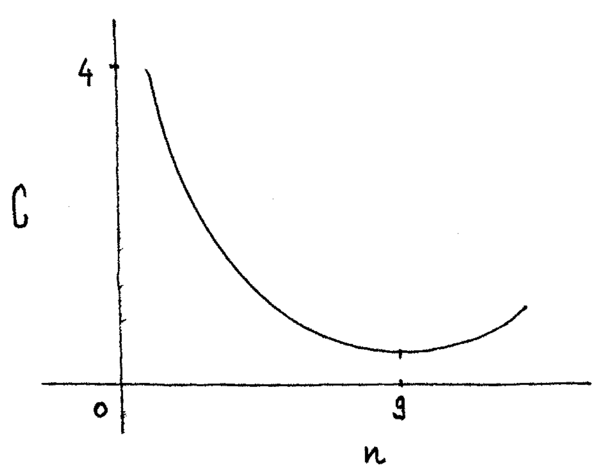
gives $n^* = \sqrt{\left(\frac{\sigma_o^2}{c}\right)} - \frac{\sigma_o^2}{\varphi^2}$, optimum sample size. Note that as expected n* is decreasing function of c. It is observed that $\frac{\partial c}{\partial n} < 0$ for n<n* and $\frac{\partial c}{\partial n} > 0$ for n>n*. Thus optimum value of n is decided by comparing the value of C at n* and n*+1.

In particular, let $\varphi^2 = 400$, $\sigma_o^2 = 4$ and $c = 0.05$ gives $n^* = 8.93 \simeq 9$

To sketch the graph of total risk as a function of sample size when all other parameters are fixed we compute it for different values of n.

The table below shows total risk for different values of n, for τ^2=400, σ_0^2= and c=0.05 (ref. example 1.2.e).

n	$\rho^*_{\delta}(x)$	C = Total risk
0	400	400
1	3.9604	4.0104
2	1.9900	2.0900
3	1.3289	1.4789
4	0.9975	1.1975
5	0.7984	1.0484
6	0.6655	0.9655
7	0.5706	0.9206
8	0.4993	0.8993
9	0.4439	0.8939
10	0.3996	0.8986
11	0.3633	0.9133



1.3 Testing As A Decision Problem :

1.3.a. Statistical decision problem when Ω, D contains two points :

Consider a statistical decision problem in which $\Omega = \{w_1, w_2\}$, $D = \{d_1, d_2\}$. The loss table is as below :

$\Omega \backslash D$	d_1	d_2
w_1	0	a_1
w_2	a_2	0

a_1, a_2 positive known constants.

Table (1.3.a)

For any decision function δ let $\alpha(\delta)$ denote the conditional probability that decision d_2 will be chosen when $w = w_1$. Also let $\beta(\delta)$ denote the conditional probability that decision d_1 will be chosen when $w = w_2$. That is $\alpha(\delta)$ and $\beta(\delta)$ denote the probabilities of wrong decisions when $w = w_1$ and $w = w_2$ respectively.

If the prior distribution of w is

$P(w = w_1) = \xi$, $(0 < \xi < 1)$ $P(\omega = \omega_2) = 1 - \xi$.

If $\xi = 0$ or 1 then the solution is trivial.

That is $\xi = 0$ we have $P(w = w_2) = 1$ decision d_2 is correct decision.

The risk $\rho(\xi, \delta)$ of the decision function δ is

$$\rho(\xi, \delta) = a_1 . \xi . \alpha(\delta) + a_2 . (1 - \xi) \beta(\delta)$$

$$= a . \alpha(\delta) + b\, \beta(\delta) \text{ where } a = a_1 \xi \text{ , } b = a_2(1 - \xi)$$

Here a and b are given positive constants.

Problem is to find a decision rule δ^* in D that minimizes $\rho(\xi, \delta)$; $\delta \in \Delta$ where Δ is the class of decision functions.

