Journal of Computer and Mathematical Sciences, *Vol.9(12),1890-1897 December 2018* *ISSN 0976-5727 (Print)*
(An International Research Journal), *www.compmath-journal.org* *ISSN 2319-8133 (Online)*

Variable Control Charts for Gumbel Distribution Based on Percentiles

K. Rosaiah*[1], B. Srinivasa Rao[2], J. Pratapa Reddy[3] and C. Chinnamamba[4]

[1]Department of Statistics,
Acharya Nagarjuna University, Guntur, INDIA.
[2]Department of Mathematics & Humanities,
R.V.R & J.C College of Engineering, Chowdavaram, Guntur, INDIA.
[3]Department of Statistics,
St. Ann's Women College, Guntur, INDIA.
[4]Department of Mathematics & Statistics,
P R Govt. College, Kakinada, INDIA.
[*]Corresponding author: rosaiah1959@gmail.com

(Received on: November 29, 2018)

ABSTRACT

A Variable quality characteristic is assumed to follow Gumbel Distribution. Based on the evaluated percentiles of sample statistic like mean, median, standard deviation, the control limits for the respective control charts are developed. The admissibility and power of the control limits is assessed in comparison with those based on the popular Shewart control limits.

Keywords: Most probable, equi-tailed, percentiles, Gumbel distribution.

1. INTRODUCTION

The well-known Shewart control charts are developed under the assumption that the quality characteristic follows a normal distribution. If $x_1, x_2, ..., x_n$ is a collection of observations of size *n* on a variable quality characteristic of a product and if t_n is a statistic based on this sample, the control limits of Shewart variable control chart are $E(t_n) \pm 3S.E(t_n)$. In quality control studies, data is always in small samples only. Since most of the distributions tend to normal distribution, it is taken as an alternative solution for all the distributions because of its central limit theorem. And if the data is assumed to follow normal distribution, the commonly used constants are Shewart control limits constants. Even if a skewed data which follows Gumbel distribution, it is not advisable to apply the Shewart constants. An alternative procedure is to be adopted. Therefore if the population is not normal there is a need to develop a separate procedure for the construction of control limits. In this paper we assume that the quality variate follows gumbel model and develop control limits for such a data on par with the presently available control limits. If a process quality characteristic is assumed to follow gumbel distribution the online process of such a quality can be controlled through the theory of Gumbel distribution. In the absence of any such specification of the population model we generally use the normal distribution and the associated constants available in all standard text books of statistical quality control. However, normality is only an assumption that is rarely verified and found to be true. Unless the sample is very large in size this assumption may not be taken for granted without proper goodness of fit test procedure. At the same time central limit theorem cannot be made use of, because central limit theorem gives only asymptotic normality for any statistic. Therefore, if a distribution other than normal is a suitable model for a quality variate, separate procedures are to be developed. We present the construction of control charts when the process variate is assumed to follow Gumbel distribution.

The probability density function (pdf) of Gumbel distribution (GD) is given by

$f(x)=\beta^{-1} e^{-(x-\mu)/\beta} \exp[-e^{-(x-\mu)/\beta}]; -\infty<x<\infty, -\infty<\mu<\infty, \beta > 0$ (1.1)

Its cumulative distribution function (cdf) is

$F(x)=\exp[-e^{-(x-\mu)/\sigma}]$ (1.2)

Gumbel distribution (GD) is a skewed, unimodal distribution on the positive real line. The distributional properties are

Mean=$\mu+\beta\gamma$ where γ is Euler-Mascheroni constant (1.3)

Median=$\mu-\beta \ln(\ln(2))$ (1.4)

Mode=μ (1.5)

Variance=$(\pi^2/6) \beta^2$ (1.6)

