C H A P T E R III

DESCRIPTIVE STATISTICS FOR NON-PARAMETRIC MODELS

3.1 Introduction :

In this chapter we discuss the descriptive statistics for non-parametric models introduced by Bickel and Lehamman (1975, 1976). In this chapter we discuss the following points

1) Non-parametric neighbourhood model.

2) Measure of location.

3) Measure of dispersion (for symmetric models).

3.2 Non-parametric neighbourhood models :

For a given observations an attempts are made to propose parametric models. In model fitting some times one can model itself. It might due to observations not coming from the specific parametric family and/or the observations might be coming from contaminated model, wherein a measure part of the observations come from the parametric model and few number of observations come from contaminated model. This can be well described by stating that the observations have come from a mixture model of "parametric models with non-parametric". Such a mixture model itself is called as non-parametric neighbourhood models. Thus a parametric model contaminated by a (small) non-parametric mixture is called as non-parametric neighbouhood model.

Note : In non-parametric neighbourhood models the parameters of interest are still the parameters of parametric part of the model. However, while estimating or testing one has to be careful about the non-parametric disturbance. In the following we discuss non-parametric models with natural parameters.

Non-parametric models with natural parameters :

In many situations by using data it is possible to propose a parametric model. But many a times, due to change in situation, proposing a suitable parametric model may not be possible. For example, for the class of all symmetric unimodal distribution the point of symmetry might a quantity of interest, that is, for $\mathbb{F} = \{ F(x-\theta) : \theta \in R$ and F is symmetric $\}$, one might be intrested in θ, the location parameter. In such situation one may prefer a totally non-parametric models. In these models there may exists natural parameters which describes an important uses of the model. The estimation or testing of such parameters is a very active field of study. Now we see about the neighbourhood of non-parametric models with natural parametrs.

3.2.1 <u>Neighbourhood of non-parametric models with natural parameters</u> :

Let us assume that we are dealing with a sample from unknown symmetric distribution F. Further assume that this F has a slight amount of asymmetric contamination. In this case the parameter of interest would be the center of symmetry. This is, an example of neighbourhood of non-parametric models with natural parameters.

<u>Note</u> : The types of the models discussed above are the part of an important characteristic that there exists natural measures of of the model under consideration; measures like location, scale and some other measures such as skewness kurtosis. There may be many possible estimators of these measures. The main problem is to choose suitable measure among these measures.

Now we shall consider the models where such measures do not exists that is the model still possess natural parameters. For example in a symmetrical model first quantity to be specified is θ the point of symmetry and the scale parameter σ can be introduced by using the fact that the distribution of $(X - \theta)/\sigma$ is symmetric about zero having unit variance (assuming the 2^{nd} order moment exists). When we are dealing with such situations some questions arise like, how to measure the location of non-symmetric distribution? or how to measure or describe its spread, skewness or kurtosis? such type of questions we have to begin with the description of a particular form of distribution. It is more convenient to answer these questions by defining in each case , a distribution G (or random variable Y with distribution G) posses the attribute under consideration more strongly than the distribution F (or random variable X with distribution F).

Definition (3.2.1) : Partial ordering

A reflexive antisymmetric and transitive relation on a set X is called a partial order on X, such relation is denoted by the symbol $\leq$.

For instance, the set theoretic relation $\subset$ defines a partial order in the set of all subsets of a given set X. We find here that there can be subsets A, B of X for which $A \subset B$ is true but $B \subset A$ need not be true (that is absence of symmetry). Also neither $A \subset B$ nor $B \subset A$ need be true at all. However, if the partial order on set X be such that for $x, y \in X$ either $x \leq y$ or $y \leq x$ is true, then this partial order is said to be a total order (or linear order). A example of total order is the familiar order relation in the set N of natural numbers the usual order relations in Q and R are also total order.

Now we turn to the conditions for θ to be measure.

3.2.2 <u>Conditions for θ to be a measure</u> :

Consider a certain proper attribute of distribution function. Let '<' be a partial ordering defined on $\mathbb{F}$, the class of all distributions. Further $F < G$ indicate that G posses the attribute under consideration is more strongly than F. Let θ be a functional satisfying the following properties.

i) $\theta(F) \leq \theta(G)$ where $F < G$.

ii) $\theta(aX + b) = a\,\theta(X) + b \quad \forall\ a,b$ (3.2.1)

If θ is measure of location then it satisfies this condition.

iii) $\theta(aX + b) = |a|\,\theta(X) \quad \forall\ a \neq 0$ and b (3.2.2)

If θ is measure of scale then it satisfies this condition.