1.3.b. Hypothesis testing as a decision problem :

Suppose H_1: $w \in W_1$ against the alternative hypothesis H_2: $w \in W_2$ (W_1 and W_2 are two mutually exclusive sets of the parameter space Ω). In this case $D = \{d_1, d_2\}$ where d_1 means the decision that $w \in W_1$ and d_2 is the decision that $w \in W_2$. That is there are just two possible actions to be taken. This is why hypothesis testing is a two decision problem, involving the two alternative decisions d_1 and d_2. Consider the loss table (table 1.3.a) For such a loss function the risk function of the decision rule (i.e. test) ∂, having the critical regions S_1 wil be,

$$\rho(w, \partial) = \begin{cases} a_1 \int_{S_1} f_w(x)\, dx, & \text{if } w \in W_1 \\ a_2 \int_{S_2} f_w(x)\, dx, & \text{if } w \in W_2 \end{cases}$$

where $S_2 = S_1^c$

Since $\partial(x) = d_1$ if $x \in S_2$ and

$\partial(x) = d_2$ if $x \in S_1$.

For a given prior distribution, represented by the probability density function (p.d.f.) ξ the average risk for $\partial$is

$$\rho(\xi, \partial) = \int_{W_1} a_1 [\int_{S_1} f_w(x)\, dx] \xi(w)\, dw + \int_{W_2} a_2 [\int_{S_2} f_w(x) dx] \xi(w) dw.$$

$$= \int_{S_1} f_\xi(x) \int_{W_1} a_1\, \xi_x(w) dw\, dx + \int_{S_2} f_\xi(x) \int_{W_2} a_2 \xi_x(w)\, dw\, dx$$

Here $f_\xi(x) = \int_\Omega \xi(w)\, f_w(x)\, dw$ and

$$\xi_x(w) = \frac{\xi(w) f_w(x)}{f_\xi(x)}$$

In order to obtain a Bayes rule, we shall have to minimize $\rho(\xi, \delta)$.

Obviously this can be achieved by taking, for any given x, the decision d_1 (or d_2)(or equivalantly, deciding that $x \in S_2$ or deciding that $x \in S_1$) according as $\int_{W_2} a_2 \xi_x(w)\, dw < (>) \int_{W_1} a_1 \xi_x(w)\, dw$.

Let

$$R_1 = \text{Risk in accepting } H_1$$

$$= \int_{W_2} a_2 \, \xi_x(w)\, dw, \quad \text{and} \tag{1.3.1}$$

$$R_2 = \text{Risk in accepting } H_2$$

$$= \int_{W_1} a_1 \, \xi_x(w)\, dw. \tag{1.3.2}$$

If $R_1 < R_2$ decision d_1 should be chosen.
If $R_1 > R_2$ decision d_2 should be chosen (1.3.3)

In case if equality holds we may take either decision. Equivalently, accept H_1 based on x provided the posterior risk (given x) in accepting H_1 is less than that of rejecting H_1 (given x).

The nature of a Bayes test may be seen to have a striking similarity with that of an MP test for a simple hypothesis H_1 against a simple alternative H_2. In case of an MP test, a given sample point x is or is not included in S_1 is decided by keeping in view the relative magnitude of the probability density of X under H_1 and the probability density under H_2. In the case of Bayes test, we consider a sort of weighted average density under H_1 and a weighted average density under H_2, both of which may be composite hypothesis, condition (1.3.3) namely reject H_1 if

$$\int_{W_2} a_2 \xi_X(w) \, dw > \int_{W_1} a_1 \xi_X(w) \, dw$$

may be taken to be Bayes critical region, For loss table defined in table (1.3.a), and for the hypothesis H_1: $w \in W_1$ against H_2: $w \in W_2$. That is

$$\int_{W_2} \xi_X(w) \, dw > \frac{a_1}{a_2} \int_{W_1} \xi_X(w) \, dw \qquad (1.3.4)$$

If in addition H_1 and H_2 are simple hypothesis the condition (1.3.a) takes the form,

$$f_{w_2}(x) \, \xi(w_2) > \frac{a_1}{a_2} f_{w_1}(x) \, \xi(w_1)$$

where H_1: $w = w_1$ and H_2: $w = w_2$.
Here $\xi(w_1)$ and $\xi(w_2) = 1 - \xi(w_1)$ are the prior probabilities attached to w_1 and w_2 respectively. This is exactly similar to the condition defining an MP test, only, the ξ's are then ignored and the constant $\frac{a_1}{a_2}$ is determined by the prescribed level of the test.

