Journal of
INDUSTRIAL
TECHNOLOGY

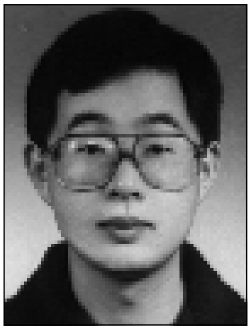
Properties Of The Half-Normal Distribution And Its Application To Quality Control

By
Dr. Chao-Yu Chou and Dr. Hui-Rong Liu

The Official Electronic Publication of the National Association of Industrial Technology

Properties Of The Half-Normal Distribution And Its Application To Quality Control

Dr. Chao-Yu Chou and Dr. Hui-Rong Liu



Chao-Yu Chou is an associate professor in the Department of Industrial Engineering and Management, National Yunlin University of Science & Technology, Touliu, Taiwan 640, R.O.C. He received a B.S. degree from Tunghai University, a M.S. and a Ph.D. from Auburn University. His research interests include applied statistics, quality management, and quality engineering.

Hui-Rong Liu is an associate professor in the Department of Foods and Nutrition, Hung-Kuang Institute of Technology, Shalu, Taiwan 433, R.O.C. She received a B.S. degree from Taipei Medical College, a M.S. from Iowa State University, and a Ph.D. from Auburn University. Her research interests include oil chemistry, food quality control, and consumer preference test.

Introduction

The purpose of industrial process control is to manufacture products whose quality is designed and maintained at lowest possible cost, while providing full customer satisfaction. Statistical methods applied to industrial process control have received much attention since the 1930s (Simon, 1949). One of these methods is the control chart technique. A typical control chart usually consists of limiting values called the lower control limit (LCL) and upper control limit (UCL). Between the LCL and the UCL is a center line, which represents the midpoint between the lower and upper control limits. Samples are taken in order of time, and a sample statistic is plotted on the control chart. A major function of control charting is to detect the occurrence of assignable causes so that the necessary corrective action can be taken before a large quantity of nonconforming product is manufactured. Actually, the control chart technique is a graphical representation of statistical hypothesis testing.

When dealing with a quality characteristic which is a numerical measurement, it is a standard practice to control both the mean of the quality characteristic and its variability (Montgomery, 1991). That is, accuracy and precision of selected properties of the product judge quality. Accuracy is measured in terms of the proximity of average measured values to target values and precision is measured in terms of the variability of measured values. In general, accuracy is controlled through the control chart for means (i.e., the X-bar chart) and the control chart for ranges (i.e., the R chart) usually monitors precision. Therefore, for a specific quantitative quality characteristic, two control charts (e.g., the X-bar and R charts) are usually needed. An important part of control charting is defining the control limits for the X-bar and R charts. The control limits can be set up by the quality engineers and are related to the Type I error probability (denoted by a) of the corresponding chart or hypothesis testing. Note that the Type I error of a control chart means that the process is concluded to be out of control when the process is really in control. By defining appropriate control limits, the quality engineers could have a good control on the false alarm rate of the control chart. Usually 3-sigma control limits are used in the X-bar and R charts (Grant and Leavenworth, 1995). "Sigma" here is referred to the standard deviation of the corresponding sample statistic, and is not the population standard deviation. If the underlying distribution of the quality characteristic is normally distributed, then the sampling distribution of the X-bar is also normally distributed. Thus, the 3-sigma control limits of an X-bar chart will result in its Type I error probability a=0.0027, and consequently its average run length (ARL) is approximately 370 when the process is in control. However, this is not the case for the R chart. Since the sampling distribution of the range (R) is not symmetrical even when the underlying distribution of the quality characteristic is normally distributed, quality engineers might experience some trouble determining the a value or ARL of an R chart when the 3-sigma control limits are used. Although Chou (1994) studied the statistical properties of R by using numerical analysis method, it is still not convenient to employ this method in on-line management.

