UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

CUMULATIVE SUM CONTROL CHARTS FOR THE FOLDED NORMAL DISTRIBUTION

by

N. L. Johnson

December 1962

Grant No. AFOSR -62-148

Methods of construction of cumulative sum control charts for folded normal variates are described. These charts are likely to be useful when the sign of an approximately normally distributed quantity is lost in measurement. Some assessment is given of the information lost by omission of the sign.

This research was supported by the Air Force Office of Scientific Research.

Institute of Statistics
Mimeo Series No. 346

CUMULATIVE SUM CONTROL CHARTS FOR THE FOLDED NORMAL DISTRIBUTION

by

N. L. Johnson[1]

University of North Carolina

1. Introduction.

The distribution of the modulus of a normal variable is known as the 'folded' normal distribution. It is reasonable to apply this distribution when the sign of a variable, which can be justifiably represented by a normally distributed random variable, is irretrievably lost in measurement. Some examples are given by Leone et al [7]. This paper also describes properties of the folded normal distribution. It also includes a discussion of methods of estimating the parameters, which are further discussed by Elandt [1] and Johnson [3].

Many standard control chart procedures are based on the assumption that the observations used (either individually or as sample arithmetic means) in plotting the chart can be represented by normally distributed random variables. It is evident that, on occasion, it may be desirable to construct control chart procedures when the appropriate distribution is that of a folded normal variable. In this paper the construction of cumulative sum control charts for such variables will be described. The methods used will be based on the ideas described by Johnson [2] and applied by Johnson and Leone [4] [5] [6].

2. Statement of Problem.

If the original variable (x', say) is normally distributed with expected value ξ and standard deviation σ, then the probability density function of the folded normal variable $x = |x'|$ is

1. This research was supported by the Mathematics Division of the Air Force Office of Scientific Research.

(1) $$p(x|\theta,\sigma) = \sigma^{-1}\sqrt{(2/\pi)}\, e^{-\frac{1}{2}[\theta^2 + (x/\sigma)^2]} \cosh[\theta x/\sigma].$$

where $\theta = \xi/\sigma$.

We will consider two situations (a) where we are trying to keep $\xi = 0$, so that the original variable has zero mean (b) where we are trying to control σ at a specified value σ_o.

In case (a) we suppose σ is known; in case (b) we suppose ξ is known.

3. Control of Mean.

If we have a sequence of m independent random variables $x_1, x_2, \ldots, x_m$, each having a folded normal distribution, then the likelihood ratio of the hypothesis H_1: $(|\xi| = \xi_1)$ against the hypothesis H_o: $(\xi = 0)$ is

(2) $$\frac{p(x_1, \ldots, x_m|\theta_1,\sigma)}{p(x_1, \ldots, x_m|\theta,\sigma)} = e^{-\frac{1}{2}m\theta_1^2} \prod_{i=1}^{m} \cosh[\theta_1 x_i/\sigma] \quad .$$

where $\theta_1 = \xi_1/\sigma$.

Hence the sequential probability ratio test discriminating between these two hypotheses has its continuation region defined by the inequalities

(3) $$\ln\left(\frac{\alpha_1}{1-\alpha_o}\right) + \frac{1}{2}m\theta_1^2 < \sum_{i=1}^{m} \ln\cosh[\theta_1 x_i/\sigma] < \ln\left(\frac{1-\alpha_1}{\alpha_o}\right) + \frac{1}{2}m\theta_1^2 \doteq$$

where $\alpha_i = \Pr[\text{accept } H_{1-i}|H_i]$ $(i = 0, 1)$.

Now we apply the method described in [2], regarding the cumulative sum control chart (CSCC) as a reversed sequential test, with a very small value for α_1, and using only the right-hand inequality of (3) which now becomes, effectively

(4) $$\sum_{i=1}^{m} \ln\cosh[\theta_1 x_i/\sigma] < \frac{1}{2}m\theta_1^2 + (-\ln\alpha_o) \quad .$$

To apply this method we must (i) transform the observed values x_i to 'scores' $y_i = \ln \cosh [\theta_1 x_i/\sigma]$ and (ii) plot the cumulative sum $\Sigma_i y_i$. Then change in the value of $|\xi|$ from zero to $\theta_1\sigma$ is indicated if any plotted points fall below the line PQ in Figure 1. Here O is the last plotted point, tan $\hat{OPQ}$ = $\frac{1}{2}\theta_1^2$ and OQ = $-\ln \alpha_o$, so that OP = $(-2 \ln \alpha_o)\theta_1^{-2}$.

