C H A P T E R - IV

OTHER DESCRIPTIVE MEASURES

4.1 Introduction :

In this chapter we introduce some new descriptive measures .In section (4.2) we describe some new measures of location which are obtained by combining two measures [Rattihalli (1996)]. In section (4.3) we discuss the measure for peakedness introduced by Paul(1983). This measure is defined even for the distributions for which moments not exist. Then we discuss some properties of this measure and obtain measure for crammer density and double exponential distributions. Generally we say that if the given distribution is normal then its kurtosis is definitely equal to 3. Even for a non-normal distribution it may be equal to 3, a class of such distributions is introduced by Kale and Sebastian (1996) which we discuss in the last section of this chapter.

4.2 Measures of Location :

In this section we discuss two measures of location which are obtained by combining the two location measures. Let n = mk and M_i be a measure of location based on a random sample of size k formed by $X_{(i-1)k+1}$, $X_{(i-1)k+2}$, ..., X_{ik} (i=1,2,...,m).

Further U be a measure of location based on M_1, M_2, ..., M_m. For example M_i be the mean of $X_{(i-1)k+1}$, $X_{(i-1)k+2}$, ..., X_{ik} (i=1,2,...,m). and U be the median of M_1, M_2, ..., M_m. We shall refer U as the median of arithmetic means of $X_{(i-1)k+1}$, $X_{(i-1)k+2}$, ..., $X_{(i)k}$. For simplicity we shall denote it by

$$MA(X) = U.M(X). \tag{4.2.1}$$

Similarly we define AM(X) (not mean to arithmetic mean of X but arithmetic mean of median of X). It is denoted by,

$$AM(X) = M.U(x) \tag{4.2.2}$$

Welknown measures of location are also of this form;

i) The sample mean : i^{th} sample is $\{X_i\}$, M_i be the mean of i^{th} sample and U is the mean of M_i´s.

ii) The sample median : i^{th} sample is $\{X_i\}$, M_i be the median of i^{th} sample, and U is the median of M_i´s.

In general any statistic W of this form can be viewed as $M_i = \{ X_i\}$ and U = W. These measure satisfy the desired properties of measure of location.

i) If Y is stochastically larger than X, then ,

$$U.M(X) \leq U.M(Y) \tag{4.2.3}$$

ii) Under change of location or scale,

$$U.M(aX + b) = a\,U.M(X) + b \tag{4.2.4}$$

iii) The measure of location change the sign under reflection with respect to the origin, that is it satisfies the condition.

$$U.M(-X) = -U.M(X) \tag{4.2.5}$$

To obtain a sample version we have to obtain F_n, the empirical distribution function. Such measures are used to define some new measures of symmetry. [Refer Rattihalli (1996)].

4.3 Measure of dispersion :

Let $-\infty \le X_1 \le X_2 \le \ldots\ldots \le X_n \le \infty$ be n given numbers, then the functional,

$$\mu(F) = \inf_{\theta} \left\{ E\,[W(X)\;|X - \theta|] \right\} = E\left\{W(X)\;|X - \theta_w|\right\} \tag{4.3.1}$$

be a measure of dispersion. where W(x) is a non-negative numbers. Let,

$$H(\theta) = E\left\{ W(X)\;|X - \theta_w| \right\}$$

$$= \int_{-\infty}^{\infty} w(x)\;|x - \theta|\;f(x)\,dx.$$

$$= \int_{-\infty}^{\theta} w(x)(\theta - x)f(x)\,dx + \int_{\theta}^{\infty} w(x)\,(x - \theta)\,f(x)\,dx \tag{4.3.2}$$

Applying Lebnit'z rule of differentiation we get,

$$\frac{d}{d\theta} H(\theta) = \int_{-\infty}^{\theta} w(x)\ f(x)\ dx. + \int_{\theta}^{\infty} w(x)\ f(x)\ dx \qquad (4.3.3)$$

If we set this equation is equal to 0 we get,

$$F(\theta) = \int_{\infty}^{\theta} f(x)\ dx.$$

$$= 1/2 \qquad (4.3.4)$$

If $w(x) = 1$ then $\mu(F)$ will be minimum when θ = median.