This article proposes the half-normal distribution method (HNDM) to control both the process central tendency and variability simultaneously. Parker et. al. (1994) used the absolute value of the deviation from targets, which is similar to the half-normal random variant, to control the quality of construction materials, such as hot mix asphalt. The HNDM provides an alternative to reduce the quality control process into only one chart and consequently avoid the problem of the a value (or ARL) in an R chart. In this article, the properties of the half-normal distribution will be studied. Then, the application of the HNDM to quality control will be illustrated, and finally a case study will be given to compare the HNDM with the traditional X-bar and R charts.

The random variables (r.v.) and their definitions used in this article are listed as follows:

Z a standardized normal r.v. (mean $\mu=0$ and variance $\sigma^2=1$)

X a normal r.v. with mean μ and variance σ^2

Y absolute value of Z (i.e., $Y = |Z|$), or called half-normal r.v.

L(n) mean of n absolute standardized normal r.v's, i.e.,

$$L(n) = (Y_1 + Y_2 + \ldots + Y_n)/n, \text{ where } Y_i = |Z_i|$$

The Half-Normal Distribution

Let Y be an absolute value of the standardized normal random variate. The probability density function of Y is (Hogg and Tanis, 1993).

$$f(y) = \frac{2}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \qquad y \geq 0$$

The mean of Y can be found by integrating $y \cdot f(y)$ over its range from 0 to ∞. The result is $E(Y) = \mu_Y = 0.79788456$. The expectation of Y^2 can be found by integrating $y^2 \cdot f(y)$ from 0 to ∞, and the result is $E(Y^2) = 1$. Then, the variance of Y is:

$$V(Y) = E(Y^2) - (\mu_Y)^2 = 1 - (2/\pi)$$

The standard deviation of Y, denoted by σ_Y, is 0.602810275. The cumulative distribution function of Y can be obtained as follows:

$$F_Y(a) = P(Y \leq a) = P(-a \leq Z \leq a) = 2 \cdot \Phi(a) - 1 \quad (1)$$

where $\Phi(a) = P(Z \leq a)$.

Let L(n) be the mean of n Y_i's, where $Y_i = |Z|$. From statistical theory (Hool and Maghsoodloo, 1980), the mean of L(n) is the same as E(Y). That is,

$$E[L(n)] = \mu_Y = 0.79788456 \qquad (2)$$

The variance of L(n) is equal to:

$$V[L(n)] = V(Y) / n = 0.363380227 / n \quad (3)$$

The standard deviation of L(n), denoted by $\sigma_{L(n)}$, is found by taking the square root on Equation (3). To use L(n) as a sample statistic in control charting, its distribution function should be obtained first. However, the probability density function of L(n) is intractable mathematically. Therefore, simulation technique is used to obtain the probability distribution of L(n) on the scale of 0.01. The simulation procedure (in the