Since hypothesis testing is just a form of decision with its own special notation and calculations, we need to make only a few changes, mostly notational.

1. We shall suppose that the data X has been observed so that the posterior probabilities $P(w/X)$ are available.
2. The hypothetical states H_1 and H_2 are the possible decisions ''accept' H_1 and reject H_1''.
3. We shall abbreviate the loss function $L(w_i, d_j)$ to l_{ij}, where d_j are decisions.

Now to generalise the testing problem we proceed as below. Let L_1 denote the average loss in accepting H_1

$$L_1 = P(W_1/X)\, l_{11} + P(W_2/X)\, l_{21}$$

Similarly the average loss in accepting H_2 is

$$L_2 = P(W_1/X)\, l_{12} + P(W_2/X)\, l_{22}$$

Choose d_1 iff $L_1 < L_2$.

That is $P(W_2/X)(l_{21} - l_{22}) < P(W_1/X)(l_{12} - l_{11})$

The bracketed quantities are called regrete r_1 and r_2.

That is $r_1 = l_{12} - l_{11}$

$r_2 = l_{21} - l_{22}$.

Thus r_1 is the extent to which l_{12} exceeds l_{11} that is extra loss incurred by the wrong decision when H_1 is true. Similarly r_2 is the extra loss incurred by wrong decision when H_2 is true.

Therefore accept H_1 iff

$$P(W_2/X)r_2 < P(W_1/X)r_1 .$$

That is

$$\frac{P(W_2/X)}{P(W_1/X)} < \frac{r_1}{r_2} .$$

that is

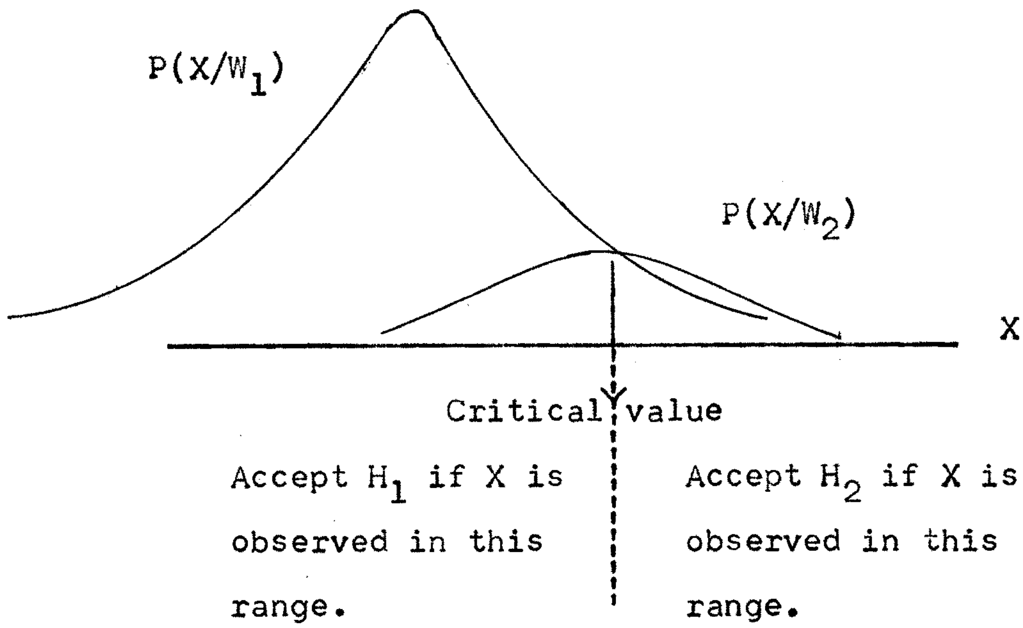
$$\frac{P(W_2)\,P(X/W_2)}{P(W_1)\,P(X/W_1)} < \frac{r_1}{r_2} .$$

