Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 4 (2023)

Study On Moderate Distribution In The Statistical Framework

[1]Susheel Kumar, [2]Dr. Mahender Singh Poonia

[1]Research Scholar, Department of Mathematics, OM Sterling Global University, Hisar, Haryana, India
[2]Professor, Department of Mathematics, OM Sterling Global University, Hisar, Haryana, India

E-mail:[1] susheelkumar170593@gmail.com; [2] drmahender@osgu.ac.in

Abstract: This essay is a personal investigation into the origins, consequences, and significance of the concepts of "distribution" that we are working to instill in children. When creating learning experiences, these factors need to be in the back of our minds. The idea of "distribution" is formed as a filter through which statisticians view the variety in data. I investigate the causes of data variation, the distinction between empirical and theoretical distributions, the characteristics of statistical models, sampling distributions, the conditional nature of distributions used for modelling, and the foundations of inference.

Keywords: Moderate Distribution, Statistical Framework

1. Introduction:

There are so many measures of dispersion. Standard deviation is the most popular and most important among all these. But mean deviation is another useful and meaningful measure of dispersion. Mean deviation can be obtained in case of mean, median and made also. In this section we will study a distribution similar to normal distribution known as moderate distribution. Moderate distribution is a modified normal distribution in which mean is used as location parameter and mean deviation is used as scale parameter. We also study some properties of this modified or alternative distribution. We will show that the probability distribution in the defined distribution is actually moderate as compared to the previous distribution (normal distribution). Mean deviation is also the most likely and logically defined measure of dispersion which contains many inherently merits over the standard deviation. If it is possible to find an alternative distribution of normal distribution by taking means deviation as scalar parameter then we can get a sound alternative distribution further alongwith this sound alternative distribution we can find out alternatives of some other distribution like chi-saware distribution, F-distribution etc. which are related to normal distribution. When normality is considered as an assumption then there is a need of some moderation in allocation of chance so that extreme scores may get higher chance of occurrence. The normal distribution can be modified by taking high value of S. D. so that the curve of the modified distribution may be flatter as compared to the original distribution. We may have longer failed normal distribution which contains M. D. as scale parameter and moderateness in probability allocation. Increasing standard deviation also means automatically increasing in M. D. and the distribution becomes moderate distribution. Because the relation b/w M. D. and S. D. is

$$M.D.(r) = \sqrt{\frac{2}{\pi}\,\sigma^2} = \sqrt{\frac{2}{\pi}}\,\sigma$$

This is the moderate distribution the following objects are satisfied.

1. M. D. is taken as scaler parameter in place of S. D.
2. The values of 'α' and 'σ' become larger than their values in normal distribution.

So, in the moderate distribution 'h' is taken as mean and 'α' is taken as mean deviation. It may also call moderate normal distribution.

Definition: - Let us take a random variable 'x', then the probability density function of a distribution is given by

$$F(x) = \frac{1}{\pi\alpha} e^{\frac{-1}{\pi}\left(\frac{x-w}{\alpha}\right)^2}, \; -\infty < x < \infty,\; -\infty < h < \infty \qquad \alpha > 0$$

The random variable 'x' can be said to follow moderate distribution with parameter 'h' and 'α' and can be denoted by x ∞ M (h, α)

The above function contains the following properties: -

$\int_{-\infty}^{\infty} F(x)\ dx = 1$ (i)

Mean = E (x) = h (ii)

Mean Deviation about mean = E (x – h) = α (iii)

Standard Deviation (σ) = $\sqrt{\frac{\pi}{2}\ \alpha^2} = \sqrt{\frac{\pi}{2}}\alpha$ (iv)

Moment Generating Function = $M_x\ (p) = e^{hp+\frac{\pi}{4}\alpha^2 p^2}$ (v)

All negative values are similar to positive values mean

f (h – x) = f (h + x) (vi)

Note: - the relation b/w the 'σ' and 'α' for a lined FDD at some value 'M'

Let mean = h, FDD = 'M' then for normal distribution N (h, σ = M) and for moderate distribution Md (h, α = M)

We know that

$\alpha = \sqrt{\frac{2}{\pi}}\ \sigma = \sqrt{\frac{2}{\pi}}\ M$ for moderate distribution σ = M

And for normal distribution

$$\sigma = \sqrt{\frac{2}{\pi}}\ \alpha = \sqrt{\frac{2}{\pi}}\ M$$

So we can say in 'N' (h, σ = M)

$$\sigma = M\ \&\ \alpha = \sqrt{\frac{2}{\pi}}\ k$$

And in Md (h, α = k)

$\sigma = \sqrt{\frac{2}{\pi}}\ M$ and α = M

From the above relation we can say that in moderate distribution for FDD the values of 'α' and 'σ' are larger than their values in normal distribution i.e. σ increase from M to $\sqrt{\frac{2}{\pi}}\ M$ and 'α' increased from $\sqrt{\frac{2}{\pi}}\ M$ to M.