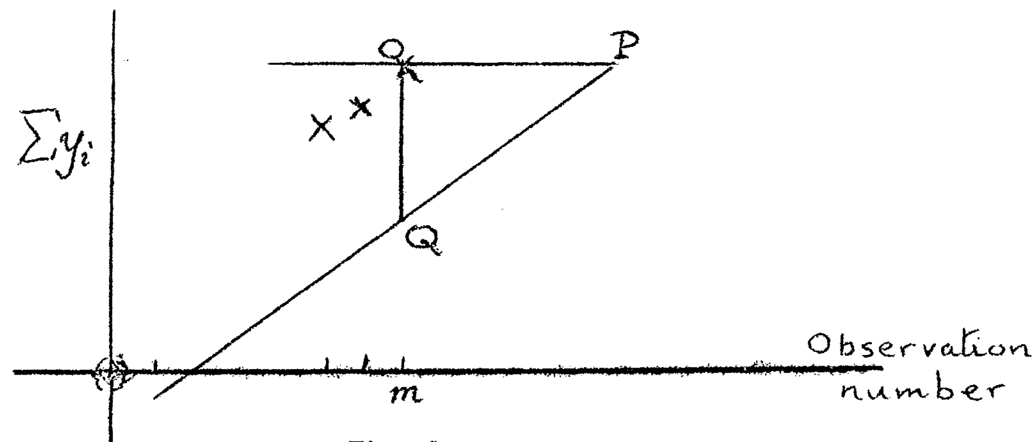

Fig. 1

With this system of plotting y_i is never negative, so the value of Σy_i is non-decreasing. To reduce the amount of paper needed for the chart it may sometimes be preferred to use, instead of y_i, the modified score

$$y_i' = y_i - \frac{1}{2}\theta_1^2 \quad .$$

Then $\Sigma\ y_i'$ is plotted and we simply use a horizontal line, $(-\ln \alpha_o)$ below the last plotted point, as boundary.

The need to transform observed vlaues x_i to scores, y_i or y_i', makes the construction of the CSCC a little troublesome. However a table of values of $\ln \cosh X$ as a function of X makes the task quite easy, while if a considerable amount of work is to be done for well-established values of $\theta_1 (= \xi_1/\sigma)$ and σ, a special table of $y = \ln \cosh(\theta_1 x/\sigma)$ can be constructed once for all. (It may be noted that when x is large, $y \sim \theta_1 x/\sigma - \ln 2$.)

4. Control of Mean-Average Sample Number.

We will restrict ourselves to consideration of the situation when the true mean has in fact shifted to $\xi_1 = \theta_1\sigma$. The expected number of observations before this will be detected is approximately $(-\ell n\ \alpha_0)E^{-1}$ where

$$E = \mathcal{E}\left[\log\left\{\frac{p(x|\theta_1,\sigma)}{p(x|\theta,\sigma)}\right\}\ |\theta_1,\sigma\right]$$

$$= -\frac{1}{2}\theta_1^2 + \mathcal{E}\left[\ell n\ \cosh(\theta_1 x/\sigma)\ |\theta_1,\sigma\right]$$

$$= -\frac{1}{2}\theta_1^2 + \sigma^{-1}\sqrt{(2/\pi)}\ e^{-\frac{1}{2}\theta_1^2}\int_0^\infty e^{-\frac{1}{2}(x/\sigma)^2}\cosh(\theta_1 x/\sigma)\ell n\ \cosh(\theta_1 x/\sigma)\cdot dx$$

(5) $$= -\frac{1}{2}\theta_1^2 + \sqrt{(2/\pi)}\ e^{-\frac{1}{2}\theta_1^2}\int_0^\infty e^{-\frac{1}{2}y^2}\cosh\theta_1 y\ \ell n\ \cosh\theta_1 y\cdot dy \quad .$$

The integral can be evaluated as an infinite series using the expansion

$$\ell n\ \cosh\theta_1 y = \ell n\left[\frac{1}{2}e^{\theta_1 y}(1 + e^{-2\theta_1 y})\right]$$

$$= \theta_1 y - \ell n\ 2 + \sum_{j=1}^{\infty}(-1)^{j+1}\ j^{-1}\ e^{-2j\theta_1 y} \quad .$$