that is

$$\frac{P(X/W_2)}{P(X/W_1)} < \frac{r_1 P(W_1)}{r_2 P(W_2)} \qquad (1.3.5)$$

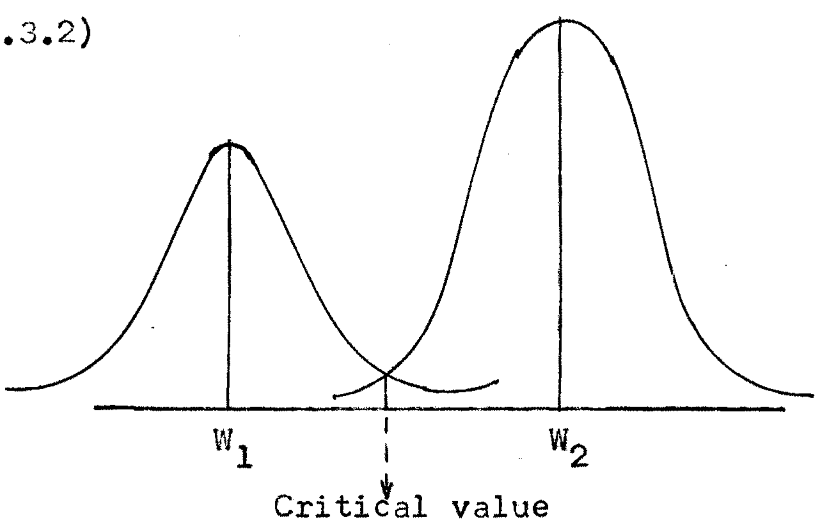
(1.3.5) is called as Bayesian likelihood criterion. This criterion is certainly reasonable because we are accepting H_1 if $P(X/W_2)$ is sufficiently less than $P(X/W_1)$ which makes the likelihood ratio small enough.

Now let us consider a case in which regrets are equal and the prior probabilities are also equal then right hand side (r.h.s.) of (1.3.5) becomes 1. Thus H_1 is accepted if the likelihood of W_1 generating the sample $P(X/W_1)$ is greater than the likelihood of W_2 generating the sample $P(X/W_2)$, otherwise H_2 is accepted. We accept whichever hypothesis is more likely to generate the observed X as shown in (fig.1.3.1).



(fig. 1.3.1)

Now let us make assumption that the two density functions centered on W_1 and W_2 have the same symmetric and unimodel shape then (1.3.5) reduces to very reasonable criterion as shown in (fig. 1.3.2)



Accept H_1 iff X is observed closer to W_1 than W_2. This can be interpreted even for n observations. Let us assume that $\bar{X}$ based on

n observations drawn from population with variance σ^2, and an unknown mean of either μ_1 or μ_2 (between which we have to decide). Then $\bar{X}$ is approximately normal (by central limit theorem) with variance σ^2/n. Let $H_1 : \mu = \mu_1$ $H_2 : \mu = \mu_2$.
So criterion for accepting H_1 becomes

$$\frac{e^{-n/2\sigma^2 (\bar{X}-\mu_2)^2}}{e^{-n/2\sigma^2 (\bar{X}-\mu_1)^2}} < \frac{r_1 P(\mu_1)}{r_2 P(\mu_2)}$$

That is $\bar{X} < \dfrac{\mu_1+\mu_2}{2} + \dfrac{\sigma^2/n}{\mu_2-\mu_1} \log \left[\dfrac{r_1 P(\mu_1)}{r_2 P(\mu_2)} \right]$

By arranging

$$\bar{X} < \frac{\mu_2^2 - \mu_1^2}{2(\mu_2-\mu_1)} + \frac{\sigma^2/n}{(\mu_2-\mu_1)} \cdot k$$

where $k = \log \left[\dfrac{r_1 P(\mu_1)}{r_2 P(\mu_2)} \right]$

Assume that the regrets are equal and prior probabilities $P(\mu_i)$ are also equal. Then criterion for accepting H_1 reduces to accept H_1 if $\bar{X} < \dfrac{\mu_1+\mu_2}{2}$.
Since $\dfrac{\mu_1+\mu_2}{2}$ is the halfway point between μ_1 and μ_2 it is similar to the criterion for accepting H_1 that we have seen in preceding case.