Now we do a comparison b/w standard normal distribution (SND) and standard moderate distribution (SMD).

Now we take h = 0 and M = 1 then in normal distribution we have

N (0, σ = 1) which implies σ = 1 & $\alpha = \sqrt{\frac{2}{\pi}}$ and in moderate distribution, we have

M (0, α = 1) then we have

$\sigma = \sqrt{\frac{2}{\pi}}\ M$ and α = 1

Again we find that in standard moderate distribution the values of 'α' and 'σ' are larger than their values are in standard normal distribution. i.e. 'σ' increases from 1 to $\sqrt{\frac{2}{\pi}}$ and 'α' has increase from $\sqrt{\frac{2}{\pi}}$ to 1. Thus we can say that M (0,1) is always higher than N (0, 1) in the first degree dispersion.

In the word 'moderate' the meaning of the word 'Normal' is already contain. The modified normal distribution is known as moderate distribution because the later probability contains the total area under the curve over the values of random variable. If FDD is fixed at some constant 'M' the moderate distribution Md (h, σ = M) have the standard deviation $\sigma = \sqrt{\frac{2}{\pi}}\ M$ is the normal distribution in which total area under the curve is modified by increasing the value of S. D. from 'M' to $\sqrt{\frac{2}{\pi}}M^2$ i.e. $\left[\left[\sqrt{\frac{\pi}{2}}-1\right]100\% \cong 25\%\right]$ it also means the value of mean deviation increased from $\sqrt{\frac{2}{\pi}}M^2$ to 'M' i.e. $\left[\left[1-\sqrt{\frac{\pi}{2}}\right]100\% \cong 20\%\right]$

From these values of random variable means the variable have high (or low) chance of occurrence in the N (h, σ = M) in relation to low (or high) chance of occurrence in Md (h, α = M). if we plot the graph of both distributions then the difference can be easily seen.

Diagram is pending

If we take some random variable 'X' and take some different ranges [like (01, 1), (1, 2), (-3, -2) etc.] then we find that in case of Md (0, α = 1) the variable has higher chance of occurrence than N (0, σ = 1) in case of all ranges except (-1, 1).

Standard Moderate Distribution: - The standard form of any distribution is the form that have location parameter zero and scale parameter is one. Here we study definition and some important properties of standard moderate distribution.

Let us consider random variable $x \sim M_d$ (h, r) then the variable z (standard moderate variable) is defined as z = $\frac{x-h}{r}$ contains the probability density function or

$$H(z) = \frac{1}{\pi}\, e^{\frac{-1}{\pi}(x-h)^2} = \frac{1}{\pi}\, e^{\frac{-1}{\pi}z^2}$$
$$where -\infty < z < \infty$$

The above distribution is called standard moderate distribution we can also write it as $z \sim M_d$ (0, 1)

2. Graphical Representation of SND and SMD

There are various properties which show the difference b/w SND and SMD probability curves as shown in the following diagram

Now we study some results from the graph.

1. At point x = 0, the value of standard normal distribution is f(x) = $\frac{1}{\sqrt{2\pi}}$ = 0.398 and the value of standard moderate distribution f(x) = $\frac{1}{\pi}$ = 0.318, so we can say that standard normal curve has higher keep than standard moderate curve.
2. Both the curves SNC and SMC interest at points (-1.1147, 0.2143) and (1.1147, 0.2143)
3. The area at these intersecting points are for SND

P (-1.1147 < x < 1.1147) = 0.735

And

For SMD

P (-1.1147 < x < 1.1147) = 0.626

Thus it can be said that for normal distribution 74% (approx.) values of standard normal variable full in this range while for standard moderate distribution 63% (approx.) values of standard normal variety fall in this range. So, it can be easily said that moderate distribution represents a moderate spread of values around mean and also have lower and broader peak.