Hence

$$\cosh\theta_1 y\ \ell n\ \cosh\theta_1 y = (\theta_1 y - \ell n\ 2)\cosh\theta_1 y + \frac{1}{2}\left[e^{-\theta_1 y} + \sum_{j=1}^{\infty}(-1)^{j+1}\left\{j(j+1)e^{(2j+1)\theta_1 y}\right\}^{-1}\right]$$

and

(6) $$\sqrt{(2/\pi)}\, e^{-\frac{1}{2}\theta_1^2} \int_0^{\infty} e^{-\frac{1}{2}y^2} \cosh \theta_1 y \, \ln \cosh \theta_1 y \cdot dy$$

$$= \theta_1 \left[\sqrt{(2/\pi)}\, e^{-\frac{1}{2}\theta_1^2} + \theta_1 \left\{ 2F(\theta_1) - 1 \right\} \right] - \ln 2 + I(\theta_1)$$

$$+ \sum_{j=1}^{\infty} (-1)^{j+1} \left\{ j(j+1) \right\}^{-1} I((2j+1)\theta_1)$$

where

$$I(r\theta_1) = \sqrt{(2/\pi)}\, e^{-\frac{1}{2}\theta_1^2} \cdot \frac{1}{2} \int_0^{\infty} e^{-\frac{1}{2}y^2 - r\theta_1 y} dy = e^{\frac{1}{2}(r^2-1)\theta_1^2} \left[1 - F(r\theta_1) \right] .$$

Finally we obtain

(7) $$E = \theta_1 \left[\sqrt{(2/\pi)}\, e^{-\frac{1}{2}\theta_1^2} + \theta_1 \left\{ 2F(\theta_1) - \frac{3}{2} \right\} \right] - \ln 2 + 1 - F(\theta_1)$$

$$+ \sum_{j=1}^{\infty} (-1)^{j+1} \left\{ j(j+1) \right\}^{-1} e^{2j(j+1)\theta_1^2} \left[1 - F(\overline{2j+1} \cdot \theta_1) \right] .$$

An approximate formula for E can be obtained by using the approximation to Mills' ratio

$$\sqrt{2\pi}\, e^{\frac{1}{2}x^2} \left[1 - F(x) \right] \doteq x^{-1} .$$

The resulting formula

(8) $$E \doteq \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta_1^2} \right) \left(2\theta_1 + \frac{\pi - 3}{\theta_1} \right) + \theta_1^2 \left[2F(\theta_1) - \frac{3}{2} \right] - \ln 2 + 1 - F(\theta_1)$$

gives useful results for θ_1 larger than 3/4.

Using further terms in the asymptotic expression

$$\sum_{i=o}^{\infty} (-1)^{i+1} a_i X^{-(2i+1)} \qquad (a_o = 1; \quad a_i = \frac{(2i-1)!}{2^{i-1}(i-1)!} \quad \text{for } i \geq 1)$$

for Mill's ratio we get the formal expression

$$E = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\Theta_1^2} \right) \left[2\Theta_1 + \sum_{i=o}^{\infty} (-1)^i a_i \Theta_1^{-(2i+1)} \left\{ \sum_{j=o}^{i} \frac{E_j \pi^{2j+1}}{(2j)! 2^{2j}} - 4i - 3 \right\} \right]$$

$$+ \Theta_1^2 \left[2F(\Theta_1) - \frac{3}{2} \right] - \ln 2 + 1 - F(\Theta_1)$$

where E_j is the j-th Euler number.

Evaluating the coefficients in the asymptotic expansion we obtain

$$(8)' \quad E = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\Theta_1^2} \right) \left[2\Theta_1 + \frac{0.14169}{\Theta_1} - \frac{0.01738}{\Theta_1^3} + \frac{0.00603}{\Theta_1^5} - \dots \right]$$

$$+ \Theta_1^2 \left[2F(\Theta_1) - \frac{3}{2} \right] - \ln 2 + 1 - F(\Theta_1) \quad .$$

Some values of E^{-1} as a function of Θ_1 are shown in Table 1. Values given by the approximate formulas (8) and (8)' (excluding the term $0.00603\Theta_1^{-5}$) are also shown.

