CHAPTER-III

INADMISSIBILITY OF STANDARD ESTIMATES AND CONFIDENCE INTERVALS.

3.0 Introduction :

In chapter I we have seen that a common formulation of statistical decision theory involves a sample space and a class of probability distributions P_Θ where Θ is a parameter. The loss depends on a action to solve the problem and is a function of the parameter Θ, Thus to choose a proper action we must'know about Θ'. Ordinarily the parameter is not considered to be a random. If this will be the case, then we shall choose a decision rule so that the risk from assuming its worst possible value. In many problems, certain extreme values of the parameter are not allowed completely but the experimenter may be allow such values of the parameter for to re-formulate its recommendations to suit their judgement. Some statisticians accept this but some are rejected, such type of different judgements there will be a incentive power to reformulate the standard decision procedure so that there may be exact use of prior information with more efficient. If nothing is known about the distribution of parameter we can do as usual formulation. In this chapter we discuss some standard estimates cealing with the admissibility in the presence of prior information.

In the next two subsections we show that certain moment conditions are essential for the admissibility of various 'good' statistical procedures which are translation invariant. Backwell (1951) first gave an example in which he proved that a best invariant decision rule may be inadmissible. After this many papers have been published dealing with the admissibility of the best invariant procedure. In case of point estimation, the problem of admissibility of a location parameter was considered by Blyth (1951), Blackwell (1951), Stein (1956), Farrell (1964), Brown (1966). The problem of admissibility of certain confidence intervals was treated by Joshi (1966). In each of the above paper, the admissibility requires the existence of one more moment than what is needed for finite risk. In the following we can show that without this extra moment inadmissibility may result.

3.1 Inadmissibility of some standard estimates :

In binomial distribution we take into account the prior information about Θ, the probability of success. We are to estimate this probability from an observation on the number X of successes in n trials. The conventional estimate is $\frac{X}{n}$. We assume that the probability of success in our trials is the value of a random variable Ⓗ. Note that if nothing is known about the distribution of Ⓗ, we can do no better than the usual formulation. However, we assume that Ⓗ has a distribution which belongs to a subclass of the distributions on [0,1] as a prior information to the experimenter. If we now regard the binomial distribution to be conditional on Ⓗ, the members of this subclass generate a family of joint distributions for X and Ⓗ. With this as background we may view our problem as a special case of conventional prediction theory. In this section we discuss the results of Skibinsky and Cote (1964). In the following we present a generalized maximum likelihood principle as applied to this example and inuestigate a class of predictors which it suggests. Under appropriate conditions to be discussed below each of these has a uniformly smaller mean square error than the conuentional estimate.

Let n, ∂, α be given, n is a positive integer, $0 < \partial \leq \frac{1}{2}$, $0 < \alpha < 1$.

The random variables $\textcircled{H}$, X the former distributed on the unit interval and the latter distributed discretely over the numbers 0,1,2,..,n.

We suppose that

Prob (X = x / $\textcircled{H}$ = θ) = $f(x, \theta)$ (1)

where

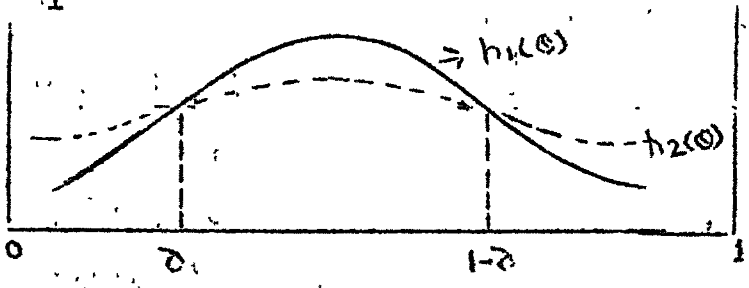
$$f(x,\theta)= \begin{cases} \binom{n}{x} \theta^x (1-\theta)^{n-x}, & x = 0,1,2,..,n,\ 0<\theta<1 \\ 1 \text{ or } 0 \text{ , according as} & x=0 \text{ or } x>0,\ \theta=0 \\ & x=n \text{ or } x<n,\ \theta=1 \end{cases}$$

is the value at x of the binomial frequency function with parameters n and θ (n is known). Let v be a c.d.f. on the unit interval. We shall write P_v to indicate any probability measure on the domain of $\textcircled{H}$ and X which satisfies (1) and has v as marginal c.d.f. for $\textcircled{H}$; and E_v for expectation relative to P_v . Let

$$m(\partial, \alpha) = [v : v(1 - \partial) - v(\partial - 0) \geq 1 - \alpha]$$

That is, $m(\partial, \alpha)$ is the class of priors whose concentration on $(\partial, 1-\partial)$ is greater than $1-\alpha$. e.g. In the following figure, Let $h_1(\theta)$ the pdf corresponding to prior distribution be such that, the area under the curve from ∂ to $1-\partial$ is 0.7 (say). For a different prior corresponding to the pd.f. $h_2(\theta)$ the corresponding area be 0.5 (say).

Let H_i be the distribution function corresponding to the pdf h_i; and assume that $1-\alpha = 0.6$ (say).



clearly $H_1(\Theta) \in m(\partial, \alpha)$,

$H_1(\Theta)$ is the cdf corresponding to $h_1(\Theta)$ and

$H_2(\Theta) \notin m(\partial, \alpha)$

<u>Theorem (3.1.1)</u>:

Let v belong to $m(\partial,\alpha)$, then for $\alpha>0$ and sufficiently small, $\frac{X}{n}$ is an inadmissible predictor of Ⓗ relative to the squared difference loss function, in the sense that there exists a predictor which is uniformly better over $m(\partial,\alpha)$. In fact there exists a mapping ξ from the range of X to the unit interval such that

$$E_v [(\xi(X) - Ⓗ)^2] < E_v [(\frac{X}{n} - Ⓗ)^2],$$

for all $v \in m(\partial, \alpha)$.

<u>Proof</u>:

The proof of this theorem consists of three stages.