2.1 Area under moderate curve and its properties: -

We know that some properties of normal and moderate distribution are similar.

i. Moderate distribution is also a unimodal distribution i. e. (mean = median = mode)
ii. Kurtosis: - $K_1 = h$, $K_2 = \frac{\pi}{2}r^2$, $K_3 = 0$, $K_4 = 0$
iii. B = 0, $B_2 = 3$
iv. Cumulative generating function is

$$C_x(t) = ht + \frac{\pi}{4}r^2t^2$$

v. Standard moderate curve is also bell shaped and symmetrical
vi. Characteristics function is

$$\phi_x(t) = e^{hit - \frac{\pi}{4} r^2 t^2}$$

vii. The p^{th} moment about mean is known as p^{th} degree dispersion.

For dispersion parameter the frequency in percentage is: -

Range	Frequency within Range (%)
	Moderate Distribution
$w \pm 1r$	57.506
$w \pm 2r$	88.945
$w \pm 3r$	98.331
$w \pm 4r$	99.993

Now if we calculate quartiles and deciles of N (0, 1) and M (0 1) then it can be seen that for M (0, 1) the quartile and decile are plotted at relatively distant points which represent relative even speed.

Quartile	N (0 1)	M (0 1)
Q_1	-0.674	-0.845
Q_2	Zero	Zero
Q_3	0.674	0.845

From the above table values, it is clear that probability distribution in SMC is more moderate than SNC, so moderate distribution can also be known as moderate normal distribution.

Decile	N (0 1)	M (0 1)
D_1	-1.281	-1.606
D_2	-0.841	-1.054
D_3	-0.524	-0.657
D_4	-0.253	-0.317
D_5	Zero	Zero
D_6	0.253	0.317
D_7	0.524	0.657
D_8	0.841	1.054
D_9	1.281	1.606

2.2 Limiting form of Moderate Distribution

i. Binomial Distribution Theorem: - If a random variable ‘x’ follows the binomial distribution having parameters ‘n’ and ‘p’ then moderate distribution is known as the limiting form of binomial distribution having parameter ‘h’ and mean deviation ‘r’

Proof: - The probability mass function of binomial distribution is given by: -

$P\,(x) = n_{c_x} p^x q^{n-x}$ (i)

$0 \leq x \leq n$

$0 \leq p \leq 1$

$Q = 1 - p$

Note: - Take x = r

The standard binomial variable is defined as: -

$$Z = \left[\frac{x - E(x)}{MD\,(x)}\right]$$

Where E(x) = np and MD (x) = $\sqrt{\frac{2}{\pi}}\, npq$

If ‘n’ is large, then

$Z = \frac{x - np}{\sqrt{\frac{2}{\pi}} npq}\, \sigma$ (ii)

$= -\sqrt{\frac{\pi}{2}\frac{np}{q}}$ If x = 0

and

$= \sqrt{\frac{\pi}{2}\frac{np}{q}}$ If x = n

Now if ‘n’ ⟶ ∞, ‘z’ have values form (-∞, ∞), thus the distribution ‘x’ is continuous over the range (-∞, ∞)

Now we apply stirlins formula on equation (i)

$$Lim\ p(x) = lim\left[\frac{\sqrt{2\pi}\ e^{-n}n^{n+\frac{1}{2}}p^{x}q^{n-x}}{\sqrt{2\pi}\ e^{-x}x^{x+\frac{1}{2}}\sqrt{2\pi}e^{-(n-x)}(n-x)^{n-x+\frac{1}{2}}}\right]$$

Note: - Take x = r

$$= Lim\ \frac{1}{\sqrt{2\pi}}\frac{1}{npq}\left(\frac{np}{r}\right)^{r+\frac{1}{2}}\left(\frac{nq}{n-r}\right)^{n-r+\frac{1}{2}}$$

Also from equation (ii) we have

X = np + z $\sqrt{\frac{2}{\pi}}\ npq$

$$\frac{x}{np} = 1 + z\sqrt{\frac{2}{\pi}\frac{q}{np}}$$

Then

$$n - x = nq - z\sqrt{\frac{2}{\pi}\ npq}$$

Also

$$dx = \sqrt{\frac{2}{\pi}\ npq\ dz}$$

So distribution of ‘z’ in the probability differential is

$$dF\ (z) = f(z)\ dz = \lim_{n\to\infty}\left[\frac{1}{\pi}\ \frac{1}{k}\right]\ dz$$