Development of statistical quality control methods skewed distributions are attempted by many authors. Some of them are Edgeman (1989) Inverse Gaussion Distribution, Gonzalez and Viles (2000) Gamma Distribution, Kantam and Sriram (2001) Gamma Distribution, Chan and Cui (2003) have developed control chart constants for skewed distributions where the constants are dependent on the coefficient of skewness of the distribution, Kantam *et al.*, (2006) Log logistic Distribution, Betul and Yaziki (2006) Burr Distribution, Subba Rao and Kantam (2008) Double exponential distribution, Kantam and Srinivasa Rao (2010) control charts for process variate, Srinivasa Rao and Sarath Babu (2012) Linear failure rate distribution, Srinivasa Rao and Kantam (2012) Half logistic distribution and references there in. GD is another situation of skewed distribution that was not paid much attention with respect to development of control charts. At the same time it is one of the probability models applicable for life testing and reliability studies. Accordingly, if a lifetime data is considered as a quality data, development of control charts for the same is desirable for the use by practitioners. Since GD is a skewed distribution, this paper makes an attempt to study in a comparative manner. The rest of the paper is organized as follows. The basic theory and the development of control charts for the statistics- mean, median, and standard deviation are presented in Section 2. The comparative study of the developed control limits in relation to Shewart limits is given in Section 3. Summary and conclusions are given in Section 4.

2. CONTROL CHART CONSTANTS THROUGH PERCENTILES

2.1 Mean Chart

Let $x_1, x_2, \ldots\ldots, x_n$ be a random sample of size *n* drawn from GD with μ=0 and σ=1. If this is considered as a subgroup of an industrial process data with a targeted population average, under repeated sampling the statistic $\bar{x}$ gives whether the process average is around the targeted mean or not. Statistically speaking, we have to find the 'most probable' limits within which $\bar{x}$ falls. Here the phrase 'most probable' is a relative concept which is to be considered in the population sense. As the existing procedures are for normal distribution only, the concept of 3σ limits is taken as the 'most probable' limits. It is well known that 3σ limits of normal distribution include 99.73% of probability. Hence, we have to search two limits of the sampling distribution of sample mean in GD such that the probability content of those limits is 0.9973. Symbolically, we have to find L, U such that

$$P(L \leq \bar{x} \leq U) = 0.9973 \quad (2.1.1)$$

where $\bar{x}$ is the mean of sample size n. Taking the equi-tailed concept L, U are respectively 0.00135 and 0.99865 percentiles of the sampling distribution of $\bar{x}$. We resorted to the empirical sampling distribution of $\bar{X}$ through simulation there by computing its percentiles. These are given in Table 2.1

Table 2.1: Percentiles of Mean in GD

n	0.99865	0.99	0.975	0.95	0.05	0.025	0.01	0.00135
2	4.49294	3.28156	2.77994	2.34975	0.21687	0.15784	0.1001	0.03733
3	3.55494	2.80484	2.43263	2.11555	0.31479	0.24256	0.17313	0.09505
4	3.12189	2.5015	2.16837	1.91927	0.38656	0.31662	0.25193	0.13846
5	2.83479	2.29378	2.02514	1.81918	0.43454	0.36449	0.28237	0.2064
6	2.68648	2.19689	1.96668	1.7675	0.4792	0.41347	0.33679	0.2426
7	2.46582	2.10638	1.87711	1.70211	0.50425	0.4381	0.37656	0.2759
8	2.36053	1.97329	1.78516	1.62555	0.53406	0.46935	0.39332	0.29866
9	2.28933	1.91166	1.73221	1.58784	0.55643	0.49979	0.43488	0.30828
10	2.23825	1.88341	1.70924	1.57272	0.57108	0.51194	0.43691	0.32988

The percentiles in the above table are used in the following manner to get the control limits for sample mean. From the distribution of $\bar{x}$, consider

$$P(Z_{0.00135} \leq \bar{x}_i \leq Z_{0.99865}) = 0.9973 \quad (2.1.2)$$

But $\bar{X}$ of sampling distribution when the mean is γ, where $\gamma \approx 0.5772$ and $\beta = 1$ for GD From equation (2.1.2) over repeated sampling, for the ith subgroup mean we can have

$$P\left(Z_{0.00135}\frac{\bar{\bar{x}}}{0.5772} \le \bar{x}_i \le Z_{0.99865}\frac{\bar{\bar{x}}}{0.5772}\right) = 0.9973 \tag{2.1.3}$$

This can be written as

$$p\left(A_{2p^*}\bar{\bar{x}} \le \bar{x}_i \le A_{2p^{**}}\bar{\bar{x}}\right) = 0.9973 \tag{2.1.4}$$

where $\bar{\bar{x}}$ is a grand mean, $\bar{x}_i$ is i[th] subgroup mean, $A_{2p^*} = \frac{Z_{0.00135}}{0.5772}$, $A_{2p^{**}} = \frac{Z_{0.99865}}{0.5772}$.