(I) <u>A maximum likelihood method for prediction of Ⓗ</u> :

We proceed in two steps :

<u>Step</u> (i) : We choose corresponding to each x, a c.d.f. $v_x \in m(\partial, \alpha)$ such that

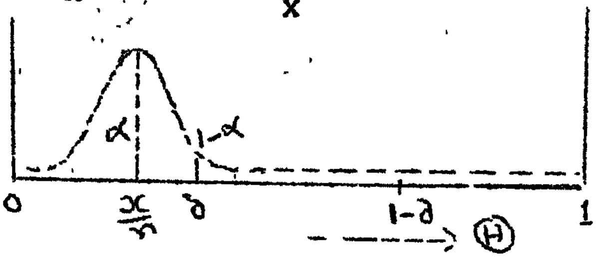
$$P_{v_x} (X = x) \geq P_v (X = x) \text{ , all } v \in m(\partial, \alpha) \qquad (2)$$

By definition of P_v and (1), we have for each c.d.f. v on [0,1] that

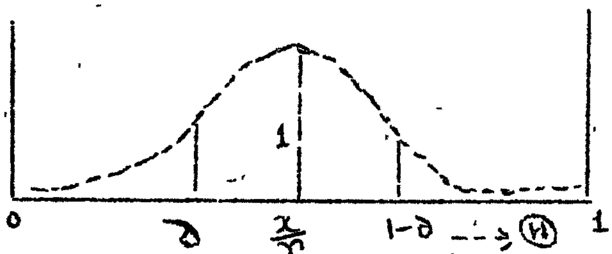
$$P_v\ (X = x) = E_v(f(x, \textcircled{H})),\ x = 0,1,2,\ldots,n.$$

But for each x, the likelihood function f(x,.) is strictly monotone on [0,1] to each side of a unique maximum at $\Theta = \frac{x}{n}$. Hence v_x defined by,

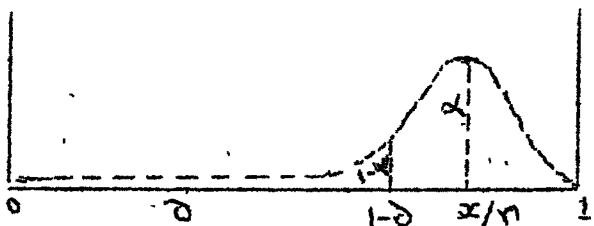
$$P_{v_x}(\textcircled{H} = \tfrac{x}{n}) = \alpha,\ P_{v_x}(\textcircled{H} = \partial) = 1-\alpha,\ x \leq n\sigma$$



$$P_{v_x}(\textcircled{H} = \tfrac{x}{n}) = 1,\quad n\partial \leq x \leq n - n\partial$$



$$P_{v_x}(\textcircled{H} = 1-\partial) = 1-\alpha,\ P_{v_x}(\textcircled{H} = \tfrac{x}{n}) = \alpha,\ x \geq n-n\partial$$



satisfies (2) uniquely for x = 0,1,2,...,n.

<u>Step</u> (ii) : We obtain corresponding to each x, a value of $\textcircled{H}$, we shall call it $\hat{\Theta}_{\partial,\alpha}(x)$, whose a posterior proba-

-bility is maximum, i.e. a value Θ which maximizes the conditional probability

$$P_{v_x}(\textcircled{H} = \Theta / X=x) = \frac{P_{v_x}(\textcircled{H} = \Theta, X=x)}{P_{v_x}(X = x)} \qquad (3)$$

The numerator of the RHS of (3) can be written as

$$P_{v_x}(\textcircled{H} = \Theta, X=x) = P_{v_x}(\textcircled{H} = \Theta) \cdot f(x,\Theta), \qquad (4)$$

and this may be interpreted for each fixed x as an ' a posterior likelihood of Θ '. Clearly, maximizing this is for fixed x equivalent to maximizing (3). In view of the definition of v_x when $x < n\partial$,

$$\begin{aligned} P_{v_x}(\textcircled{H} = \Theta).f(x, \Theta) &= P_{v_x}(\textcircled{H} = \tfrac{x}{n}).f(x,\tfrac{x}{n}), & \Theta = \tfrac{x}{n} \\ &= \alpha\, f(x, \tfrac{x}{n}) & , \Theta = \tfrac{x}{n} \\ &= (1 - \alpha)\, f(x, \partial) & , \Theta = \partial \\ &= 0 & , \text{o.w.} \end{aligned}$$

Similarly, when $x > n - n\partial$ then,

$$\begin{aligned} P_{v_x}(\textcircled{H} = \Theta).\ f(x, \Theta) &= \alpha\, f(x, \tfrac{x}{n}), & \Theta = \tfrac{x}{n} \\ &= (1 - \alpha)\, f(x, 1 - \partial), & \Theta = 1 - \partial \\ &= 0 & , \text{o.w.} \end{aligned}$$

When $n\partial \le x \le n - n\partial$ then,

$$P_{v_x}(Ⓗ = \theta)\ f(x,\theta) = f(x, \tfrac{x}{n}) \quad , \quad \theta = \tfrac{x}{n}$$
$$= 0 \quad , \quad \text{o.w.}$$

Now let $\eta(x,\partial) = f(x, x/n) / f(x,\partial)$

define $\hat{\theta}_{\partial,\alpha}(x)$ as follows,

when $x < n\partial$

$$\hat{\theta}_{\partial,x}(x) = \frac{x}{n} \quad , \quad \eta(x,\partial) > \frac{1-\alpha}{\alpha}$$
$$= \partial \quad , \quad \eta(x,\partial) \le \frac{1-\alpha}{\alpha}$$

When $x > n - n\partial$

$$\hat{\theta}_{\partial\alpha}(x) = \frac{x}{n} \quad , \quad \eta(x, 1-\partial) > \frac{1-\alpha}{\alpha}$$
$$= 1 - \partial \quad , \quad \eta(x, 1-\partial) \le \frac{1-\alpha}{\alpha}$$

When $n\partial \le x \le n - n\partial$

$$\hat{\theta}_{\partial,\alpha}(x) = \frac{x}{n}$$

It is clear from the above description of (4) that for each x .

$$P_{v_x}(Ⓗ = \hat{\theta}_{\partial,\alpha}(x)/ X=x) \ge (P_{v_x}(Ⓗ = \theta/ X=x), \text{ all } \theta \quad ..(5)$$

To find a simple expression for $\hat{\theta}_{\partial,\alpha}(x)$, we proceed as follows.

Let b denotes the largest integer less than $n\partial$. It can be shown that for $x = 0,1,\ldots,b$, $\eta(x,\partial)$ is strictly decreasing in x and bounded below by 1. It follows that

$$\frac{1}{1+\eta(x,\partial)} = C(x, \partial), \text{ (say)} \qquad (6)$$

is strictly increasing for these x, with $C(b, \eth) \leq \frac{1}{2}$.

By the above discussion and simple consideration of symmetry,

$$\hat{\theta}_{\eth,\alpha}(x) = \frac{x}{n}, \text{ when } \alpha > C(b,\eth).$$

For $\alpha \leq C(b,\eth)$, if we define a to be the smallest non-negative integer x such that $\eta(x,\eth) \leq 1-\alpha / \alpha$ (or equivalently, such that $C(x,\eth) \geq \alpha$), then

$$\hat{\theta}_{\eth,\alpha}(x) = \begin{cases} \eth, & a \leq x \leq b \\ 1 - \eth & n-b \leq x \leq n-a \\ \frac{x}{n} & \text{o.w.} \end{cases} \qquad (7)$$

$\hat{\theta}_{\eth,\alpha}$ is uniquely optimum in the sense of (5) unless $C(x,\eth) = \alpha$, for some $x \leq b$, in which case it may be modified at x, by replacing its value $\eth$, there with $\frac{x}{n}$ without affecting the value of the left hand side of (5), Thus $\hat{\theta}_{\eth,\alpha}(X)$ is optimal as a predictor of Ⓗ relative to the class $m(\eth, \alpha)$ is given by (5) and (2).