Where the value of ‘k’ is

$$k = \left(\frac{r}{np}\right)^{r+\frac{1}{2}}\left(\frac{n-r}{nq}\right)^{n-r+\frac{1}{2}}$$

To solve the above equation, take log both sides, then

$$Log\ k = \left(r + \frac{1}{2}\right)\log\left(\frac{r}{np}\right) + \left(n - r + \frac{1}{2}\right)\log\left(\frac{n-r}{nq}\right)$$

$$= \left[np + z\sqrt{\frac{2}{\pi}\ npq} + \frac{1}{2}\right]\log\left[1 + z\sqrt{\frac{2}{\pi}\frac{q}{np}}\right] + \left[nq + z\sqrt{\frac{2}{\pi}\ npq} + \frac{1}{2}\right]\ log\left[1 - z\sqrt{\frac{2}{\pi}\frac{p}{nq}}\right]$$

Where

$$z = \frac{x-h}{r} = \frac{x-np}{npq}$$

In a simple way we can write

$$Log\ k = \frac{\left(\frac{x-np}{npq}\right)^{2}}{\pi} + \left(Order\ in\ n^{-\frac{1}{2}}\right)$$

RHS of above expression shows a bounded constant, now take $\lim_{n\to\infty}$ *both sides*

$$\lim_{n\to\infty} Log\ k = \lim_{n\to\infty}\left[\frac{z^{2}}{\pi} + Order\ in\ n^{-\frac{1}{2}}\right]$$

$$= e^{\frac{z^{2}}{\pi}}$$

We put this value in probability differential and get

$$dF(z) = f(z)dz = \lim_{n\to\infty} \frac{1}{\pi} e^{\frac{1}{\pi}z^2} dz$$

So the probability density function is

$$f(z) = \frac{1}{\pi} e^{\frac{-1}{\pi}z^2}; -\infty < z < \infty$$

The above equation is known as pdf of standard moderate distribution with h = 0 & r = 1

Statement: - if x ~ p(x) then the limiting form of poison distribution is a moderate distribution with M (h, r)

Proof: - We know that poison distribution is a discrete distribution and its probability mass function is given by

$p(x) = \frac{e^{-\lambda}\lambda^x}{x}$ x = 0, 1, 2, 3, ………

λ > 0

= 0 otherwise

We know that mean and variance of poison distribution are same, so,

E(x) = λ and if n is large then

$MD\ (x) = \sqrt{\frac{2}{\pi}}\lambda$

Now let us study the standard poison variable

$$z = \frac{x - E\ (x)}{MD\ (x)} = \frac{x - \lambda}{\sqrt{\frac{2}{\pi}}\lambda}$$

If we calculate mean and mean deviation of above then

E (z) = 0 = mean & MD (z) = 1

Now the moment generating function of poison distribution is

$$M_x(x) = e^{\lambda e^x - \lambda}$$

Now we find moment generating function of standard poison variate i.e.

$$M_z(x) = E\left[e^x\left[\frac{x - x}{\frac{2}{\pi}\lambda}\right]\right] = e^{-x\sqrt{\frac{\lambda\pi}{2}}} M_x\left[\frac{x}{\sqrt{\frac{2\lambda}{\pi}}}\right]$$

$$= e^{-\lambda - x}\sqrt{\frac{\lambda\pi}{2}} + \lambda e\frac{x}{\sqrt{\frac{2\lambda}{\pi}}}$$

After simplification the result is

$$M_z(x) = e^{\frac{\pi}{4}x^2} + \frac{x^3}{6\left(\frac{2}{\pi}\right)^{\frac{3}{2}}\sqrt{\lambda}} + f\left(\lambda^{\frac{-1}{2}}\right)$$

Where $f\left(\lambda^{\frac{-1}{2}}\right)$ some function having parameter λ.

If λ ⟶ ∞ then the above equation becomes

$$M_z(x) = e^{\frac{\pi}{4}x^2}$$

Which is similar to the moment generating function of standard moderate variate.

2.3 Role of Goodness of Fit Test in Moderate Distribution: -

Some scientist has used Kolmogorov – Smirnov test to check if a sample has comes from the population that follows a specific distribution.