Thus A_{2p^*}, $A_{2p^{**}}$ are the percentile constants of $\bar{x}$ chart for GD are given in Table 2.2

Table 2.2: Percentiles constants of $\bar{x}$-chart

n	$A_{2p^{**}}$	A_{2p^*}
2	7.784026334	0.064674
3	6.158939709	0.164674
4	5.408679834	0.239882
5	4.911278586	0.357588
6	4.654331254	0.420305
7	4.272037422	0.477997
8	4.089622315	0.517429
9	3.966268191	0.534096
10	3.877772003	0.571518

2.2 Median- chart

We have to search for two limits of the sampling distribution of sample median in GD such that the probability content of these limits is 0.9973. Symbolically, we have to find L, U such that

$$p\left(L \le m_i \le U\right) = 0.9973 \tag{2.2.1}$$

where m_i is the median of sample size n. Through simulation, the percentiles observed are given in the Table 2.3

Table 2.3 : Percentiles of Median in GD

n	0.99865	0.99	0.975	0.95	0.05	0.025	0.01	0.00135
2	4.51438	3.34557	2.78541	2.36693	0.22174	0.15706	0.09739	0.0315
3	3.66287	2.84391	2.31339	1.90947	0.18735	0.13193	0.08449	0.02232
4	2.96377	2.38903	2.02779	1.72377	0.27399	0.21349	0.15117	0.07817
5	2.84802	2.17209	1.86293	1.61718	0.26616	0.20882	0.15367	0.06945
6	2.34899	1.90425	1.66053	1.48187	0.32487	0.26343	0.19804	0.10901
7	2.43683	1.89707	1.64797	1.44212	0.31518	0.25841	0.20538	0.10799
8	2.24865	1.75012	1.54627	1.3702	0.34838	0.29114	0.23052	0.15505
9	2.17742	1.72561	1.49229	1.33545	0.35316	0.29714	0.23462	0.14736
10	1.93656	1.59428	1.41709	1.28948	0.38201	0.32823	0.27153	0.16096

The percentiles in the above table are used in the following manner to get the control limits for median. From the distribution of m, consider

$$p\left(Z_{0.00135} \leq m_i \leq Z_{0.99865}\right) = 0.9973 \quad (2.2.2)$$

But median of sampling distribution when μ-β ln(ln(2)) is 0.3665 for GD. From the equation (2.2.2) over repeated sampling, for i[th] subgroup median, we can have

$$P\left(Z_{0.00135}\frac{\bar{m}}{0.3665} \leq m_i \leq Z_{0.99865}\frac{\bar{m}}{0.3665}\right) = 0.9973 \quad (2.2.3)$$

This can be written as

$$p\left(A_{7p^*}\bar{m} \leq \bar{m}_i \leq A_{7p^{**}}\bar{m}\right) = 0.9973 \quad (2.2.4)$$

where $\bar{m}$ is mean of subgroup medians. Thus $A_{7p^*} = \frac{Z_{0.00135}}{0.3665}$, $A_{7p^{**}} = \frac{Z_{0.99865}}{0.3665}$ are the percentile constants of $\bar{m}$ chart for GD and given in Table 2.4

Table 2.4 Percentile constants of median chart

n	$A_{7p^{**}}$	A_{7p^*}
2	12.31754434	0.085948
3	9.994188267	0.0609
4	8.086684857	0.213288
5	7.770859482	0.189495
6	6.409249659	0.297435
7	6.648922237	0.294652
8	6.135470668	0.423056
9	5.94111869	0.402074
10	5.283929059	0.439181

2.3 Standard deviation Chart

We have to search two limits of the sampling distribution of sample standard deviation in GD such that the probability content of these limits is 0.9973. Symbolically, we have to find L, U such that

$$P(L \leq s \leq U) \quad (2.3.1)$$

where s_i is the standard deviation of sample size *n*. Through simulation the percentiles obtained and are given in the Table 2.5