(II) <u>A class of predictors for Ⓗ</u> :

We consider the following class of predictors for Ⓗ which are suggested by the maximum likelihood predictor $\hat{\theta}_{\eth,\alpha}(X)$. Define ξ_j for $j = 0,1,2,\ldots,b$ on the range of X by

$$\xi_j(X) = \begin{cases} \eth & , j \leq x \leq b \\ 1 - \eth & , n-b \leq x \leq n-j \\ \frac{x}{n} & , \text{o.w.} \end{cases} \qquad (8)$$

The relationship of $\hat{\theta}_{\eth,\alpha}$ to the ξ_j follows directly from the definition of a. Indeed for $j = 0, 1, .., b$, we have

$$\hat{\theta}_{\partial,\alpha}(x) \equiv \xi_j(x), \text{ when } C(j-1,\partial) < \alpha \leq C(j,\partial) \quad (9)$$

We take $C(-1,\partial) = 0$.

(III) Now we shall compare $\xi_j(X)$ with X/n as a predictor of Ⓗ, we examine for $v \in m(\partial,\alpha)$ the difference,

$$E_v [(\xi_j(X) - Ⓗ)^2] - E_v[\frac{X}{n} - Ⓗ)^2] = E_v H_j (Ⓗ) \quad (10)$$

where H_j(Ⓗ) is the conditional expectation given, Ⓗ of the difference between the two squares. It can be shown that

$$H_j(\theta) = h_j(\theta) + h_j(1 - \theta), \text{ where}$$

$$h_j(\theta) = \sum_{x=j}^{b} (\partial - \frac{x}{n}) (\partial + \frac{x}{n} - 2\theta) f(x, \theta).$$

It can be seen that H_j are polynomials in θ each of which is symmetric about $\theta = \frac{1}{2}$. Also for $j = 0,1,2,..,b$, $H_j(\theta) < 0$ when $\partial \leq \theta \leq 1-\partial$ and $H_j(\theta) > 0$ for sufficiently small $\theta > 0$, also note that $H_0(0) = \partial^2$ and $H_j(\theta) = 0$, $j = 1,2,..,b$.

Thus using the above results, the largest value attained by (10) for any $v \in m(\partial,\alpha)$ is

$$\alpha \max_{0 \leq \theta < \partial} H_j(\theta) + (1-\alpha) \max_{\sigma \leq \theta \leq 1/2} H_j(\theta) \quad (11)$$

The first term in (11) is positive and second term is negative. For sufficiently small $\alpha > 0$, (11) is negative.

Thus

$E_V\ [(\xi_j(X) - Ⓗ)^2] < E_V[(\frac{x}{n} - Ⓗ)^2], v \in m(\partial, \alpha)$. Clearly, any one of the predictors $\hat{\xi}_j(X)$ is uniformly better over $m(\partial,\alpha)$ than the standard estimate $\frac{x}{n}$ relative to the squared difference loss function provided that α is sufficiently small. Hence $\frac{X}{n}$ is inadmissible predictor of Ⓗ .

□

3.2 Inadmissibility of the best invariant estimate of a location parameter

In this section we shall show that a unique best translation invariant estimate may be inadmissible if a certain moment condition fails to be satisfied.

Let the loss function be is given by

$L(\theta, a) = \omega(\theta\ a) = |\theta - a|^k, k \geq 1$

According to Brown's (1966) theorem, a unique best invariant estimate is admissible if the following moment condition is satisfied.

$$E\ |X|^{\alpha}\ W\ (X) < \infty \qquad \text{for } \alpha = 1 \qquad (1)$$

It is interesting to see whether this is the weakest moment condition we can have. Brown (1966) gave a partial answer to this question by giving an example. He gave a probability density function such that (1) is valid for $0 < \alpha < \frac{k}{2^k - 1}$. But the unique best invariant estimate is inadmissible. Now question is, a unique best invariant estimate is admissible if the moment condition (1) satisfied for $\frac{k}{2^k-1} < \alpha < 1$. We answer this question by the following example in the form of theorem due to Perng (1970).

<u>Theorem (3.2.1)</u> :

In the fixed sample size case, if the loss function is $W(t) = |t|^k$ for $k > 1$ then for every α ($0 \leq \alpha < 1$) there exists a family of probability densities such that $E|X|^{\alpha}W(X)<\infty$ and the best invariant estimate of the real location parameter is unique but it is inadmissible.

<u>Proof</u>:

Let θ be an unknown real parameter $-\infty < \theta < \infty$, Y be a random variable according to the known distribution G such that

$$\begin{aligned} dG(y) &= \frac{C}{y^{k+2-\eta}}\, dy , \quad y > 1 \\ &= 0 , \quad \text{o.w.} \end{aligned} \qquad (2)$$

where η, C are positive constants and $\eta < 1$. Assume that X given y is distributed according to $F(x-\theta/Y)$ where

$$\begin{aligned} dF(x-\theta/y) &= \frac{1}{y} \cdot \frac{1}{2b}\, dx , \text{ for } \left|\frac{x-\theta}{y}\right| \leq b \\ &= 0 , \quad \text{otherwise} \end{aligned}$$

and b is a positive constant.

It can be seen that the unique best invariant estimate of Θ is X. Also it can be seen to be $E\,|X|^{\alpha}\,W(X) < \infty$ for $0 \le \alpha < 1-\eta$. Now we shall show that X is inadmissible.

Consider the estimate of the form

$$\phi(x, y) = y\,\psi\left(\frac{x}{y}\right) \tag{3}$$

where

$$\psi\left(\frac{x}{y}\right) = \frac{x}{y} + f\left(\frac{x}{y}\right) \tag{4}$$

$$\begin{aligned} f(z) &= -\epsilon\,\partial z \quad \text{if} \quad |z| \le \frac{1}{\epsilon} \\ &= 0 \quad , \quad \text{otherwise} \end{aligned} \tag{5}$$

ϵ, ∂ are constants such that $0 < \partial < \epsilon < \frac{1}{b}$.