<u>Remark</u> : Once these conditions have been laid down, there are many functionals satisfying above properties. Then a natural question arises is which one of these is to be selected. To answer this question we have to define a functional

<u>Definition</u> (3.2.2) : <u>Functional</u>

A functional is a real valued function defined on domain, class of distribution functions. For example mean (whenever exist), median are the functionals.

Further if $\mathbb{P}$ is class of all normal densities with mean 0 and variance σ^2, then

$\phi(p) = E_p[g\ (x)]$

where X has the distribution belonging to $\mathbb{P}$ is a functional. In location case, the functional

$$\theta(F) = \int_0^1 F^{-1}(t)\ dk(t) \qquad (3.2.3)$$

where K is any distribution function on (0,1) and symmetric with respect to 1/2 defines a large class of such measures.

Remark :

i) Specifically if K is U(0,1) then $\theta(F)$ is our usual mean, median corresponds to K(t) degenerate at 1/2.

ii) If K is discrete uniform over (1/4,3/4) then $\theta(F)$ is our usual mean,median corresponds to K(t) degenerate at 1/2.

ii) If K is discrete uniform over (1/4,3/4) then $\theta(F)$ gives an average of 1^{st} and 3^{rd} quartile.

iii) Q_1 corresponds to K(t) degenerate at 1/4.

iv) Q_3 corresponds to K(t) degenerate at 3/4.

In general if K is degenerate at q we obtain the q^{th} percentile of F. In the following we discuss in detail about location parameter.

3.3 Measure of location (location parameter) :

Measure of location means a functional $\mu(F)$ defined over a suitably large class of distributions which satisfies the following three conditions

Consider a random variable X with distribution F and $\mu(X)$ be a functional then $\mu(X)$ will be a measure of location if it satisfies the following three conditions;

i) If X is stochastically smaller than Y then

$$\mu(X) \leq \mu(Y) \tag{3.3.1}$$

ii) Under change of location or scale .

$$\mu(aX + b) = a\mu(X) + b \text{ if } a > 0$$

iii) It must satisfy

$$\mu(-X) = -\mu(X). \tag{3.3.2}$$

Remark : Condition (i)-(iii) are very natural conditions. But some reachers have objected these conditions. Because according to them, some time location considered only the central part of a distribution. Bickel and Lehamman (1975) give the following example to illustrate that there is no reasonable version of (ii). Because truncation of two stochastically ordered distributions on a common point may reverse the ordering.

Let $G(t) = t \qquad \text{for } 0 < t < 1$

and

$F(t) = t \qquad \text{for } 0 < t < t_o \ (< 1)$

$$= t_o + \frac{B - t_o}{A - t_o}(t - t_o). \quad \text{for } t_o \le t \le t_1$$

where,

$$t_1 = t_o + \frac{B - t_o}{A - t_o}(1 - to)$$

$$= 1 \qquad \text{for } t > t_1 \text{ and } t_o < A < B < 1.$$

Then, $G(t) < F(t)$ for $t_o < t < 1$.

Let F^* and G^* denote the conditional distribution given that the random variable is $< A$. Then,

$F^*(t) < G^*(t)$ for $0 < t < A$.

Related to this remark Bickel and Lehamann (1975) introduces the following three classes of measures of location which we will discuss after the following theorem.

<u>Theorem</u> (3.3.1) : Conditions (i)-(iii) given in section (3.3) imply the following four additional desirable requirements.

1. If F is symmetric with respect to θ, then, $\mu(X) = \theta$,
2. In particular if $X = C$, with probability 1, $\mu(X) = C$,

3. If $a \leq X \leq b$ with probability 1, then $a \leq \mu(X) \leq b$, and

4. If X is stochastically positive then $\mu(X) \geq 0$.

<u>Proof</u> : Without loss of generality assume that, F is symmetric about zero. We know that if F is symmetric about zero, then the corresponding random variable X and that of (−X) have the same distribution.

Now from condition (iii) we have,

$\mu(-X) = -\mu(X)$

and further X and (−X) have the same distributions, we have,

$\mu(X) = \mu(-X)$

Hence,

$\mu(X) = -\mu(X)$

That is $\mu(X) = 0$ (3.3.3)

In particular if we take $P(X = C) = 1$ then we get,

$\mu(X) = c.$ (3.3.4)

We have $P (a \leq X \leq b) = 1$ then X is stochastically larger than constant variable a and smaller than b. We have, the result that if F is symmetric about θ then $\mu(F) = \theta$. and if X = C with probability 1 . Then the result follows from these two statements.