Remark (1.3.1) :

Although Bayesian methods are more complicated than classical methods they are often satisfactory. A Bayesian test uses all the information [that is used in a classical test] and also exploits the prior distribution P(W) and the loss function. A classical test sets the level of significance at 5 % or 1 %, sometimes arbitrarily, sometimes with implicit reference to vague considerations of loss and loss belief. Bayesians would argue that these considerations should be introduced explicitely with all assumptions exposed, and open to criticism and improvement.

Example (1.3.1) :

Let X be a normal random variable with mean θ and variance σ_o^2 (known) and the prior density of θ be normal with mean μ and variance τ^2. Then the posterior distribution of θ is normal with mean $\hat{\theta}(x)$ and variance σ_1^2 where

$$\hat{\theta}(x) = \left(\frac{\mu}{\tau^2} + \frac{n\bar{x}}{\sigma_o^2} \right) \sigma_1^2$$

and $$\sigma_1^2 = \left(\frac{1}{\tau^2} + \frac{n}{\sigma_o^2} \right)^{-1}$$

when $x_1, x_2, \ldots, x_n$ is observed.

Consider the problem of testing H_1: $\theta \leq 0$ against H_2: $\theta > 0$. based on a single observation x. That is $W_1 = (-\infty, 0)$ and $W_2 = (0, \infty)$. From equation (1.3.3) the Bayes rule for the problem is accept H_1 if $R_1 \leq R_2$. The loss considered here assumes regrets $r_1 = \theta$ and $r_2 = -\theta$

Therefore
$$R_1 = \int_0^\infty r_1 . \, p(\theta/x) \, d\theta.$$
$$= \int_0^\infty \theta \, p(\theta/x) d\theta.$$

and $$R_2 = \int_{-\infty}^{0} (-\theta)\ p(\theta/x)d\theta .$$

We have, accept H_1 if $R_1 - R_2 < 0$.

$$R_1 - R_2 = \int_{-\infty}^{\infty} \theta \frac{1}{\sqrt{2\pi}\,\sigma_1} \cdot \exp - \frac{1}{2\sigma_1^2} (\theta - \hat{\theta}(x))^2\, d\theta.$$

$$= \hat{\theta}(x)$$

$$= \left(\frac{\mu}{\tau^2} + \frac{x}{\sigma_o^2} \right) \sigma_1^2$$

Therefore accept H_1 if

$$x < - \frac{\mu}{\tau^2} \cdot \sigma_o^2$$

observe as $\tau^2 \rightarrow \infty$ the prior distribution tends to a uniform distribution (uniform over $-\infty, \infty$). Thus an expected Bayes rule will be, accept H_1 if $X < 0$. This procedure can be represented in the form of classical test procedure as

$$\phi(x) = \begin{cases} 1, & \text{if } x \leq 0 \\ 0, & \text{if } x > 0 \end{cases}$$

Example (1.3.2) :

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with an unknown value of the mean θ and an unknown value of variance $1/\sigma'$. Suppose also that the prior joint distribution of θ and $1/\sigma'$ is as follows : The conditional distribution of θ when $\sigma' = \sigma (\sigma > 0)$ is a normal distribution with mean μ and variance $1/\tau\sigma$ such that $-\infty < \mu < \infty$ and $\tau > 0$ and marginal distribution of $1/\sigma'$ is

gamma distribution with parameter α such that $\alpha > 0$.
Then the posterior joint distribution of θ and $1/\sigma'$ when $X_i = x_i$ $(i = 1,2,..,n)$ is as follows :
The conditional distribution of θ when $1/\sigma' = 1/\sigma$ is a normal distribution with mean μ' and variance $1/(\Upsilon+n)\sigma$ where $\mu' = \frac{\mu+n\bar{x}}{\Upsilon+n}$. And marginal distribution of $1/\sigma$ is a gamma distribution with parameter α. In particular $\Upsilon = 1$. The marginal posterior density of θ is given by