Then the risk of ϕ is

$$R(\phi, \theta) = E[|\phi - \theta|^{k}]$$

$$= \int_{1}^{\infty}\int_{-by+\theta}^{\infty} |\phi - \theta|^{k} \frac{C}{y^{k+2-\eta}} \frac{1}{2by}\, dx\, dy$$

$$= \int_{1}^{\infty}\int_{-by+\theta}^{by+\theta} \left|y\left[\frac{x}{y} + f\left(\frac{x}{y}\right)\right] - \theta\right|^{k} \frac{C}{2b\, y^{k+3-\eta}}\, dx\, dy$$

$$= \int_{1}^{\infty}\int_{-by+\theta}^{by+\theta} y^{k} \left|\psi\left(\frac{x}{y}\right) - \frac{\theta}{y}\right|^{k} \frac{C}{2b\, y^{k+3-\eta}}\, dx\, dy \tag{6}$$

Let $\frac{x}{y} = z$ and $\frac{\theta}{y} = \tau$ Thus for $\theta > 0$, then (6) becomes

$$R(\phi,\theta) = \frac{C\theta^{\eta-1}}{2b} \int_{0}^{\theta} \left[\int_{\tau-b}^{\tau+b} |\psi(z) - \tau|^{k}\, dz\right] \tau^{-\eta}\, d\tau \tag{7}$$

By (4) replace $\psi(z)$ by $z + f(z)$ and $z - \tau = \omega$. We have

$$R(\phi;\theta) = \frac{C\theta^{\eta-1}}{2b} \int_{0}^{\theta} \left[\int_{-b}^{b} |f(\omega+\tau)+\omega|^{k}\, d\omega\right] \tau^{-\eta}\, d\tau \tag{8}$$

Now we will evaluate the inner integral in (8).

For $k > 1$, $|w| \le b$ and $|f| \le \partial < \infty$.

consider

$$|w + f(w+\tau)|^k = |w|^k + f(w+\tau)\, k\, |w|^{k-1} \operatorname{sgn} w + O(f(w+\tau)) \qquad ..(9)$$

Hence

$$\int_{-b}^{b} |f(w+\tau)+w|^k \, dw =$$

$$\doteq \int_{-b}^{b} |w|^k \, dw + \int_{-b}^{b} f(w+\tau)\, k\, |w|^{k-1} \operatorname{sgn} w \, dw + + \int_{-b}^{b} O\,(f(w+\tau)\, dw$$

$$= \frac{2b^{k+1}}{k+1} + k \int_{-b}^{b} f(w+\tau)\, |w|^{k-1} (\operatorname{sgn} w) dw + + O\,(\sup_{|w|\le b} |f(w+\tau)|) \qquad (10)$$

Therefore (8) becomes

$$R(\emptyset,\theta) = \frac{C\theta^{\eta-1}}{2b} \int_0^{\theta} [\, \frac{2b^{k+1}}{k+1} + k \int_{-b}^{b} f(w+\tau)|w|^{k-1} (\operatorname{sgn} w)\, dw + + O\,(\sup_{|w|\le b} |f(w+\tau)|)]\, \tau^{-\eta}\, d\tau$$

$$= \frac{Cb^k}{(k+1)(1-\eta)} + \frac{C\theta^{\eta-1}}{2b} \left\{ k \int_0^{\theta} \tau^{-\eta}\, d\tau \int_{-b}^{b} f(w+\tau)|w|^{k-1}(\operatorname{sgn} w)\, dw + \int_0^{\theta} O\,(\sup_{|w|\le b}|f(w+\tau)|)\, \tau^{-\eta}\, d\tau \right. \qquad ..(11)$$

It has been shown that the risk of $\emptyset$ is

$$R(\emptyset,\theta) \le \frac{Cb^k}{(k+1)(1-\eta)} + \frac{C\theta^{\eta-1}}{2b} \left\{ \frac{-2\epsilon\, \partial\, \theta^{1-\eta} b^{k+1}}{(1-\eta)} + O(\epsilon\,\partial\theta) \right\}$$

$$\text{if } 0 < \theta < \epsilon^{-1} - b \qquad (12)$$

and

$$R(\phi,\theta) \leq \frac{Cb^k}{(k+1)(1-\eta)} + \frac{C\theta^{\eta-1}}{2b}\left\{\frac{-2\epsilon\partial\eta b^k}{1-\eta}[\min(\epsilon^{-1}-b,\theta)]^{1-\eta} + O(\epsilon^{\eta})\epsilon\,\partial + O\left(\min\left(\frac{\partial}{\epsilon},\ \partial\epsilon\theta\right)\right)\right\}$$

$$\text{if } \theta \geq \epsilon^{-1} - b \qquad (13)$$

Hence by choosing ϵ sufficiently small and then choosing ∂ so that $\frac{\partial}{\epsilon}$ is sufficiently small, then we have

$$R(\phi,\theta) < \frac{Cb^k}{(k+1)(1-\eta)} \quad \text{for all } \theta > 0. \qquad (14)$$

By the symmetry of the problem, we can show (14) still holds for $\theta < 0$. For $\theta = 0$ we can prove (14) by direct computation. Thus for fixed $\eta \in (0,1)$, $k > 1$, $0 < \partial < \epsilon < b^{-1}$ and ϵ, $\frac{\partial}{\epsilon}$ sufficiently small we have,

$$R(\phi,\theta) < \frac{Cb^k}{(k+1)\ (1-\eta)}, \quad -\infty < \theta < \infty \qquad (15)$$

It can be seen that the risk of the best invariant estimate X of θ is,

$$R(X,\ \theta) = \frac{Cb^k}{(k+1)\ (1-\eta)} \qquad (16)$$

Hence we have $R(\phi,\ \theta) < R(X,\ \theta)$, $-\infty < \theta < \infty$

This shows that a unique best invariant estimate X is inadmissible. □

3.3 Admissibility of Confidence intervals :

In section 2 , we have seen the admissibility of estimators of a location parameter in continuous frequency functions. Here the analogous question is considered regarding confidence intervals. The admissiblity of confidence intervals is proved for the location parameter in a wide class of continuous frequency functions which includes the normal and some other commonly occuring ones. An application of the result is that the usual symmetrical confidence intervals for the mean of a normal population are seen to be admissible whether the population variance is known or not. In the following, X denotes a real random variable, with a density function involving a parameter Θ which assumes values in a set Θ of the real line.

Let $x_1, x_2, \ldots, x_n$ be independent observations of X, and $X = (x_1, \ldots, x_n)$ a point in the sample space $\mathcal{X}$.

Lebesgue measure is defined on $\mathcal{X}$ and Θ. Let a(x), b(x) denote measurable functions defined on $\mathcal{X}$ and (a(x), b(x)) denotes the set of confidence intervals $[a(x) \leq \Theta \leq b(x)]$. We define admissibility of confidence intervals as below:

Definition (3.3.1) :

A set of confidence intervals [a(x), b(x)] is said to be admissible if and only if, there exists no other set of confidence intervals $(a_1(x), b_1(x))$ satisfying,

(i) $b_1(x)-a_1(x) \leq b(x) - a(x)$ for almost all $x \in \mathfrak{X}$ and

ii) $P\,(a_1(x) \leq \theta \leq b_1(x)/\theta) \geq P(a(x) \leq \theta \leq b(x)/\theta)$

for all $\theta \in Ⓗ$

for all $\theta \in Ⓗ$ where Ⓗ is the parameter space, the strict inequality in (ii) holding for at least one $\theta \in Ⓗ$.