To prove the fourth condition first we have to define stochastically positive random variable.

Definition (3.3.1) : Stochastically positive random variable

X is said to be stochastically positive if there exists a random variable U symmetric about zero and such that X is stochastically larger than U.

Therefore by the first part of the theorem $\mu(U) = 0$. Hence the result follows from the condition (i).

Example (3.3.1) :

The functional $\mu(F) = F^{-1}(a)$ with $0 < a < 1$ $a \neq 1/2$ satisfies (i) and (ii) but not (iii). Consider a random variable X with distribution function F and $\mu(X)$ be a functional,

$$\mu(X) = F^{-1}(X) \quad \text{with} \quad 0 < X < 1 \text{ and } X \neq 1/2 \qquad (3.3.5)$$

We know that,

$$F^{-1}(1/2) = \text{Median}.$$

But we are given that $X \neq 1/2$. Hence the distribution is not symmetric. Therefore the condition $\mu(-X) = -\mu(X)$ do not satisfies.

Now X is stochastically positive therefore condition (i) follows from the requirements 4 of the above theorem. For the second condition we proceed as follows:

$$\mu(X) = F^{-1}(X) \qquad (3.3.6)$$

Replacing X by aX + b in the above equation, we get

$$\mu(aX + b) = a\,\mu(X) + b \tag{3.3.7}$$

Thus the functional

$\mu(X) = F^{-1}(X)$ for $0 < X < 1$ and $X \neq 1/2$.

satisfies the conditions (i) and (ii) but not (iii).

<u>Example</u> (3.3.2) : The functional

$\mu(F) = 2E(X) - F^{-1}(1/2)$

satisfies (ii) and (iii) but not (i).

Consider the random variable Y which is stochastically larger than X having distribution

$Y \sim G = 1/2\ U(0,\ 1/2) + 1/2\ \Delta(1)$

then,

$E(Y) = 1/2\ (1/4 + 1)$

and

$m(Y) = 1$; where, $m(Y)$ is the median of Y. Therefore,

$$\mu(Y) = 2\ E(Y) - m(Y)$$

$$= 5/4 - 1$$

$$= 1/4.$$

Hence, $\mu(X) \leq \mu(Y)$ is not satisfied.

Now using, $\mu(X) = 2\ E(X) - m(X)$ we have,

$$\mu(aX + b) = 2\ [\ a\ E(X) + b] - (am + b)$$

$$= a\mu(X) + b.$$

and

$$\mu(-X) = 2\ E(-X) - m(-X)$$

$$= -2\ E(X) + m(X)$$

$$= -\mu(X)$$

Thus the functional given above satisfies (ii), (iii) but not (i).

<u>Example</u> (3.3.3) :

$\mu(X) = [\ E(X)\ ]^3$ satisfies (i) and (iii) but not (ii).

We know the result,

$$\mu(X) = E(X) = \int_0^{+\infty} \left[\ 1 - F(x)\ \right] dx - \int_{-\infty}^{0} F(x)\ dx \qquad (3.3.8)$$

Since, the random variable X is stochastically smaller than Y.

$F_X(x) > F_Y(x)$

$1-F_X(x) < 1- F_Y(x)$

Therefore using equation (3.3.8) we can write,

$$EX = \int_{o}^{+\infty} \left[1 - F_X(x)\right]dx - \int_{-\infty}^{o} F_X(x)dx \le \int_{o}^{+\infty} \left[1 - F_Y(x)\right]dx - \int_{-\infty}^{o} F_Y(x)dx = EY$$

That is, $E(X) \le E(Y)$

Hence,

$$[\ E(X)\ ]^3 \le [\ E(Y)\ ]^3 \qquad (3.3.9)$$

Thus the condition (i) is satisfied.

Consider,

$$\mu(aX + b) = \left[\ E(aX + b)\ \right]^3$$

$$= E(aX)^3 + 2\ E(aX)^2 b + E(aX)b^2 + E(aX)^2 b + 2\ E(aX)b^2 + b^3$$

$$\ne a\ E(X)^3 + b \qquad (3.3.10)$$

Hence the condition (ii) is not satisfied.