As we know that Kolmogorov – Smirnov test is based on the empirical distribution function. When 'N' ordered date points are given as z_1, z_2, ……… z_n then the empirical distribution function is defined as: -

$E_N = \frac{n_{(i)}}{N}$, where z_i are ordered from smallest to largest value and $n_{(i)}$ is the number of points less than z_i. The EDF is a step function that increases b_j 1/N.
Practically, the critical values of Kolmogorov – Smirnov test are determined by simulation under K – S test the hypothesis is taken as: -
H_0: null hypothesis: the data follow a specific distribution
H_1: alternative hypothesis: the data do not follow any specific distribution
The K – S test statistic is given as: -
$D = \max_{1 \le i \le N} \left| F\,(Z_i) - \frac{i}{N} \right|$ where 'F' is some fully specific theoretical cumulative distribution of the given specific distribution.
The above equation can also be written as: -

$$D = \max_{1 \le i \le N} \left| F\,(Z_i) - \frac{i}{N}, \frac{i}{N} - F\,(Z_i) \right|$$

Thus we can say that an upper bound is $\frac{i}{N}$ and for equal data the difference is less than the upper bound.

Now, to test whether the data is fitted to moderate distribution we generate some different random numbers.
Null Hypothesis: - The data is moderate distributed.
Alternative Hypothesis: - The data is not moderately distributed.
About the acceptance and rejection of null hypothesis we use SPSS software. We also check here that for different sample size whether the data approaches the moderate distribution or not by using the simulation technique.

Table 1: Goodness of fit by K – S test for moderate distribution

Mean	Mean Deviation	Sample size (H)	K – S Statistic	Level of Significance (P-Value)
5.7	1.44	50	0.07	0.2
		100	0.04	0.2
		500	0.039	0.06
		5000	0.009	0.2
		10000	0.005	0.2
8.8	1.84	50	0.06	0.2
		100	0.05	0.2
		500	0.019	0.2
		1000	0.019	0.2
		5000	0.007	0.2
		10000	0.006	0.2
23.1	5.58	50	0.14	0.009
		100	0.07	0.2
		500	0.028	0.2
		5000	0.01	0.14
		10000	0.5	0.2
25	3.59	50	0.09	0.2
		100	0.06	0.2
		500	0.032	0.2
		1000	0.014	0.2
		5000	0.009	0.2
		10000	0.007	0.2

3. Conclusion

In this paper a special norm of normal distribution i.e. moderate distribution has defined. In normal distribution we take mean and standard deviation is considered. Here we have proved that efficiency of mean deviation is better than the efficiency of standard deviation. Here we have studied that the shape of standard moderate distribution is more flat than the standard normal curve. The moderate curve is divided into 4 sigma limits and gives more accurate results in comparison of normal curve. By moderate distribution we come to know that mean deviation is also an standard measure of dispersion. Limiting case of binomial and poisson distribution are also studied in case of moderate distribution. Some theorems related to variance of independent variates are also studied. We have also discussed that if sample size vary then the moderate distribution is either monotonically increasing or decreasing. At last while applying goodness of fit test we have discussed about the acceptance of rejection at null hypothesis based on sample size. Central limit theorem, moment generating function, percentile etc. can also be discussed for moderate distribution and comparison can be done with normal distribution.

References

[1] Stacy, E. W. (1962). A generalization of the gamma distribution. The Annals of Mathematical Statistics, 33, 11871192.
[2] Jaggia, S. (1991). Specification tests based on the heterogeneous generalized gamma model of duration: with an application to Kennan's strike data. Journal of Applied Econometrics, 6, 169-180.
[3] Kiefer, N. M. (1984). Simple test for heterogeneity in exponential models of duration. Journal of Labour Economics, 539 549.
[4] Lancaster, T. (1979). Econometric methods for the duration of unemployment. Econometrica, 939 - 956.
[5] Sarmento, R. P. and Costa, V. (2019). An Overview of Statistical Data Analysis. Short studies, Research Gate, 1-29.
[6] Mishra, P., Pandey, C. M., Singh, U., Keshri, A., and Sabaretnam, M. (2019). Selection of appropriate statistical methods for data analysis. Annals of Cardiac Anaesthesia. 22(3), 297-301.
[7] Seufert, M. (2021). Statistical methods and models based on quality of experience distributions. Quality and User Experience, 6(3), 1-27.
[8] Savsani, M., and Ghosh, D. (2017). Bayesian inference for moderate distribution. Int. J. Agricult. Stat. Sci., 13(1), 303-317.
[9] Peligrad, M., Sang, H., Zhong, Y., and Wu, W. B. (2018). Exact Moderate and Large Deviations for Linear Processes. *Math. St*, 1-28.