Theorem (3.3.1) :

If $x_1, x_2, \ldots, x_n$ are independent observations from a known frequency function $f(x,\theta)$ containing an unknown parameter θ, $-\infty < \theta < +\infty$, and if,

(a) $f(x,\theta)$ admits a sufficient statistic $T(x)$ for θ, where $T(x)$ is a function of $x_1, x_2, \ldots, x_n$ with a frequency function of the form $P(T - \theta)$ i.e. the distribution of $(T - \theta)$ given θ is independent of θ for $-\infty < \theta < +\infty$.

(b) the frequency function $p(t)$ in (a) strictly decreases for $t \geq 0$ as t increases, and for $t \leq 0$ as t decreases, is continuous for all t and is such that

$$\int_0^\infty \left[\int_{t_1}^\infty P(t)\, dt + \int_{-\infty}^{-t_1} P(t)\, dt \right] dt_1$$

converges

(c) the frequency function $f(x,\theta)$ of x is positive (>0) and continuous in x for all $x = (x_1, x_2, \ldots, x_n)$ and all θ, $-\infty < \theta < +\infty$.

(d) $v_1(x), v_2(x)$ are non-negative statistics distributed independently of T and θ such that for every $x \in \mathfrak{X}$ and $P(t)$ in (a),

$$P(-v_2(x)) = P(v_1(x))$$

and further such that $v(x) = \max(v_1(x), v_2(x))$ has finite expectation and variance, then the confidence interval for θ :

$[T - v_1(x) \leq \theta \leq T + v_2(x)]$ are admissible according to the definition (3.3.1).

For a proof refer Joshi (1966).

Example (3.3.1):

Let $X_1, X_2, \ldots, X_n$ be iid $N(\theta,1)$. In this case $\bar{X}$ is sufficient and the distribution of which is normal with mean θ and variance $\frac{1}{n}$. It is easy to verify the above conditions (a), (b) and (c). We need to verify the condition (d). Since the density P(.) is symmetric about zero. Define

$$v_1(x) = ns^2 = v_2(x)$$

which are independently distributed of T and θ, and the distribution of ns^2 is χ^2_{n-1}.

Now $\quad v(x) = \max(v_1(x), v_2(x)) = ns^2$

$$\therefore \quad E\, v(x) = E(ns^2) = n-1$$

and $\operatorname{var} v(x) = \operatorname{var}(ns^2) = 2(n-1)$

hence the condition (d) is satisfied.

hence the confidence interval for θ :

$$[\bar{x} - ns^2 \leq \theta \leq \bar{x} + ns^2]$$

is admissible according to the theorem 3.3.1.. □

The above theorem 3.3.1 which gave a set of sufficient conditions for the admissibility of certain confidence interval procedures for a location parameter. Now we shall give the statement of a theorem due to Perng (1970).

Theorem (3.3.2):

Suppose X, Y have joint distribution given $h(x,y) = f(x-\theta/y)\,g(y)$.

Suppose $f(-v_2(y)/y) = f(v_1(y)/y)$, where $v_1(y)$, $v_2(y)$ are two non-negative statistics, and $f(t/y)$ is strictly decreasing in $|t|$ on the set $f(t/y) > 0$.

Suppose $\int g(y)\,dy \int |t|\, f(t/y)\,dt < \infty$ (1)

Then the confidence interval procedure given by $x - v_1(y) \leq \theta \leq x + v_2(y)$ is admissible according to the definition 3.3.1.

For a proof we refer to Perng (1970).

We have seen in section (2) of this chapter, that a unique best translation invariant estimate may be inadmissible if a certain moment condition fails to be satisfied. Now we notice that the moment condition (1) in theorem 3.3.2 is quite similar to the moment conditions in the estimation problem as in section (2). Here we shall see whether the moment condition is also essential for the admissibility of the specified confidence interval procedure; and these results are due to Perng (1970).

In the following we shall show the moment condition is necessary for the admissibility of confidence interval.

Theorem (3.3.3):

For every α $(0 \leq \alpha < 1)$ there is a family of probability density functions such that $E|X|^{\alpha} < \infty$, and a confidence interval procedure I such that pdf satisfy all but the moment condition (1), and I is inadmissible.

Proof:

Let Θ be an unknown real-valued parameter $-\infty<\Theta<\infty$, Y be a random variable according to the known density function.

$$g(y) = \frac{C_1}{y^{2-\eta}} \quad , \quad y > 1$$
$$= 0 \quad , \quad \text{otherwise} \qquad (2)$$

where $0 < \eta < 1$, $C_1 > 0$ are constants.

Let x given y have density function

$$P(x-\Theta/y) = \frac{C_2}{y} \left(b-1 \left|\frac{x-\Theta}{y}\right|\right), \text{if } \left|\frac{x-\Theta}{y}\right| \leq b$$
$$= 0 \qquad , \text{ otherwise} \qquad (3)$$

where C_2, b, l are proper positive constants and $b > 2$.

We define,

$$I(x,y) = [x-y,\ x+y] \qquad (4)$$

$$I^*(x,y) = I(x,y) \qquad \text{if } y < \epsilon|x|+1$$
$$= [x(1-\epsilon\partial)-y,\ x(1-\epsilon\partial)+y],$$
$$\text{if } y \geq \epsilon|x|+1 \qquad (5)$$

Where ϵ, ∂ are constants such that $0<\partial<\epsilon < \frac{1}{b}$.

It can be seen to be $E|X|^{\alpha} < \infty$ for $0 \leq \alpha < 1-\eta$ and diverges if $\alpha = 1-\eta$. Clearly, all but the moment condition (1) are satisfied.

Now we shall show that $I^*(x,y)$ dominates $I(x,y)$ in the sense of definition for sufficient small ϵ, ∂ and 1. Clearly the length of $I(x,y)$ is equal to the length of $I^*(x,y)$ for every (x,y).

Hence we need only to show that

$$P_{\theta}(\theta \in I(x,y)) \leq P_{\theta}(\theta \in I^*(x,y)) \quad (6)$$

for all θ and strict inequality holds for some θ. For $\theta = 0$ clearly (6) holds.