Table 2.5: Percentiles of standard deviation in GD

n	0.99865	0.99	0.975	0.95	0.05	0.025	0.01	0.00135
2	5.37205	3.3453	2.61989	2.11406	0.03885	0.02096	0.00639	0.00076
3	3.82928	2.83337	2.34875	1.96658	0.13391	0.09662	0.06146	0.02402
4	3.62724	2.56152	2.15643	1.83734	0.2161	0.16586	0.12321	0.06485
5	3.30895	2.49537	2.10329	1.79416	0.255	0.20562	0.16226	0.08895
6	3.00897	2.34808	2.03393	1.76397	0.30581	0.25261	0.19728	0.12805
7	3.09693	2.29428	1.98778	1.7232	0.32835	0.28134	0.23774	0.16448
8	2.8583	2.1975	1.90033	1.67033	0.35748	0.30651	0.24996	0.19047
9	2.78492	2.18342	1.86648	1.65392	0.381	0.32981	0.28698	0.20518
10	2.67594	2.09922	1.82671	1.61967	0.40169	0.34807	0.29902	0.21765

The percentiles from the above table are used in the following manner to get the control limits for sample standard deviation. From the distribution of *s*, consider

$$P(Z_{0.00135} \leq s \leq Z_{0.99865}) = 0.9973 \qquad (2.3.2)$$

But S.D of sampling distribution when standard deviation is 1.2825 for GD. From equation (2.3.2), for the ith subgroup standard deviation we can have

$$P\left(Z_{0.00135} \frac{\bar{s}}{1.2825} \leq s_i \leq Z_{0.99865} \frac{\bar{s}}{1.2825}\right) = 0.9973 \qquad (2.3.3)$$

This can be written as

$$P(B^*_{3p} \times \bar{s} \leq \bar{s}_i \leq B^{**}_{4p} \times \bar{s}) = 0.9973 \qquad (2.3.4)$$

where $\bar{s}$ is mean of standard deviation, s_i is ith subgroup Standard deviation. Thus $B_{3p^*} = \frac{Z_{0.00135}}{1.2825}$, $B_{4p^{**}} = \frac{Z_{0.99865}}{1.2825}$ are the percentile constants of $\bar{s}$ chart for GD are given in Table 2.6

Table 2.6: Percentile constants of S.D-Chart

n	$B_{4p^{**}}$	B_{3p^*}
2	4.188732943	0.000593
3	2.985793372	0.018729
4	2.82825731	0.050565
5	2.580077973	0.069357
6	2.346175439	0.099844
7	2.414760234	0.12825
8	2.228693957	0.148515
9	2.171477583	0.159984
10	2.086502924	0.169708

3. COMPARATIVE STUDY

The control chart constants for the statistics mean, median, and standard deviation developed in Section 2 are based on the population described by GD. In order to use this for a data, the data is confirmed to follow GD. Therefore, the power of the control limits can be assessed through their application for a true GD data in relation to the application of Shewart limits. With this back drop we have made this comparative study simulating random samples of size n = 2(1) 10 from GD and calculated the control limits using the constants of Section 2 for mean, median, and standard deviation in succession. The number of statistic values that have fallen within the respective control limits is evaluated and is named as GD coverage probability. Similar count for control limits using Shewart constants evaluable in quality control manuals are also calculated. These are named as Shewart coverage probability. The coverage probabilities under the two schemes namely true GD, Shewart limits are presented in the following Tables 3.1, 3.2, and 3.3.

Table 3.1: Coverage Probabilities of Mean-Chart

	Shewart limits			**Percentile limits**		
n	$\bar{x}$ + $A_2\bar{R}$	$\bar{x}$ - $A_2\bar{R}$	Count	A_{2p}** $\bar{\bar{x}}$	A_{2p}* $\bar{\bar{x}}$	Count
2	2.824364	-0.79533	0.9770	7.8970193	0.065613102	0.9954
3	2.521127	-0.47878	0.9802	6.2893491	0.168161103	0.9909
4	2.301005	-0.26827	0.9819	5.4972091	0.243808582	0.9913
5	2.18805	-0.15303	0.9854	4.9972800	0.363850087	0.9853
6	2.083538	-0.0515	0.9841	4.7288843	0.427037365	0.9733
7	2.020824	0.009402	0.9861	4.3366007	0.4852212	0.9603
8	1.953069	0.073959	0.9888	4.1448895	0.524421502	0.9565
9	1.908033	0.120591	0.9896	4.0230334	0.541739611	0.9573
10	1.878877	0.157587	0.9896	3.9484715	0.581937582	0.9436