continued on page 6

(a) n=5

a	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0									0	.0001
0.1	.0001	.0002	.0002	.0003	.0004	.0006	.0007	.0009	.0013	.0017
0.2	.0023	.0030	.0037	.0047	.0057	.0068	.0082	.0099	.0118	.0137
0.3	.0157	.0184	.0209	.0236	.0270	.0305	.0345	.0385	.0432	.0484
0.4	.0537	.0597	.0660	.0728	.0801	.0870	.0948	.1032	.1118	.1211
0.5	.1304	.1405	.1504	.1613	.1721	.1832	.1945	.2064	.2181	.2298
0.6	.2428	.2559	.2692	.2834	.2980	.3120	.3260	.3408	.3559	.3702
0.7	.3847	.3993	.4142	.4288	.4439	.4589	.4740	.4893	.5047	.5191
0.8	.5336	.5476	.5614	.5761	.5898	.6035	.6172	.6308	.6436	.6568
0.9	.6695	.6819	.6939	.7059	.7173	.7283	.7398	.7500	.7610	.7715
1.0	.7816	.7910	.8003	.8093	.8182	.8271	.8348	.8429	.8507	.8573
1.1	.8647	.8716	.8779	.8840	.8901	.8957	.9016	.9065	.9118	.9167
1.2	.9212	.9258	.9304	.9343	.9383	.9420	.9455	.9488	.9518	.9546
1.3	.9575	.9601	.9628	.9652	.9675	.9695	.9713	.9732	.9751	.9769
1.4	.9783	.9801	.9813	.9827	.9840	.9854	.9863	.9874	.9882	.9892
1.5	.9898	.9904	.9911	.9917	.9923	.9928	.9934	.9939	.9943	.9948
1.6	.9952	.9956	.9959	.9963	.9966	.9968	.9972	.9974	.9977	.9978
1.7	.9980	.9982	.9984	.9986	.9987	.9988	.9989	.9990	.9991	.9992
1.8	.9992	.9993	.9994	.9995	.9995	.9996	.9996	.9997	.9997	.9998
1.9	.9998	.9998	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999
2.0	1									

(b) n=10

a	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.2				0	.0001	.0001	.0002	.0003	.0005	.0007
0.3	.0009	.0012	.0016	.0021	.0027	.0034	.0043	.0053	.0066	.0079
0.4	.0096	.0118	.0140	.0169	.0199	.0236	.0275	.0317	.0369	.0425
0.5	.0486	.0554	.0627	.0707	.0799	.0891	.0995	.1109	.1218	.1342
0.6	.1481	.1623	.1771	.1926	.2093	.2263	.2440	.2621	.2802	.2994
0.7	.3199	.3387	.3591	.3794	.3999	.4210	.4418	.4627	.4836	.5052
0.8	.5263	.5464	.5661	.5865	.6067	.6264	.6452	.6634	.6816	.6987
0.9	.7150	.7320	.7475	.7627	.7777	.7922	.8055	.8186	.8307	.8423
1.0	.8532	.8634	.8736	.8827	.8914	.8997	.9075	.9154	.9226	.9289
1.1	.9348	.9405	.9456	.9503	.9550	.9590	.9627	.9663	.9694	.9721
1.2	.9748	.9771	.9794	.9816	.9837	.9854	.9868	.9882	.9893	.9906
1.3	.9917	.9925	.9934	.9941	.9949	.9956	.9961	.9966	.9971	.9974
1.4	.9977	.9980	.9983	.9985	.9986	.9988	.9990	.9991	.9992	.9993
1.5	.9994	.9995	.9996	.9996	.9997	.9997	.9998	.9998	.9998	.9999
1.6	.9999	.9999	.9999	1						

(c) n=15

a	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.2									0	.0001
0.3	.0001	.0001	.0001	.0002	.0002	.0003	.0004	.0006	.0009	.0012
0.4	.0016	.0020	.0029	.0038	.0049	.0064	.0082	.0104	.0132	.0162
0.5	.0196	.0235	.0285	.0336	.0399	.0469	.0549	.0639	.0737	.0847
0.6	.0961	.1094	.1237	.1388	.1550	.1728	.1913	.2112	.2313	.2526
0.7	.2750	.2986	.3222	.3463	.3706	.3961	.4214	.4473	.4723	.4976
0.8	.5238	.5486	.5735	.5983	.6228	.6470	.6695	.6918	.7135	.7343
0.9	.7533	.7715	.7891	.8061	.8222	.8373	.8520	.8650	.8769	.8889
1.0	.8993	.9093	.9186	.9268	.9342	.9409	.9469	.9527	.9583	.9634
1.1	.9674	.9712	.9745	.9778	.9804	.9829	.9851	.9870	.9888	.9906
1.2	.9919	.9930	.9939	.9948	.9957	.9963	.9968	.9973	.9977	.9980
1.3	.9983	.9985	.9987	.9990	.9991	.9992	.9993	.9994	.9995	.9996
1.4	.9997	.9998	.9998	.9999	.9999	.9999	1			