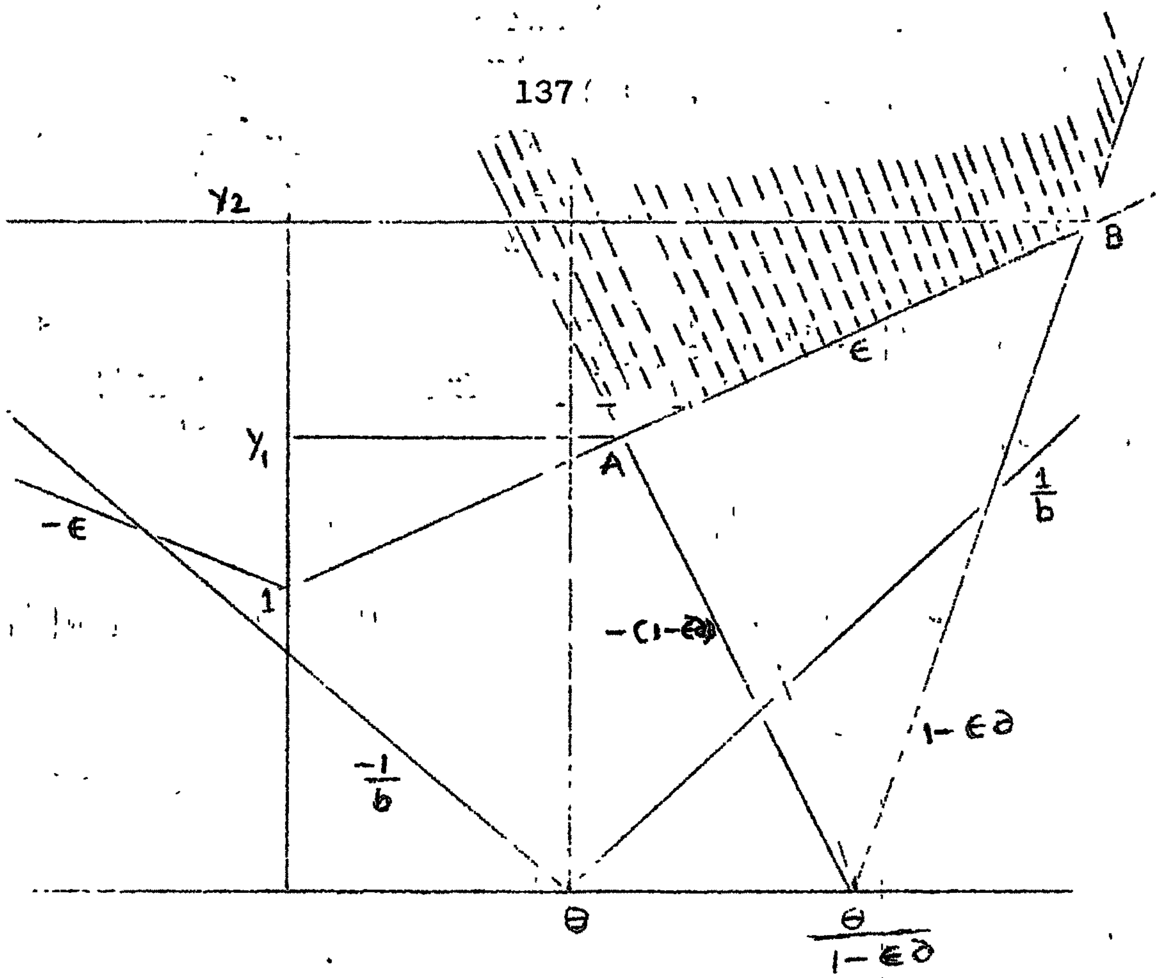
By the summetry of the problem, we need only to consider the case $\theta > 0$. That is we wish to show

$$P_{\theta}(\theta \in I(x,y)) < P_{\theta}(\theta \in I^*(x,y)), \text{ for all } \theta > 0 \quad (7)$$

Showing (7) is equivalent to showing

$$P_{\theta}(\theta \in I(x,y) \text{ and } y \geq \epsilon|x|+1) <$$
$$< P_{\theta}(\theta \in I^*(x,y) \text{ and } y \geq \epsilon|x|+1) \text{ for all } \theta > 0 \quad ..(8)$$

Now we shall evaluate the two probabilities in (8).



$$P_\theta(\theta \in I^*(x,y) \text{ and } y \geq \epsilon|x|+1) = \int_{\frac{1+\epsilon\theta-\epsilon\partial}{1+\epsilon-\epsilon\partial}}^{\frac{1+\epsilon\theta-\epsilon\partial}{1-\epsilon-\epsilon\partial}} \int_{\frac{\theta-y}{1-\epsilon\partial}}^{\frac{y-1}{\epsilon}} \frac{C_1C_2}{y^{2-\eta}} \frac{1}{y} \left(b-1\left|\frac{x-\theta}{y}\right|\right) dx\, dy \; +$$

$$+ \int_{\frac{1+\epsilon\theta-\epsilon\partial}{1-\epsilon\partial-\epsilon}}^{\infty} \int_{\frac{\theta-y}{1-\epsilon\partial}}^{\frac{\theta+y}{1-\epsilon\partial}} \frac{C_1C_2}{y^{2-\eta}} \frac{1}{y} \left(b-1\left|\frac{x-\theta}{y}\right|\right) dx\, dy \qquad (9)$$

and

$$P_\theta(\theta \in I(x,y) \text{ and } y \geq \epsilon|x|+1) =$$

$$= \int_{\frac{1+\epsilon\theta}{1+\epsilon}}^{\frac{1+\epsilon\theta}{1-\epsilon}} \int_{\theta-y}^{\frac{y-1}{\epsilon}} \frac{C_1C_2}{y^{2-\eta}} \frac{1}{y} (b-1|\frac{x-\theta}{y}|)\, dx\, dy +$$

$$+ \int_{\frac{1+\epsilon\theta}{1-\epsilon}}^{\infty} \int_{\theta-y}^{\theta+y} \frac{C_1C_2}{y^{2-\eta}} \frac{1}{y} (b-1|\frac{x-\theta}{y}|)\, dx\, dy \qquad (10)$$

Using the dominated convergence theorem we have

$$\lim_{1\to 0} P_\theta (\theta \in I^*(x,y) \text{ and } y \geq \epsilon|x|+1) =$$

$$= \underbrace{\int_{\frac{1+\epsilon\theta-\epsilon\partial}{1+\epsilon-\epsilon\partial}}^{\frac{1+\epsilon\theta-\epsilon\partial}{1-\epsilon\partial-\epsilon}} \int_{\frac{\theta-y}{1-\epsilon\partial}}^{\frac{y-1}{\epsilon}} \frac{C_1C_2b}{y^{3-\eta}}\, dx\, dy}_{(I_1)} +$$

$$+ \underbrace{\int_{\frac{1+\epsilon\theta-\epsilon\partial}{1-\epsilon\partial-\epsilon}}^{\infty} \int_{\frac{\theta-y}{1-\epsilon\partial}}^{\frac{\theta+y}{1-\epsilon\partial}} \frac{C_1C_2b}{y^{3-\eta}}\, dx\, dy}_{(I_2)} . \qquad (11)$$

consider

$$I_2 = \int_{\frac{1+\epsilon\theta-\epsilon\partial}{1-\epsilon\partial-\epsilon}}^{\infty} \int_{\frac{\theta-y}{1-\epsilon\partial}}^{\frac{\theta+y}{1-\epsilon\partial}} \frac{C_1C_2b}{y^{3-\eta}}\, dx\, dy$$

$$= \frac{-2C_1C_2b}{(\eta-1)(1-\epsilon\partial)} (1+\epsilon\theta-\epsilon\partial)^{\eta-1}(1-\epsilon-\epsilon\partial)^{1-\eta}$$

and

$$I_1 = \int_{\frac{1+\epsilon\theta-\epsilon\partial}{1+\epsilon-\epsilon\partial}}^{\frac{1+\epsilon\theta-\epsilon\partial}{1-\epsilon\partial-\epsilon}} \int_{\frac{\theta-y}{1-\epsilon\partial}}^{\frac{y-1}{\epsilon}} \frac{C_1C_2b}{y^{3-\eta}}\, dx\, dy$$