Table 3.2: Coverage Probabilities of Median-Chart

	Shewart limits			**percentile limits**		
n	$\bar{m}$ + $A_7\bar{R}$	$\bar{m}$ – $A_7\bar{R}$	Count	A_{7p}** $\bar{m}$	A_{7p}* $\bar{m}$	Count
2	2.833286	-0.7969	0.9414	12.54166206	0.087511985	0.9919
3	2.335323	-0.61251	0.8398	8.60903375	0.052459856	0.9955
4	2.141508	-0.42679	0.7401	6.933183955	0.182864051	0.9831
5	1.982656	-0.34359	0.6365	6.368460242	0.155297211	0.9897
6	1.877977	-0.25226	0.5542	5.209797041	0.241771985	0.9808
7	1.790386	-0.20432	0.4656	5.2728081	0.233668556	0.9835
8	1.73555	-0.15304	0.3935	4.854721844	0.334745123	0.9571
9	1.682925	-0.11753	0.3252	4.650095776	0.314701855	0.9692
10	1.636247	-0.06723	0.2764	4.145284618	0.344541358	0.9682

Table 3.3: Coverage Probabilities of S.D-Chart

	Shewart limits			**Percentile limits**		
n	B_4 X$\bar{s}$	B_3 X$\bar{s}$	Coverage probability	B_{4p}**X $\bar{s}$r	B_{3p}*X $\bar{s}$	Coverage probability
2	2.833286	-0.7969	0.3823	4.277546648	0.000605157	0.9973
3	2.335323	-0.61251	0.4113	3.050970256	0.019137881	0.9944
4	2.141508	-0.42679	0.4284	2.875067797	0.051402208	0.9976
5	1.982656	-0.34359	0.4313	2.604900903	0.070024006	0.9964
6	1.877977	-0.25226	0.4718	2.388350288	0.101638851	0.9963
7	1.790386	-0.20432	0.5580	2.450645986	0.130155429	0.9985
8	1.73555	-0.15304	0.6522	2.264444437	0.150896943	0.9971
9	1.682925	-0.11753	0.7265	2.199480958	0.162047564	0.9984
10	1.636247	-0.06723	0.7910	2.116654978	0.172160047	0.9977

4. SUMMARY AND CONCLUSIONS

In most of the quality control applications, the data is assumed to follow normal distribution and the Shewart constants are used rigorously. These tables show that for a true GD if the Shewart limits are used in a mechanical way it would result in less confidence coefficient about the decision of process variation for mean, median, and standard deviation charts. Hence if a data is confirmed to follow GD, the usage of Shewart constants is all the above charts is not advisable and exclusive evaluation and application of GD constants is preferable in statistical quality control.

REFERENCES

1. Betul Kan and Berna Yaziki. The individual control charts of BURR distributed data, *Proceedings of the ninth WSEAS international conference on applied mathematics*, Istanbul, Turkey, Pages 645-649, (2006).
2. Chan.Lai.K. and Heng.J.Cui Skewness Correction xbar and R-Charts for Skewed distributions, *Naval Research Logistics*, 50:1-19, (2003).
3. Edgeman.R.L. Inverse Gaussian control charts, *Australian Journal of Statistics*, 31(1): 435-446, (1989).
4. Gonzales.I. and Viles.I. Semi-Economic design of mean control charts assuming Gamma distribution, *Economic Quality Control*, 15:109-118, (2000).
5. Kantam.R.R.L. and Sriram.B. Variable control charts based on Gamma distribution, *IAPQR Transactions*, 26(2):63-67, (2001).
6. Kantam.R.R.L., Vasudeva Rao.A. and Srinivasa Rao.G. Control charts for Log, Logistics Distribution, *Economic Quality Control*, 21(1):77-86, (2006).
7. Kantam.R.R.L. and Srinivasa Rao.B. An improved dispersion control charts for process variate, *International Journal of Mathematics and Applied Statistics*, 1(1):19-24, (2010).
8. Srinivasa Rao. B. and Sarath Babu.G., Variable Control Charts based on Linear Failure Rate Model, *International Journal of statistics and Systems*, 7(3):331-341, (2012).
9. Srinivasa Rao. B. and Kantam. R.R.L. Mean and Range Charts for Skewed distributions-A comparison based on half logistic distribution, *Pakistan Journal of Statistics*, 28(4):437-444, (2012).
10. Subba Rao. R. and Kantam. R.R.L. Variable control charts for process mean with reference to double exponential distribution, *Acta. Cinica Indica*, 34(4):1925-1930, (2008).