Table 1. The Cumulative Distribution of L(n), (P[L(N) $\leq$ a])

continued from page 5

example of n = 3) is summarized as follows:

Step 1. Generate 3 standardized normal random variates by using Polar algorithm (Banks and Carson, 1984).
Step 2. Take absolute values of the 3 normal random variates.
Step 3. Obtain one L(n) by averaging the 3 absolute normal random variates.
Step 4. Determine the interval in which this L(n) is located. (L(n) is on the scale of 0.01.)
Step 5. Repeat Step 1 to Step 4 a hundred thousand times to generate 100,000 L(n)'s.
Step 6. Count the number in each interval, and then, divide this number by 100,000 to get the probability value for each interval.

Table 1 (page 5) shows some resulting cumulative distributions of L(n) for n = 5, 10, and 15.

From Table 1(a), for example, P[L(5) £ 1] = 0.7816 and P[L(5) £ 1.66] = 0.9972. These simulation results permit the applications of quality control analysis, which will be illustrated in the next section. The accuracy of simulation depends on the number of random variates generated and the random number seed. Table 2 compares the mean and variance of L(n) from the simulated results and the theoretical values in equations (2) and (3). It can be observed that the differences between the theoretical and simulated values are negligible.

Application to Quality Control

Assuming the underlying distribution of the quality characteristic is normally distributed, it is essential to determine the values of m and s. If there is a target value for mean, this target value will be used as m; otherwise, the mean value needs to be estimated by the sample average from historical data. Since s is usually unknown, historical data are also required to estimate s. There are many estimators available for s. For example, one widely used estimator of s, presented by Nelson (1990), is:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{m} (n_i - 1) S_i^2}{N - m}}$$

where N is the total number of measurements, m is the number of samples, n_i is the sample size of the i^th sample, and S_i is the sample standard deviation of the i^th sample. Once m and s are determined, the value of L(n) can easily be calculated from each sample.

An important part of control charting is the control limits. The control limits are determined by a, the Type I error probability. The Type I error probability is the probability that the process is concluded to be out of control when the process is really in control. One can define the 1-a quantile of L(n), denoted by $L(n)_{1-a}$, as:

$$P \{ L(n) £ L(n)_{1-a} \} = 1-a$$

Thus, as n = 5, from Table 1(a), $L(n)_{.9545}$ = 1.2896, $L(n)_{.9876}$ = 1.4725, and $L(n)_{.9973}$ = 1.6650. (These quantiles are obtained by interpolation.) Note that $L(n)_{.9545}$, $L(n)_{.9876}$ and $L(n)_{.9973}$ correspond to the 2s, 2.5s and 3s two-sided control limits of an X-bar chart. Table 3 (page ?) lists several important and useful quantiles of L(n) for various sample sizes.

The following example illustrates the application of the half-normal distribution to quality control. Assume the target value of a specific quality characteristic is 5%. Historical data show its standard deviation is 0.218%. Five measurements are collected and their values are 5.275%, 4.856%, 4.597%, 5.560% and 5.343%. The L(5) can be calculated as follows:

Y_1 = | (5.275-5.000)/0.218 | = 1.26147,
Y_2 = | (4.856-5.000)/0.218 | = 0.66055,
Y_3 = 1.84862, Y_4 = 2.56881, and Y_5 = 1.57339;
L(5) = (1.26147+0.66055+1.84862+ 2.56881+1.57339)/5 = 1.58257