That is

$$I_1 = \frac{C_1C_2b(1-\epsilon\partial+\epsilon)(1+\epsilon\theta-\epsilon\partial)^{\eta-1}}{\epsilon(1-\epsilon\partial)(\eta-1)}\left\{(1-\epsilon-\epsilon\partial)^{1-\eta}-(1+\epsilon-\epsilon\partial)^{1-\eta}\right\}$$

$$-\frac{C_1C_2b(1+\epsilon\theta-\epsilon\partial)^{\eta-1}}{\epsilon(1-\epsilon\partial)(\eta-2)}\left\{(1-\epsilon-\epsilon\partial)^{2-\eta}-(1+\epsilon-\epsilon\partial)^{2-\eta}\right\}$$

Therefore (9) becomes

$$\lim_{1\to 0} P_\theta(\theta\in I^*(x,y) \text{ and } y\geq\epsilon|x|+1) = I_1+I_2$$

$$= \frac{-2C_1C_2b}{(1-\epsilon\partial)(\eta-1)}\left(\frac{1+\epsilon\theta-\epsilon\partial}{1-\epsilon\partial-\epsilon}\right)^{\eta-1}+$$

$$+\frac{C_1C_2b(1-\epsilon\partial+\epsilon)(1+\epsilon\theta-\epsilon\partial)^{\eta-1}}{\epsilon(1-\epsilon\partial)(\eta-1)}\left\{(1-\epsilon-\epsilon\partial)^{1-\eta}-(1+\epsilon-\epsilon\partial)^{1-\eta}\right\}$$

$$-\frac{C_1C_2b(1+\epsilon\theta-\epsilon\partial)^{\eta-1}}{\epsilon(1-\epsilon\partial)(\eta-2)}\left\{(1-\epsilon-\epsilon\partial)^{2-\eta}-(1+\epsilon-\epsilon\partial)^{2-\eta}\right\}$$

$$= \frac{C_1C_2b(1+\epsilon\theta-\epsilon\partial)^{\eta-1}}{\epsilon(1-\epsilon\partial)(\eta-1)(\eta-2)}\Big\{-2\epsilon(\eta-2)(1-\epsilon\partial-\epsilon)^{1-\eta}+$$

$$+(1-\epsilon\partial+\epsilon)(1-\epsilon-\epsilon\partial)^{1-\eta}(\eta-2)-(1+\epsilon-\epsilon\partial)^{1-\eta}(1-\epsilon\partial+\epsilon)(\eta-2)$$

$$-(\eta-1)(1-\epsilon-\epsilon\partial)^{2-\eta}+(\eta-1)(1+\epsilon-\epsilon\partial)^{2-\eta}\Big\}$$

$$= \frac{C_1C_2b(1+\epsilon\theta-\epsilon\partial)^{\eta-1}}{\epsilon(1-\epsilon\partial)\ (\eta-1)(\eta-2)} \left\{(1+\epsilon-\epsilon\partial)^{2-\eta} - (1-\epsilon-\epsilon\partial)^{2-\eta}\right\}$$

$$= \frac{C_1C_2b\left\{(1+\epsilon-\epsilon\partial)^{2-\eta} - (1-\epsilon\partial-\epsilon)^{2-\eta}\right\}}{\epsilon(1-\epsilon\partial)\ (1-\eta)\ (2-\eta)\ (1+\epsilon\theta-\epsilon\partial)^{1-\eta}} \qquad (12)$$

Similarly,

$$\lim_{1\to 0} P_\theta(\theta \in I(x,y) \text{ and } y \geq \epsilon|x|+1) =$$

$$= \underbrace{\int_{\frac{1+\epsilon\theta}{1+\epsilon}}^{\frac{1+\epsilon\theta}{1-\epsilon}} \int_{\theta-y}^{\frac{y-1}{\epsilon}} \frac{C_1C_2b}{y^{3-\eta}}\, dx\, dy}_{(I_3)} + \underbrace{\int_{\frac{1+\epsilon\theta}{1-\epsilon}}^{\infty} \int_{\theta-y}^{\theta+y} \frac{C_1C_2b}{y^{3-\eta}}\, dx\, dy}_{(I_4)} \qquad ..(13)$$

Consider,

$$I_3 = \int_{\frac{1+\epsilon\theta}{1+\epsilon}}^{\frac{1+\epsilon\theta}{1-\epsilon}} \frac{C_1C_2b}{y^{3-\eta}} \left(\frac{y-1}{\epsilon} - (\theta-y)\right) dy$$

$$= \frac{b\,C_1C_2(1+\epsilon)}{\epsilon(\eta-1)(1+\epsilon\theta)^{1-\eta}} \left\{(1+\epsilon)^{1-\eta} - (1+\epsilon)^{1-\eta}\right\} -$$

$$- \frac{C_1C_2b(1+\epsilon\theta)}{\epsilon(\eta-2)(1+\epsilon\theta)^{2-\eta}} \left\{(1-\epsilon)^{2-\eta} - (1+\epsilon)^{2-\eta}\right\}$$

and

$$I_4 = \int_{\frac{1+\epsilon\theta}{1-\epsilon}}^{\infty} \frac{C_1C_2b}{y^{3-\eta}} (y+\theta-\theta+y)\, dy$$

$$= \frac{-2C_1C_2b}{\eta-1} \left(\frac{1+\epsilon\theta}{1-\epsilon}\right)^{\eta-1}$$

Therefore (13) becomes,

$$= \frac{C_1C_2b \left\{(1+\epsilon)^{2-\eta} - (1-\epsilon)^{2-\eta}\right\}}{\epsilon(1-\eta)\,(1+\epsilon\theta)^{1-\eta}\,(2-\eta)} \tag{14}$$