To prove the third condition consider,

$$\mu(-X) = [\ E(-X)\ ]^3$$

$$= -\ [\ E(X)\ ]^3$$

$$= -\ \mu(X). \qquad (3.3.11)$$

Hence the condition (iii) is satisfied.

From the above three examples one may say that the conditions (i)-(iii) are independent of each other.

<u>Remark</u> : Standard measures of location namely mean and median satisfies these three conditions.

1) The mean : EX :

$\mu(X) = E(X)$

Consider the random variable X is stochastically smaller than the random variable Y. Therefore,

$F_X(x) > G_Y(x)$

$1 - F_Y(x) < 1 - G_Y(x)$

Hence,

$$EX = \int_0^{+\infty} \left[1 - F_X(x)\right] dx - \int_{-\infty}^{0} F_X(x) dx$$

$$\leq \int_0^{+\infty} \left[1 - F_Y(x)\right] dx - \int_{-\infty}^{0} F_Y(x) dx$$

$$= EY$$

Hence,

$$E(X) \leq E(Y) \qquad (3.3.12)$$

Now.

$$\mu(aX + b) = \left[E(aX + b) \right]$$

$$= a\ E(X) + b$$

$$= a\ \mu(X) + b \qquad (3.3.13)$$

Consider, $\mu(-X) = E(-X)$

$$= -E(X)$$

$$= -\mu(X) \qquad (3.3.14)$$

Thus, from equation (3.3.12), (3.3.13) and (3.3.14) the mean satisfies the all three conditions.

2) Median :

Let $\mu(X) = m(X)$ be the median of random variable X with distribution function F and Y be the random variable with distribution function G which is stochastically larger than X. Then,

$F(x) > G(x)$

$F^{-1}(t) < G^{-1}(t) \qquad \forall\ 0 < t < 1.$

In particular t = 1/2 we get,

$F^{-1}(1/2) < G^{-1}(1/2)$

This implies that,

$$m(X) < m(Y) \tag{3.3.15}$$

Let, $\mu(-X) = m(-X)$

$$= -m(X)$$

$$= -\mu(X) \tag{3.3.16}$$

and $\mu(aX + b) = m(aX + b)$

$$= am(X)+b$$

$$= a\mu(X)+b \tag{3.3.17}$$

Thus, form equation (3.3.15), (3.3.16), and (3.3.17) the median also satisfies the all three conditions.

The doubly trimed mean defined by,

$$E_\alpha(X) = \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x \, dF(x) \tag{3.3.18}$$

is a class of location measures.

The mean and median are limiting cases with respect to $\alpha \to 0$ and $\alpha \to 1/2$.

Let,

$$I = \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x \, dF(x) \qquad (3.3.19)$$

Let us assume that differentiation and integration may interchangable. Differentiating (3.3.19) with respect to α we get,

$$\frac{dI}{d\alpha} = \frac{1}{-2} \frac{d}{d\alpha} \left[H(1-\alpha) - H(\alpha) \right] \qquad (3.3.20)$$

Taking limit as $\alpha \longrightarrow 1/2$ we get,

$$\lim_{\alpha \to 1/2} \frac{dI}{d\alpha} = \lim_{\alpha \to 1/2} \frac{1}{-2} \frac{d}{d\alpha} \left[H(1-\alpha) - H(\alpha) \right]$$

$$= \frac{1}{-2} [-m - m]$$

$$= m$$

This implies that, as $\alpha \longrightarrow 1/2$ equation (3.3.19) gives the median of F and at $\alpha = 0$ we get,

$$I = \int_{F^{-1}(0)}^{F^{-1}(1)} x \, dF(x) \qquad (3.3.21)$$

which is the mean of X. Thus, mean and median are limiting cases corresponds to $\alpha = 0$ and $\alpha = 1/2$.

<u>Definition</u> (3.3.2) : <u>Pseudomedian</u>

Let X_1, X_2 are independently distributed according to F. Then the median of distribution of $\frac{1}{2}(X_1 + X_2)$ denoted by $\mu_4(F)$ is called as Pseudo median.