4.4 A New measure for Peakedness :

Kurtoss is used to characterize the peakedness of a density. Sometimes, the kurtosis does not exist. A natural question in such a situation is how to measure the peakedness? In this section we discuss a measure of peakedness which exist for all densities. The numerical measure of peakedness would be helpful to make more precise statements about the peakedness of any distribution. The well known measure of kurtosis is

$$\beta_2 = \mu_4/\mu_2^2 \qquad (4.4.1)$$

This measure does not help to make more precise statements about the peakedness of any density because it does not exist for the Caushy distribution. A measure for peakedness which exist for all symmetric unimodal densities would make comparisons between densities more meaningful.

<u>Definition</u> (4.4.1) : <u>Measure for peakedness</u>

Let f(.) be a symmetric unimodal (say about 0) the density function, and let F(.) be the corresponding distribution function. Consider a rectangle in the X-Y plane which is formed by the following lines:

i) $X = 0$; $Y = 0$.

ii) $X = F^{-1}(p + 0.5)$; $Y = f(0)$ for some $0 < p < 0.5$.

Let us call this rectangle $R_p(f)$. The area of this rectangle is given by ;

$$A_p(f) = f(0)\, F^{-1}(p + 0.5) \qquad (4.4.2)$$

Thus, the measure of peakednesss would be the number ;

$$mt_p(f) = 1 - \frac{p}{A_p(f)} \qquad (4.4.3)$$

where $\frac{p}{A_p(f)}$ is the proportion of area of $R_p(f)$ covered by the density f(.) $A_p(f) = f(0).F^{-1}(p + 0.5)$.

Remark :

1) The area under the density contained in $R_p(f)$ is equal to p for all f(.).

2) If $\frac{p}{A_p(f)}$ is near to 1, then naturally most of the density is under rectangle $R_p(f)$ and therefore f(.) is looking rectangle not very peaked.

Thus we can write,

i) If $mt_p(f) = o$, the density is rectangular.

ii) $mt_p(f)$ is not exactly one but near to 1 f(.) looks like spike with a long tail for all p : $0 < p < 0.5$

Thus smallness of p relative to $A_p(f)$ is an indicative of being more peakedness of density.

Example (4.4.1) :

1) The crammer density is defined as,

$$f(x) = \frac{\theta}{2(1+\theta|x|^2)} \qquad -\infty < X < \infty,\ \theta > 0. \qquad (4.4.4)$$

$$= 0 \qquad \text{otherwise.}$$

From the density given we get, $f(0) = \theta/2$. Now find

$U = F^{-1}(p + 0.5)$.

The value of U can be find by solving the following integral as follows;

$$\int_{o}^{u} \frac{\theta}{2(1+\theta x)^2}\, dx = p$$

That is,

$$\frac{1}{2}\left\{1 - \frac{1}{1+\theta u}\right\} = p$$

Which implies that,

$$1 - \frac{1}{1+\theta u} = 2p$$

or

$$u = 2p / \theta(1 - 2p). \qquad (4.4.5)$$

Thus,

$$A_p(f) = \frac{p}{1- 2p} \qquad (4.4.6)$$

Form (4.4.2) measure for peakedness for the given density is,

$$m\!\!t_p(f) = 1 - p / A_p(f)$$

Substituting value of $A_p(f)$ from equation (4.4.6) in (4.4.3), we get,

$$m\!\!t_p(f) = 2p. \qquad (4.4.7)$$

Example (4.4.2) : The double exponential (Laplace) density is defined as

$$f(x) = \frac{1}{2} \text{ expo} \left\{ - | x | \right\} \qquad -\infty < x < \infty \qquad (4.4.8)$$

$$= 0 \qquad \text{otherwise.}$$

Clearly, f(0) = 1/2. The value of $F^{-1}(p + 0.5)$ is found by solving the following integral.

$$\frac{1}{2} \int_{o}^{u} \text{expo}\{-x\} \, dx = p$$

where $u = F^{-1}(p + 0.5)$. This gives,

$$\frac{1}{2} \left[\frac{\text{expo}\{-x\}}{-1} \right]_0^u = p$$

Equivalentaly we have,

$$\frac{1}{2} \left[1 - \text{expo } \{ -u \} \right] = p$$

or

$$u = \log\left[1 / (1 - 2p)\right] \qquad (4.4.9)$$

Thus,

$$A_p(f) = f(0).F^{-1}(p + 0.5)$$

$$= 1/ 2 \log(1 - 2p) \qquad (4.4.10)$$

Therefore from (4.4.3) and (4.4.10) the measure for peakedness is

$m_p(f) = 1 - p / \Lambda_p(f)$

That is,

$$m_p(f) = p / 2\log(1- 2p). \qquad (4.4.11)$$

Lemma (4.4.1) : The measure $m_p(f)$for peakedness is free form the scale.