From (12) and (14) we have

$$\lim_{l\to 0} \Big\{ P_\theta(\theta \in I^*(x,y) \text{ and } y \geq \epsilon|x|+1) - P_\theta(\theta \in I(x,y) \text{ and } y \geq \epsilon|x|+1) \Big\}$$

$$\frac{C_1C_2b\left\{(1+\epsilon-\epsilon\partial)^{2-\eta} - (1-\epsilon\partial-\epsilon)^{2-\eta}\right\}}{\epsilon(1-\epsilon\partial)\,(1-\eta)\,(2-\eta)\,(1+\epsilon\theta-\epsilon\partial)^{1-\eta}} - \frac{C_1C_2b\left\{(1+\epsilon)^{2-\eta} - (1-\epsilon)^{2-\eta}\right\}}{\epsilon(1-\eta)\,(1+\epsilon\theta)^{1-\eta}\,(2-\eta)}$$

$$= \frac{C_1C_2b}{\epsilon(1-\epsilon\partial)(1-\eta)(2-\eta)\,(1+\epsilon\theta-\epsilon\partial)^{1-\eta}(1+\epsilon\theta)^{1-\eta}}$$

$$\Big\{(1+\epsilon\theta)^{1-\eta}(1+\epsilon-\epsilon\partial)^{2-\eta} - (1+\epsilon\theta)^{1-\eta}(1-\epsilon\partial-\epsilon)^{2-\eta} - (1-\epsilon\partial)(1+\epsilon\theta-\epsilon\partial)^{1-\eta}(1+\epsilon)^{2-\eta} + (1-\epsilon\partial)(1+\epsilon\theta-\epsilon\partial)^{1-\eta}(1-\epsilon)^{2-\eta}\Big\}$$

$$= \frac{C_1C_2b}{D}\Big\{[(1+\epsilon-\epsilon\partial)^{2-\eta} - (1-\epsilon-\epsilon\partial)^{2-\eta}](1+\epsilon\theta)^{1-\eta} - [(1+\epsilon)^{2-\eta} - (1-\epsilon)^{2-\eta}](1+\epsilon\theta-\epsilon\partial)^{1-\eta}(1-\epsilon\partial)\Big\} \tag{15}$$

where $D = \epsilon(1-\epsilon\partial)(1-\eta)(2-\eta)(1+\epsilon\theta-\epsilon\partial)^{1-\eta}\,(1+\epsilon\theta)^{1-\eta}$

We have to show that the RHS of (15) is positive for all θ, it is sufficient to show that the term in the braces of the numerator is positive for all $\theta > 0$.

Expand $(1+\epsilon-\epsilon\partial)^{2-\eta}$, $(1-\epsilon-\epsilon\partial)^{2-\eta}$ and $(1+\epsilon\theta-\epsilon\partial)^{1-\eta}$ by Taylor's expansion then we have,

$$\left.\begin{aligned}(1+\epsilon-\epsilon\partial)^{2-\eta} &= (1+\epsilon)^{2-\eta}-(2-\eta)(1+\epsilon)^{1-\eta}\epsilon\partial+O(\epsilon\partial)\\(1-\epsilon-\epsilon\partial)^{2-\eta} &= (1-\epsilon)^{2-\eta}-(2-\eta)(1-\epsilon)^{1-\eta}\epsilon\partial+O(\epsilon\partial)\\(1+\epsilon\theta-\epsilon\partial)^{1-\eta} &= (1+\epsilon\theta)^{1-\eta}-(1-\eta)(1+\epsilon\theta)^{-\eta}\epsilon\partial+O(\epsilon\partial)\end{aligned}\right\}\quad(16)$$

Using (16), the numerator of (15) can be written as

$$[(1+\epsilon-\epsilon\partial)^{2-\eta}-(1-\epsilon-\epsilon\partial)^{2-\eta}](1+\epsilon\theta)^{1-\eta}-$$
$$-[(1+\epsilon)^{2-\eta})-(1-\epsilon)^{2-\eta}](1+\epsilon\theta-\epsilon\partial)^{1-\eta}(1-\epsilon\partial)$$

$$\geq [-2(2-\eta)(1+\epsilon)^{1-\eta}\epsilon\partial+(2-\eta)(1-\epsilon-\epsilon\partial)^{1-\eta}\epsilon\partial](1+\epsilon\theta)^{1-\eta}+$$
$$+[(1+\epsilon)^{2-\eta}-(1-\epsilon)^{2-\eta}](1+\epsilon\theta)^{1-\eta}\epsilon\partial-O(\epsilon\partial)\quad ..\ (17)$$

Expand $(1+\epsilon)^{1-\eta}$, $(1-\epsilon-\epsilon\partial)^{1-\eta}$, $(1+\epsilon)^{2-\eta}$ and $(1-\epsilon)^{2-\eta}$ by Taylor's expansion then we have

$$\left.\begin{aligned}(1+\epsilon)^{1-\eta} &= 1+(1-\eta)+O(\epsilon)\\(1-\epsilon-\epsilon\partial)^{1-\eta} &= 1-(1-\eta)\epsilon(1-\partial)+O(\epsilon)\\(1+\epsilon)^{2-\eta} &= 1+(2-\eta)\epsilon+O(\epsilon)\\(1-\epsilon)^{2-\eta} &= 1-(2-\eta)\epsilon+O(\epsilon)\end{aligned}\right\}\quad(18)$$

Substituting (18) into (17), we have

$$[(1+\epsilon-\epsilon\partial)^{2-\eta}-(1-\epsilon-\epsilon\partial)^{2-\eta}](1+\epsilon\theta)^{1-\eta}-$$
$$-[(1+\epsilon)^{2-\eta}-(1-\epsilon)^{2-\eta}](1+\epsilon\theta-\epsilon\partial)^{1-\eta}(1-\epsilon\partial)$$

$$\geq \epsilon\, \partial\, (1+\epsilon\theta)^{1-\eta}\, (2-\eta)\, \epsilon \left\{2\eta-\partial(1-\eta) - O(\epsilon)\right\} \qquad (19)$$

For given $\eta>0$, it is possible to choose ϵ, ∂, sufficiently small so that RHS of (19) is positive for all $\theta > 0$.

Equivalently, for sufficiently small ϵ, ∂,

$$\lim_{l \to 0} [P(\theta \in I^*(x,y) \text{ and } y \geq \epsilon|x|+l) - P(\theta \in I(x,y) \text{ and } y \geq \epsilon|x|+l)] > 0 \quad .. (20)$$

for all $\theta > 0$.

(20) implies that there exists a positive l_o such that

$$P_\theta(\theta \in I^*(x,y) \text{ and } y \geq \epsilon|x|+l) > P_\theta(\theta \in I(x,y) \text{ and } y \geq \epsilon|x|+l)$$

for sufficiently small ϵ, ∂ and all $\theta > 0$.

But from (4) and (5) the lengths of I and I* are equal. Hence from the definition 3.3.1, I is inadmissible.

□