<u>Lemma</u> (3.3.1) : If F is continuous then, $\mu_4(F)$, the pseudo median is the solution of equation

$$\int F(2\theta - x)\, dF(x) = 1/2 \tag{3.3.22}$$

<u>Proof</u> : Consider,

$$Y = \frac{1}{2}(X_1 + X_2) \tag{3.3.23}$$

then by the definition of median if θ is the median of Y we have ,

$$P\left[\frac{1}{2}(X_1 + X_2) < \theta\right] = \frac{1}{2} \tag{3.3.24}$$

That is,

$$P\left[(X_1 + X_2) < 2\theta\right] = \frac{1}{2}$$

$$P\left[X_1 < 2\theta - X_2\right] = \frac{1}{2} \tag{3.3.25}$$

Thus,

$$\int_{-\infty}^{o} F(2\theta - x)\, dF(x) = 1/2$$

Which implies that, θ is the median of $Y = \frac{1}{2}(X_1 + X_2)$.

Hence the proof.

<u>Classes of Location Parameter</u> :

A) If $\mu(X)$ satisfies conditions (i) and (ii) of section (3.4) but not (iii) then, $\frac{1}{2}\left[\mu(X) - \mu(-X) \right]$ satisfies all these three conditions that is the above function can be considered as location parameter.

i) $\mu(X) \leq \mu(X)$

$-\mu(X) \geq -\mu(Y)$

$-\mu(-X) \leq -\mu(-Y)$ since $-\mu(X) = \mu(-X)$

$\mu(X) - \mu(-X) \leq \mu(Y) - \mu(-Y)$

$$\frac{1}{2}\left[\mu(X) - \mu(-X) \right] \leq \frac{1}{2}\left[\mu(Y) - \mu(-Y) \right]$$

ii) $\frac{1}{2}\left[\mu(aX + b) - \mu(-(aX + b)) \right]$

$$= \frac{1}{2}\left[a\mu(X) + b - a\mu(-X) + b \right]$$

$$= \frac{a}{2}\left[\mu(X) - \mu(-X) \right] + b.$$

Lemma (3.2.1) : If $\left\{\mu_i(x)\right\}$ is a countable collection of functional which satisfies (i)-(iii) then so does $\Sigma\alpha_i\mu_i(x)$ for any α non-negative, also satisfies (i)-(iii).

Proof : Proof is trivial.

Consider the functional $\mu(x) = F^{-1}(a)$ for some fixed a between 0 and 1. This functional satisfies the conditions (i) and (ii). Now $\mu(-X) = -F^{-1}(1-a)$. This follows that $\frac{1}{2}\left[F^{-1}(a) + F^{-1}(1 - a)\right]$ is also a location parameter for any $0 < a < 1$.

B) To find the second class of location consider the probability

$$
\begin{aligned}
P (| X - \theta | \leq x) &= p(-x \leq (x-\theta) \leq x) \\
&= P(\theta - x \leq X \leq x + \theta) \\
&= F(x + \theta) - F(-x + \theta)
\end{aligned}
$$

Therefore θ can be defined as the center of F for which the above probabilities are as large as possible.

Let L be an increasing convex function on [0,1] which is bounded and such that L(0)=0. Define $\mu(F)$ as the number θ which maximizes

$$
\int \left\{ L\,[F(x + \theta) - F(-x + \theta)] - L[F(x) - F(-x)]\right\} dx \qquad (3.3.26)
$$

The subtraction of $L[F(x) - F(-x)]$ under the integral sign is intended to aid convergence.

Remark :

1) L' is bounded in absolute value.

2) It is more convenient to extend L so that it is an even function on [-1,1], then, it is equivalent to maximize

$$\int_{\infty}^{\infty} \left\{ L\,[F(x+\theta) - F(-x+\theta)] - L[F(x) - F(-x)] \right\} dx$$

$$= \int_{-\infty}^{\infty} \left\{ L\,[F(x) - F(2\theta - x)] - L[F(x) - F(-x)] \right\} dx \qquad (3.3.27)$$

If we suppose F has density f, L is continuously differentiable, that $M = L'$ and that above equation can be differentiated under the integral sign.

C) The third class is obtained by using Huber's M-estimator.Hence first we describe the definition of Huber's M-estimator.

These estimators are just a slight generalization of MLE's so it is also known as generalized maximum likelihood type estimators.

Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x,\theta)$ where unknown parameter θ belongs to some parameter space Θ .Consider estimators of θ which are functional (real valued statistics)

$T_n = T_n\ (X_1, X_2, \ldots, X_n)$ based on sample. The estimators of type M are the solutions of the more general structure proposed by Huber (1964).

$$\sum_{i=1}^{n} \rho(x_i, T_n), \qquad (3.3.28)$$

where ρ is some function on $\mathcal{X}$. $X, \theta, \mathcal{X}$ being the sample space.