Proof : Let X be the random variable with density f(x). Consider the random variable

$$Y = \sigma X. \qquad (4.4.12)$$

Then,

$f(y)=1/[\sigma f(y/\sigma)](4.4.13)$

Now $F^{-1}(p + 0.5)$ can be obtained by solving the equation,

$P(Y < t) = p$

That is,

$P(X < t/\sigma) = p$

This gives,

$$t = F_Y^{-1}(p + 0.5). \qquad (4.4.14)$$

Similarly,

$F_X^{-1}(p + 0.5) = t/\sigma.$

or

$$\sigma F_X^{-1}(p + 0.5) = t \qquad (4.4.15)$$

Therefore we can write,

$$\sigma F_X^{-1}(p + 0.5) = F_Y^{-1}(p + 0.5) \qquad (4.4.16)$$

Then the measure for peakedness is

$$m_Y(p) = 1 - \frac{p}{\frac{1}{\sigma} f(0).F_Y^{-1}(p+0.5)}$$

$$= 1 - \frac{p}{\frac{1}{\sigma} f(0).F_X^{-1}(p+0.5)\sigma}$$

$$= 1 - \frac{p}{f(0).F_X^{-1}(p+0.5)}$$

$$= m_X(p)$$

which implies that $mt_p(f)$ is free from the scale.

4.5 Non-normal distributions with Kurtosis equal to 3 :

In this section we introduce a wide class of non-normal symmetric distributions which have kurtosis 3. This can be obtained by considering a mixture of two symmetric non-normal densities of which one has kurtosis strictly less than 3 and of the other has kurtosis strictly greater than 3. The p.d.f.s can be very much different from the normal density. This class was introdiced by Kale and Sebastian (1996). If f(x) and g(x) are two symmetric distributions with $\mu_f = \mu_g = 0$ and variances σ_f^2 σ_g^2 respectively, and $\beta_2(g) < 3$ and $\beta_2(f) > 3$. Then there exist a unique mixture of f and g such that $\beta_2(\text{mixture}) = 3$.

<u>Theorem</u> (4.5.1) : Let G be the class of all probability distribution functions symmetric around 0 and $\beta_2 < 3$. Let F be the class of all probability distribution functions symmetric around 0 with $\beta_2 > 3$. Then for every pair of probability distribution functions $g \in G$ and $f \in F$ there is unique $\alpha \in (0,1)$ such that $\beta_2(h_\alpha) = 3$, where,

$$h_\alpha(x) = \alpha\, g(x) + (1 - \alpha)\, f(x) \qquad (4.5.1)$$

<u>Proof</u> : Given that F and G are symmetric distributions around zero. If μ_f and μ_g are means of F and G respectively, then, mean of $h_\alpha(x)$ can be founded as follows :

$$h_\alpha(x) = \alpha\, g(x) + (1 - \alpha)\, f(x)$$

Taking expectation on both sides of the above equation, we get,

$E\left[h_\alpha(x)\right] = \alpha E\left[g(x)\right] + (1 - \alpha) E\left[f(x)\right]$

$$= \alpha \mu_g + (1 - \alpha) \mu_f \qquad (4.5.2)$$

Since $\mu_g = \mu_f = 0$, $E\left[h_\alpha(x)\right] = 0$.

Therefore, if μ_α be the mean of $h_\alpha(x)$, then, $\mu_\alpha = 0$. Hence $h_\alpha(x)$ is also symmetric about zero, for all $\alpha \in [0,1]$.