If the a value is set to be 0.0027, then, from Table 3, the UCL of L(5) is 1.6650, which is the value of $L(5)_{.9973}$. Since L(n) is a sample statistic taken absolute value, its LCL is always zero. Because the value of L(5) from the

n	Mean Theoretical	Mean Simulation	Variance Theoretical	Variance Simulation
2	0.7979	0.7979	0.1817	0.1813
3	0.7979	0.7974	0.1211	0.1212
4	0.7979	0.7982	0.0908	0.0907
5	0.7979	0.7981	0.0727	0.0725
6	0.7979	0.7976	0.0606	0.0607
7	0.7979	0.7974	0.0519	0.0519
8	0.7979	0.7973	0.0454	0.0454
9	0.7979	0.7980	0.0404	0.0406
10	0.7979	0.7981	0.0363	0.0364
12	0.7979	0.7982	0.0303	0.0302
15	0.7979	0.7974	0.0242	0.0241
20	0.7979	0.7977	0.0182	0.0184

Table 2. Mean and Variance of L(n)

n	$L(n)_{.50}$	$L(n)_{.80}$	$L(n)_{.90}$	$L(n)_{.95}$	$L(n)_{.9545}$	$L(n)_{.9876}$	$L(n)_{.99}$	$L(n)_{.9973}$
2	0.7439	1.1447	1.3774	1.5778	1.6054	1.9350	1.9850	2.2600
3	0.7628	1.0817	1.2659	1.4271	1.4465	1.7014	1.7450	1.9550
4	0.7721	1.0450	1.2015	1.3367	1.3532	1.5675	1.6000	1.7733
5	0.7769	1.0197	1.1573	1.2740	1.2896	1.4725	1.5033	1.6650
6	0.7808	1.0002	1.1239	1.2291	1.2429	1.4133	1.4388	1.5900
7	0.7822	0.9857	1.0987	1.1976	1.2103	1.3625	1.3856	1.5100
8	0.7848	0.9737	1.0788	1.1700	1.1816	1.3223	1.3431	1.4600
9	0.7859	0.9643	1.0632	1.1480	1.1588	1.3946	1.3145	1.4200
10	0.7876	0.9559	1.0504	1.1294	1.1389	1.2657	1.2854	1.3867
12	0.7895	0.9428	1.0253	1.0979	1.1065	1.2200	1.2385	1.3375
15	0.7909	0.9264	1.0007	1.0653	1.0732	1.1733	1.1867	1.2700
20	0.7921	0.9105	0.9749	1.0303	1.0369	1.1200	1.1333	1.2080

Table 3. Some Useful Quantiles of L(n)

Sample	Time	Data (n=5)	X-bar	R	L(n=5)
1	8:30	507, 503, 496, 505, 501	502.4	11	0.615
2	9:00	502, 497, 495, 503, 506	500.6	11	0.585
3	9:30	488, 505, 499, 500, 498	498.0	17	0.615
4	10:00	515, 511, 504, 516, 509	511.0*	12	1.692*
5	10:30	493, 501, 504, 496, 505	499.8	12	0.708
6	11:00	500, 490, 503, 498, 513	500.8	23	0.862
7	11:30	507, 496, 482, 488, 515	497.6	33*	1.723*
8	12:00	493, 502, 510, 498, 507	502.0	17	0.862

"*" indicates that the process is out of control.

Table 4. Numerical Data from the Plant Producing Packed Grape Juice

above sample is 1.58257, which is less than UCL, the process is judged to be in control. The quality control analysis procedure using the half-normal distribution method (HNDM) is summarized as follows:

Step 1. Determine the a value such that the UCL of L(n) can be obtained from Table 3. For example, as n=5, if a=0.0027, then the UCL of L(5) is 1.6650; if a=0.01, then the UCL of L(5) is 1.5033. The LCL of L(n) is always zero.

Step 2. Determine the mean and standard deviation of the quality characteristic such that the value of L(n) can be calculated from the sample.

Step 3. If L(n) is less than or equal to UCL, the process is judged to be in control. Otherwise, the process is said to be out of control.

A Case Study of Quality Control

The following case study is given to illustrate how the HNDM works and to compare the HNDM with the traditional two-chart (X-bar and R charts) method.