Suppose that ρ has a derivative $\phi(x,\theta) = \frac{\delta}{\delta\theta}\rho(x,\theta)$, so the estimate T_n satisfies the implicit equation

$$\sum_{i=1}^{n} \phi(x_i, T_n) = 0 \qquad (3.3.29)$$

Any estimator defined by (1) or (2) is called an M-estimator.

<u>Remark</u> : If H_n is the empirical distribution function corresponding to the sample, then the solution T_n of (2) can also be written as $T(H_n)$, where T is functional given by

$$\int \phi(x, T(H))\ dH(x) = 0 \qquad (3.3.30)$$

Thus for each distributions define functional H_n for which the integral

$$\int \phi(x, T(H))\ dH(x) = 0 \qquad (3.3.31)$$

is defined [Ref. Hample (1986) page 101].

Another class of measures of location is obtained from the quantities estimated by Huber's M-estimator. These are quantities $\theta = \mu(F)$ which minimize

$$I = \int \rho(x - \theta) dF(x) \qquad (3.3.32)$$

where we shall assume ρ to be positive, even convex and twice differentiable with derivative $\rho' = \phi$. Alternatively I can be viewed as risk function corresponding to the loss function $\rho(x - \theta) = L(x, \theta(x))$, the loss incurred by proposing θ as location, when x is observed.

In the following theorem Bickel and Lehmman (1975) show that only one parameter sub class of (c) satisfies the conditions given for θ to be a measure.

<u>Theorem</u> (3.3.2) : Suppose that $\mu(F)$ is defined as minimizing

$$\int \rho(x - \theta) dF(x) \qquad (3.3.32)$$

on set $\mathbb{F}$, which is convex, contains all point masses, is closed under changes of scale, and contains a distribution F^{o} symmetric about 0 such that,

$\int \phi'(x) dF^{o}(x + t) < \infty$ for all t and

Differentiating with respect to x we get,

$$\frac{d}{dt} \int \phi(x - t)dF^{o}(x) = -\int \phi'(x - t)dF^{o}(x)$$

Let $Y = x/\sigma$ and F_{σ} be the distribution of Y. Suppose that

$$\mu(F_{\sigma}) = \sigma\mu(F) \text{ for all } F \in \mathbb{F}, \sigma > 0 \qquad (3.3.33)$$

Then,

$$\phi(x) = c\,|x|^{\alpha} \operatorname{sgn} x \text{ for some } \alpha > 0, c > 0$$

<u>Proof</u> : The measure $\mu(F)$ is the solution of the equation

$$\int\phi(x - \theta)dF(x) = 0$$

Using assumptions, we have,

$$\lim_{\varepsilon \to 0} \frac{\mu\left[(1 - \varepsilon)F^{o} + \varepsilon\sigma_{x}\right] - \mu(F^{o})}{\varepsilon} = \frac{\phi(x)}{\int_{-\infty}^{\infty}\phi'(x)dF^{o}(x)}$$

and similarly,

$$\lim_{\varepsilon \to 0} \frac{\mu\left[(1 - \varepsilon)F^{o} + \varepsilon\delta_{x}\right] - \mu(F^{o})}{\varepsilon} = \frac{\phi(\sigma x)}{\int_{-\infty}^{\infty}\phi'(\sigma y)dF^{o}(y)}$$

Form equation (3.3.29) we get,

$$\frac{\phi(\sigma x)}{\int_{-\infty}^{\infty}\phi'(\sigma y)dF^{o}(y)} = \frac{\sigma\ \phi(x)}{\int_{-\infty}^{\infty}\phi'(y)dF^{o}(y)}$$

Differentiate both the sides with respect to x for x > 0 leads to

$$\frac{\phi'(\sigma x)}{\int_{-\infty}^{\infty}\phi'(\sigma y)dF^{o}(y)} = \frac{\phi'(x)}{\int_{-\infty}^{\infty}\phi'(y)dF^{o}(y)}$$

and hence

$$\frac{\phi'(\sigma x)}{\phi'(x)} = \frac{\phi(\sigma x)}{\sigma\phi(x)} \qquad \text{for some } x > 0.$$

put x = 1 in the above equation we get,

$$\frac{\phi'(\sigma)}{\phi(\sigma)} = \frac{\phi'(1)}{\sigma\phi(1)}$$

$$\text{or} \quad \log\phi = \frac{\phi'(1)}{\phi(1)} \log \sigma + c$$

Hence the proof.