$$f(\theta/x) = \int_0^\infty \frac{\sqrt{(n+1)\sigma}}{\sqrt{2\pi}\,\Gamma\alpha} \cdot \exp - \frac{(1+n)\sigma}{2} (\theta-\mu')^2 .\ e^{-\sigma} .\ \sigma^{\alpha-1}\, d\sigma$$

$$f(\theta/x) = \frac{\sqrt{n+1}}{\sqrt{2\pi}\,\Gamma\alpha} \cdot \int_0^\infty e^{-\sigma[\frac{n+1}{2}.(\theta-\mu')^2+1]} .\ \sigma^{\alpha-\frac{1}{2}}\, d\sigma.$$

Take $S = \frac{n+1}{2}(\theta-\mu')^2+1$

Therefore $f(\theta/x) = \frac{\sqrt{n+1}}{\sqrt{2\pi}\,\Gamma\alpha} \int_0^\infty e^{-S\sigma} \cdot \sigma^{\alpha-\frac{1}{2}}\, d\sigma$.

Put $S\sigma = Z$

$S d\sigma = dZ$

$d\sigma = dZ/S$.

Therefore

$$f(\theta/x) = \frac{\sqrt{n+1}}{\sqrt{2\pi}\,\Gamma\alpha} \int_0^\infty e^{-Z} \left(\frac{Z}{S}\right)^{\alpha-\frac{1}{2}} \frac{dZ}{S}$$

$$= \frac{\sqrt{n+1}}{\sqrt{2\pi}\,\Gamma\alpha} \int_0^\infty e^{-Z}\, Z^{\alpha-\frac{1}{2}} \left(\frac{1}{S}\right)^{\alpha+\frac{1}{2}} dZ.$$

$$= \frac{\sqrt{n+1}}{\sqrt{2\pi}\,\Gamma\alpha} \int_{0}^{\infty} e^{-Z} Z^{\alpha+\frac{1}{2}-1} \, dZ . (1/S)^{\alpha+\frac{1}{2}} \, dZ.$$

$$= \frac{\sqrt{n+1}}{\sqrt{2\pi}\,\Gamma\alpha} \Gamma\left(\alpha + \frac{1}{2}\right) (1/S)^{\alpha+\frac{1}{2}}$$

$$= \frac{\sqrt{n+1}}{\sqrt{2\pi}\,\Gamma\alpha} \frac{\Gamma(2\alpha+1)}{2} \left[\frac{n+1}{2} (\theta - \mu')^2 + 1 \right]^{-\left(\frac{2\alpha+1}{2}\right)}$$

$$= \frac{\sqrt{n+1}}{\sqrt{2\pi}\,\Gamma\alpha} \frac{\Gamma(2\alpha+1)}{2} \left[1 + \frac{\alpha(n+1)(\theta-\mu')^2}{2\alpha}\right]^{-\left(\frac{2\alpha+1}{2}\right)}$$

Therefore $f(\theta/x)$ follows (t distribution with 2α degrees of freedom (d.f.), location parameter μ' and scale parameter $\frac{1}{\alpha(n+1)}$.

Consider the problem of testing H_1: $\theta \leq 0$ against H_2: $\theta > 0$ based on n observations. The regrets given are $r_1 = \theta$ and $r_2 = -\theta$.

Following the notations in equation (1.3.2).

$$R_1 = \int_{0}^{\infty} \theta + f(\theta/x) \, d\theta \quad \text{and}$$

$$R_2 = \int_{-\infty}^{0} (-\theta) \, f(\theta/x) \, d\theta.$$

Accept H_1 iff $R_1 < R_2$

$$R_1 - R_2 = \mu' = \frac{\mu+n\bar{x}}{1+n}$$

Therefore accept H_1 iff

$$\frac{\mu+n\bar{x}}{1+n} < 0.$$

That is $\bar{x} < -\mu/n$.