Data are collected from a plant producing packed grape juice every 30 minutes. At each time, five bottles of packed grape juice are selected at random. The key quality characteristic is the quantity of content (QOC), measured by “cc”. The target of QOC is 500 cc for each bottle.

Historical data indicate the standard deviation of QOC is approximately 6.5 cc. The data collected from 8 samples are shown in Table 4.

Suppose that the 3-sigma control limits are used in the X-bar and R charts to monitor the process. The control limits for the X-bar chart are:

$$UCL_{X\text{-bar}} = \text{target} + 3s/n^{1/2} = 500 + 3(6.5)/5^{1/2} = 508.7$$

$$LCL_{X\text{-bar}} = \text{target} - 3s/n^{1/2} = 500 - 3(6.5)/5^{1/2} = 491.3$$

and the control limits for the R chart are:

$$UCL_R = D_2\, s = 4.918\ (6.5) = 31.97$$

$$LCL_R = D_1\, s = 0\ (6.5) = 0$$

where the values of D_1 and D_2 can be found in most books of Quality Control. From the columns of X-bar and R in Table 4, it can be seen that the fourth and seventh samples are out of the control limits. The fourth sample indicates that a shift occurs in process mean; meanwhile, the seventh sample shows a shift in variability.

Instead of using the X-bar and R charts, the HNDM can be applied to monitor the process. The sample statistic L(n) is calculated and the results are also shown in Table 4. If the a is set to be 0.0027 (which is the same as the 3-sigma limits of the X-bar chart), then the UCL of L(n) is 1.6650 (from Table 3) and the LCL is zero. From Table 4, the L(n)’s in the fourth and seventh samples are greater than the UCL, which indicates the process is out of control in the corresponding samples. This conclusion is consistent with that from the traditional two-chart method.

Conclusions

Traditionally, the X-bar and R charts are used in process monitoring. Therefore, two sample statistics (X-bar and R) are usually needed to be concerned and the quality engineers always have no idea about the Type I error probability in the R chart. This article proposes the half-normal distribution method (HNDM) to monitor the process. The advantages of the HNDM are that the traditional two sample statistics can be reduced to one (i.e., L(n)) and that the Type I error probability of the hypothesis testing can be set by the users. From the case study presented, the HNDM provides an alternative way to detect “when” a shift occurs as what the traditional two-chart method does. Although the HNDM does not tell the shift is due to central tendency or variability, the users can get this information easily from the data pattern of the corresponding sample(s).

References

Banks, J. and Carson, J.S., (1984). *Discrete-Event System Simulation*, Pretice-Hall, Inc., New York, NY, 315-316.

Chou, C.-Y., (1994). Evaluating the Estimators of Standard Deviation in Control Charting. *Proceedings of the Taipei International Quality Conference*, Taipei, Taiwan, 463-468.

Grant, E. L. and Leavenworth, R. S., (1995). *Statistical Quality Control*, 7th ed., McGraw-Hill Book Company, New York, NY.

Hogg, R.V. and Tanis, E.A., (1993). *Probability and Statistical Inference*, 4th ed., Macmillan Publishing Co., Inc., New York, NY.

Hool, J. and Maghsoodloo, S., (1980). Normal Approximation to Linear Combinations of Independently Distributed Random Variables. *AIIE Transactions*, (12), 140-144.

Montgomery, D. C., (1991). *Introduction to Statistical Quality Control*, 2nd ed., John Wiley & Sons, New York, NY.

Nelson, L.S., (1990). Setting Up a Control Chart Using Subgroups of Varying Sizes. *Journal of Quality Technology*, (22), 245-246.

Parker, F., Brown, E.R. and Vecellio, R.L., (1994). Development of New Criteria for Control of Hot Mix Asphalt Construction. *Project Report of Harbert Engineering Center*, Auburn University, Auburn, AL.

Simon, L.E., (1949). The Advancing Frontier of Quality Control. *Industrial Quality Control*, (5), 5-7.