<u>Remark</u> : This theorem also holds under a weaker set of conditions not including differentiability.

3.4. Measure of dispersion :

A measure of dispersion is defined as a functional which satisfies the certain equivariance and order conditions. Here we discuss measures of dispersion for symmetric distributions. This functional also has an addition property that it assigns a larger value to G than to F if G is more dispersed than F.

Let X be a random variable with distribution function F which is symmetric about zero. Then the dispersion can be interpreted in terms of distance of X from the center μ. In terms of magnitude.

Definition (3.4.1) :

Let Y be a random variable with distribution function G symmetric about ν. Then Y is said to be more dispersed about ν than X about μ if,

$$| Y - \nu | \overset{s.t}{\leq} | X - \mu | \qquad (3.4.1)$$

That is magnitude $| Y - \nu |$ is stochastically smaller than the magnitude $| X - \mu |$.

Remark :

i) Any symmetric random variable is more dispersed than a constant.

ii) If $a > 1$, then, aX is more dispersed than X.

iii) If we take $\mu = 0$ and $\nu = 0$ in the definition 3.4.1 then, we get,

$$P(|Y| \leq x) < P(|X| \leq x)$$

Hence,

$$g(x) = P(-x \leq Y \leq x) < P(-x \leq X \leq x) = f(x)$$

Therefore, $g(x)/f(x)$ is increasing function in x for $x > 0$.

<u>Lemma</u> (3.4.1) : Let X be a random variable with the distribution function F and Y be the random variable with the distribution function G. Consider both the random variables symmetric about zero and G is more dispersed than F. Then the random variable Z having distribution function $H_\theta(x)$ as,

$$H_\theta(x) = \theta\, G(x) + (1 - \theta)\, F(x) \qquad (3.4.2)$$

is more dispersed than F for any $0 < \theta < 1$ and less dispersed than that of G.

<u>Proof</u> : Proof is trivial.

<u>Theorem</u> (3.4.1) : Let X_i Y_i $(i = 1,2)$ be independent random variables with distributions F_i, G_i $(i = 1,2)$ which are symmetric about zero. Suppose that,

i) Y_i is more dispersed than X_i ; $i = 1,2$. and

ii) F_1 and G_1 are unimodal densities and possibly some probability mass at zero.

Then $Y_1 + Y_2$ is more dispersed than $X_1 + X_2$.

Proof : Consider the probability,

$$P(|X_1 + X_2| \le C) = P(-C < (X_1 + X_2) < C)$$

$$= P(C - X < X < C + X)$$

$$= 2 \int_0^\infty \left[F_1(x + c) - F_1(x - c)\right] dF_2(x) \qquad (3.4.3)$$

We are given that F_1 is the unimodal density, therefore, the integrand on R.H.S. in the above integral is decreasing function of X. Since Y_i is more dispersed than X_i it satisfies

$|X - \mu| \le | Y - \nu|$

Therefore if we replace F_2 by G_2 in above integral it is also a decreasing function. Thus,

$$P(|X_1 + X_2| \le C) \ge 2 \int_0^\infty \left[F_1(x + c) - F_1(x - c)\right] dG_2(x)$$

This implies that,

$$P(|X_1 + X_2| \le C) = 2 \int_0^\infty \left[G_2(x + c) - G_2(x - c)\right] dF_1(x)$$

Since G_2 is unimodal density here again we got R.H.S, is decreasing function, that is,

$$P(|X_1 + X_2| \leq C) \geq P(|Y_1 + Y_2| \leq C) \qquad (3.4.4)$$

Which implies that, $(X_1 + X_2)$ is stochastically smaller than $(Y_1 + Y_2)$, that is, $(Y_1 + Y_2)$ is more dispersed than $(X_1 + X_2)$.
Hence the proof. ■.

Definition (3.4.2) : Measure of dispersion

Let X be a random variable with distribution function F. Consider $\tau(F)$ be a non-negative functional defined over a sufficiently large class of symmetric distributions which is closed under changes of location and scale. Then the functional $\tau(F)$ is said to be a measure of dispersion if it satisfies the following conditions.

i) $\tau(ax) = |a|\ \tau(x)$ for $a > 0$.

ii) $\tau(ax + b) = \tau(x)$ for all b.

iii) $\tau(F) \leq \tau(G)$ whenever G is more dispersed than F.