Example (1.3.3) :

Let $X \sim B_i(k, \theta)$, k(known) prior distribution of θ is Beta (a,b). Then posterior distribution of θ is Beta (a+T, b+nk−T) for a sample $x_1, x_2, \ldots, x_n$ where $T = \sum_1^n X_i$

Consider the problem of testing of hypothesis

$H_1 : \theta < 1/2$ against $H_2 : \theta \geq 1/2$ where the regrets $r_1 = \theta - 1/2$ (for $\theta > 1/2$) and $r_2 = 1/2 - \theta$ (for $\theta < 1/2$).

Following the notations of equation (1.3.2) and (1.3.3) we get

$$R_1 = \int_{1/2}^{1} (\theta - 1/2)\, \beta(a+T, b+nk-T)\, d\theta.$$

$$R_2 = (-1)\int_{0}^{1/2} (\theta - 1/2)\, \beta(a+T, b+nk-T)\, d\theta.$$

Consider

$$R_1 - R_2 = \int_{0}^{1} (\theta - 1/2)\, \beta\, (a+T, b+nk-T)\, d\theta.$$

$$= \frac{a+T}{a+b+nk} - 1/2.$$

We have by equation (1.3.3) accept H_1 iff $R_1 - R_2 < 0$

That is $\frac{a+T}{a+b+nk} - 1/2 < 0$

gives $T < \frac{b+nk-a}{2}$

For a = b $T < \frac{nk}{2}$.

Mathematical Result :

Suppose f is symmetric about m and unimodal.

Consider $L_2(y) \downarrow$

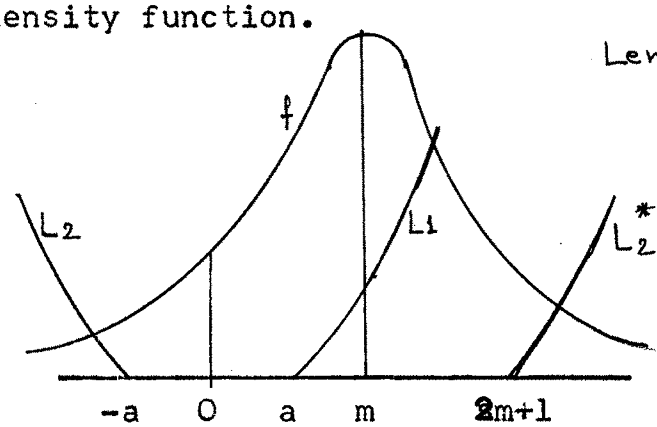
$L_2(y) = 0$ for $y \geq -a$ and

$L_1(y) = L_2(-y)$ with $a = 0$ consider a problem of testing $H_1 : m \leq 0$ against $H_2 : m > 0$.

In this case L_1 can be interepreted as a loss in accepting H_1 and L_2 a loss in accepting H_2, where f as posterior density function.

Lemma (i) $R_2 \leq R_1$ if $m \geq 0$

(ii) $R_2 \geq R_1$ if $m \leq 0$.



$$R_2 = \int_{-\infty}^{\infty} L_2(y)\, f(y)\, dy.$$

$L_2^*(m+y) = L_2(m-y)$ for all y,

put $y = m-x$.

Therefore

$$R_2 = \int_{-\infty}^{\infty} L_2^*(m+x)\, f(m+x)\, dx \quad \text{since } f(m+x) = f(m-x)$$

Put $m+x = t$

Therefore $R_2 = \int_{-\infty}^{\infty} L_2^*(t)\, f(t)\, dt.$

We have $L_2^*(t) \leq L_1(t)$ for all t, $m \geq 0$.

Therefore $\int_{-\infty}^{\infty} L_2^*(t) f(t)\, dt \leq \int_{-\infty}^{\infty} L_1(t) f(t)\, dt$.

That is $R_2 \leq R_1$.

Similarly it can be proved for case (ii)

In example (1.3.1) and (1.3.2) the lemma can be applied directly, since the posterior density is normal and t which is symmetric and the terms R_1 and R_2 are nothing but risks in accepting H_1 and H_2.