Similarly, the variance of

$h\alpha(x) = \alpha\ g(x) + (1 - \alpha)\ f(x)$

is given as

$$\sigma^2 h\alpha(x) = \alpha\ \sigma_g^2 + (1 - \alpha)\ \sigma_f^2 \qquad (4.5.3)$$

Now the measures of kurtosis for both the distributions F and G is given by,

$$\beta_2(g) = \mu_4 / \sigma_g^4 \qquad (4.5.4)$$

and

$$\beta_2(f) = \mu_4 / \sigma_f^4 \qquad (4.5.5)$$

Consider μ_4 of $h_\alpha(x)$ distribution ,it can be found as

$$\mu_4\ (h_\alpha) = \alpha\ \mu_4\ (g) + (1 - \alpha)\ \mu_4\ (f) \qquad (4.5.6)$$

Therefore we can write,

$$\mu_4 \ (h_\alpha) = \alpha\beta_2(g)\ \sigma_g^4 + (1 - \alpha)\ \beta_2(f)\sigma_f^4 \tag{4.5.7}$$

Therefore kurtosis of $h_\alpha(x) = \alpha\ g(x) + (1 - \alpha)\ f(x)$ is given by,

$$\beta_2(h_\alpha) = \frac{\alpha\ \beta_2(g)\ \sigma_g^4 + (1-\alpha)\ \beta_2(f)\ \sigma_f^4}{[\alpha\ \sigma_g^2 + (1-\alpha)\ \sigma_f^2]^2} \tag{4.5.8}$$

Here we are interesting finding α such that $\beta_2(h\alpha) = 3$.

Therefore from the equation (4.5.8)we can write,

$$\frac{\alpha\ \beta_2(g)\ \sigma_g^4 + (1-\alpha)\ \beta_2(f)\ \sigma_f^4}{[\alpha\ \sigma_g^2 + (1-\alpha)\ \sigma_f^2]^2} = 3 \tag{4.5.9}$$

equivalentaly,

$$\alpha\beta_2(g)\ \sigma_g^4 + (1 - \alpha)\ \beta_2(f)\sigma_f^4 = 3\left[\alpha\ \sigma_g^2 + (1 - \alpha)\ \sigma_f^2\right]^2$$

or

$$\alpha\beta_2(g)\ \sigma_g^4 + (1 - \alpha)\ \beta_2(f)\sigma_f^4 - 3\left[\alpha\ \sigma_g^2 + (1 - \alpha)\ \sigma_f^2\right]^2 = 0$$

Solving this equation and and by substituing $\Delta = \sigma_g^2/\sigma_f^2$ we get the quadratic form

$$3(1 - \Delta^2)\ \alpha^2 + \left[6(\Delta - 1) - \beta_2(g)\Delta^2 + \beta_2(f)\right]\alpha + \left[3 - \beta_2(f)\right] = 0 \tag{4.5.10}$$

At $\alpha = 0$ from the above equation we get,

$$[3 - \beta_2(f)] < 0. \qquad (4.5.11)$$

and for $\alpha = 1$, we have,

$$[3 - \beta_2(g)] > 0 \qquad (4.5.12)$$

Since the equation (4.5.10) is increasing, there is unique root α to the above equation in the interval (0,1).

Remark :

The mixing coefficient α in $h_\alpha(x)$ can be obtained from the (4.5.10) equation. For illuatration Kale and Sebastian (1996) give an example of double gamma probabiliyt density functions. Consider double gamma probability density function with

$$f_p(x) = [2\,\Gamma(p)]^{-1}|x|^{(p-1)} \exp\{-|x|\},\ x \in R_1,\ p > 0. \qquad (4.5.13)$$

Now $\beta_2(f_p)$ for thid pdf is given by,

$$\beta_2(f_p) = \mu_4/\mu^2{}_2 \qquad (4.5.14)$$

By our regular calculations we get,

$$\beta_2(f_p) = \frac{(3+p)\ (2+p)}{p(1+p)} \qquad (4.5.15)$$

For $\beta_2(f_p) = 3$ we get,

$$p = (\sqrt{13} + 1)/2. \tag{4.5.16}$$

The above pdf an be considered as a mixture of two gamma distributions in equal propertions, one on the positive side and one on the negative side with the same shape parameter p. Further Kale and Sebastian (1996) remarks that similar results can be obtained for the following pdfs:

$$m_p(x) = \frac{(p+1)}{2p} \{ 1 - |x|^p \}, \ |x| < 1, \ p > 0 \tag{4.5.17}$$

and

$$g_q(x) = \frac{(q-1)}{2} |x|^{-q}, \ |x| > 1, \tag{4.5.18}$$

For the probability density function given in equation (4.5.17) $\beta_2(m_p)$ is equal to 3 when $p = \sqrt{10} - 3$ and for the probability desity dunction given in (4.5.18) $\beta_2(g_q) = 3$ for $q = 3 + \sqrt{6}$.

* * * * *