Remark :

i) If F is symmetric then, $\tau(-x) = \tau(x)$.

ii) If C is any constant then from conditions (i) and (ii) it follows that, $\tau(c) = 0$.

iii) If $X = C$ with probability 1 then, $\tau(x) = \tau(c) = 0$.

A wide class of measures of dispersion is obtained by the functionals,

$$\tau(F) = \left\{ \int_{0}^{1} [F_*^{-1}(t)]^{\gamma} d\Lambda(t) \right\}^{1/\gamma} \tag{3.4.5}$$

where F is symmetric about μ, F_* denote the distribution of $|X - \mu|$ and Λ is any probability distribution on (0.1) is be any positive number.

iv) The functional defined above satisfies the conditions (i) and (ii).

v) Let X be a random variable with distribution function F and G Y be the random variable with distribution function G. Consider X be stochastically smaller than Y, then from (iii) in definition 3.4.2 it follows that

$F_*^{-1}(t) \le G^{-1}(t)$

vi) If $\gamma = 2$ and λ be the uniform over (0,1) then, the function $\tau(F)$ gives the standard deviation of F which is defined as

$$S.D. = \left[\int (x - \mu)^2 \, dF(x) \right]^{1/2}$$

<u>Classes of measures of dispersion :</u>

There are three classes of measures of dispersion which are obtained as a special cases of $\tau(F)$ given by (3.4.5).

A) A generalization of the standard deviation is the p^{th} power deviation. This is obtained by replacing γ by p in $\tau(F)$ and letting λ be the uniform over (0,1). It is denoted by $\tau(F;p)$.

B) Let Λ be the uniform over $(\alpha,1-\beta)$ with γ =2 then doubly trimmed standard deviation $\tau(F;\alpha,\beta)$ is the class of measures of dispersion.The definition of doubly trimmed standard deviation is given in definition (3.4.3).

C) The third class is obtained by considering that Λ assigns the probability 1 to the point α. We get α^{th} quantile and the resulting measure is independent of γ.

<u>Note</u> : The standard deviation is a member of classes defined in (A) and (B). The α^{th}quatile is the limit of the doubly trimmed standard deviation.

<u>Definition</u> (3.4.3) : <u>Doubly trimmed standard deviation.</u>

Let X be a random variable distributed independently with the distribution function F, which is symmetric about origin.Consider $Y = X^2$ having distribution G, then we can write

$$\tau^2(F;\alpha,\beta) = \frac{1}{1-\alpha-\beta}\int_{\alpha}^{1-\beta} G^{-1}(t)\,dt$$

Thus,

$$\tau^2(F;\alpha,\beta) = \frac{1}{1-\alpha-\beta}\int_{\mu_\alpha}^{\mu_1-\beta} G^{-1}(t)\,dt$$

where $\mu\alpha$ is α^{th} percentile of G.

Then $\tau^2(F;\alpha,\beta)$ defined above is called as doubly standard deviation.

Now consider,

$G(t^2) = P\,[\,X^2 < t^2]$

$= p\,[|X^2| \leq t^2]$

$= p\,[|X| \leq t]$

$= F^*(t)$

That is,

$G(t^2) = F^*(t) = u(say).$

Hence,

$t^2 = G^{-1}(u)$

and

$t = F^{*-1}(u).$

Therefore,

$G^{-1}(u) = [F^{*-1}(u)]^2.$

Hence,

$$\tau^2(F;\alpha,\beta) = \frac{1}{1-\alpha-\beta}\int_{\alpha}^{1-\beta} [F^{*-1}(u)]^2 \, du.$$

$$= \frac{1}{1-\alpha-\beta}\int_{\alpha}^{1-\beta} G^{-1}(u) \, du.$$

<u>Example</u> (3.4.1) : Let

$$F(t) = \begin{cases} 0 & t < -1 \\ \frac{t+1}{2} & -1 < t < 1 \\ 1 & t \geq 1 \end{cases}$$

Then

$$F(u) = \begin{cases} 0 & u < -1 \\ u & 1 < u < 1 \\ 1 & u \geq 1 \end{cases}$$

Then, $G(u) = F^{*}(v^{1/2})$

Hence,

$$\tau^2(F;\alpha,\beta) = \frac{1}{1-\alpha-\beta}\int_{\alpha}^{1-\beta} u \, du.$$

$$= \frac{1-\beta+\alpha}{2} \quad .$$

* * * * *