TABLE 1. Values of $E^{-1} = (-\ln \alpha_o)^{-1}$ x (Expected number of observations)

Θ_1	E^{-1}	Values from (8)	Values from (8)'	Θ_1	E^{-1}	Value from (8) or (8)'
0.25	(980)			1.25	2.95	2.93
0.50	73.5	(26.4)	(112)	1.50	1.67	1.67
0.75	16.6	14.6	17.9	1.75	1.06	1.06
1.00	6.13	6.01	6.17	2.00	0.73	0.73

When θ_1 is large, $E \sim \frac{1}{2}\theta_1^2 - \ln 2$ (for $\theta_1 = 2$, $E = 1.36735$ and $(\frac{1}{2}\theta_1^2 - \ln 2 = 1.30685)$. The average number of observations is then approximately $(-2 \ln \alpha_o)[\theta_1^2 - \ln 4]^{-1}$, as compared with $(-2 \ln \alpha_o)\theta_1^{-2}$ (= OP) for the CSCC for means of normal variables (see [2]).

The average sample number for the folded normal distribution is thus greater by a factor of

$$(1 - \theta_1^{-2} \ln 4)^{-1}$$

approximately. This may be regarded as reflecting the loss in information about the signs of the variables in the folded normal. As would be expected, the ratio tends to 1 as θ_1 increases.

5. Control of Standard Deviation.

We now suppose that ξ is known and that we wish to control the value of σ at σ_o. If we desire to detect a change in the value of σ from σ_o to σ_1 (for definiteness we will suppose $\sigma_1 > \sigma_o$) then the appropriate likelihood ratio is

$$(9)\quad \frac{p(x_1,x_2,\ldots,x_m|\xi,\sigma_1)}{p(x_1,x_2,\ldots,x_m|\xi,\sigma_o)} = \left(\frac{\sigma_o}{\sigma_1}\right)^m e^{\frac{1}{2}(m\xi^2 + \sum_{i=1}^m x_i^2)(\sigma_o^{-2}-\sigma_1^{-2})} \prod_{i=1}^m \left[\frac{\cosh(x_i\xi/\sigma_1^2)}{\cosh(x_i\xi/\sigma_o^2)}\right] .$$

We will infer that there has been a change in σ if the inequality

$$(10)\quad m\left[\frac{1}{2}\xi^2(\sigma_o^{-2}-\sigma_1^{-2}) - \ln(\sigma_1/\sigma_o)\right] + \left(\frac{1}{2}\sum_{i=1}^m x_i^2\right)(\sigma_o^{-2}-\sigma_1^{-2}) + \sum_{i=1}^m \ln\left[\frac{\cosh(x_i\xi/\sigma_1^2)}{\cosh(x_i\xi/\sigma_o^2)}\right] < -\ln \alpha_o$$

is not satisfied (the x's being numbered backwards, starting from the last observation). Examination of (10) is more troublesome than applying the condition (4) for the control of mean value. However,(10) can be written in the form

$$(11) \qquad \sum_{i=1}^{m} (y_i^{(1)} - y_i^{(o)}) < (-\ln \alpha_o) + m\left[\ln(\sigma_1/\sigma_o) - \tfrac{1}{2}\, \xi^2(\sigma_o^{-2} - \sigma_1^{-2})\right] \quad .$$

So if tables of 'scores'

$$y^{(j)} = \ln \cosh(x\, \xi\, \sigma_j^{-2}) - \tfrac{1}{2}(x/\sigma_j)^2 \qquad (j = 0,\ 1)$$

are available we can plot $\Sigma(y^{(1)} - y^{(o)})$ against number of observations to form a CSCC. To check whether there is evidence of a change in the value of σ we see whether any points fall below the line PQ (see Figure 1) where, as before, O is the last plotted point and OQ = $(-\ln \alpha_o)$. In the present case $\tan \hat{OPQ} =$ $\ln(\sigma_o/\sigma_1) - \frac{1}{2}\, \xi^2(\sigma_o^{-2} - \sigma_1^{-2})$.

It should be noted that $\tan \hat{OPQ}$ can be negative. If $\xi^2 > \left[2 \ln(\sigma_o/\sigma_1)\right] (\sigma_o^{-2} - \sigma_1^{-2})^{-1}$ then the point P will be to the left of O, as in Figure 2, which exhibits a situation where a change in the value of σ would be indicated at T

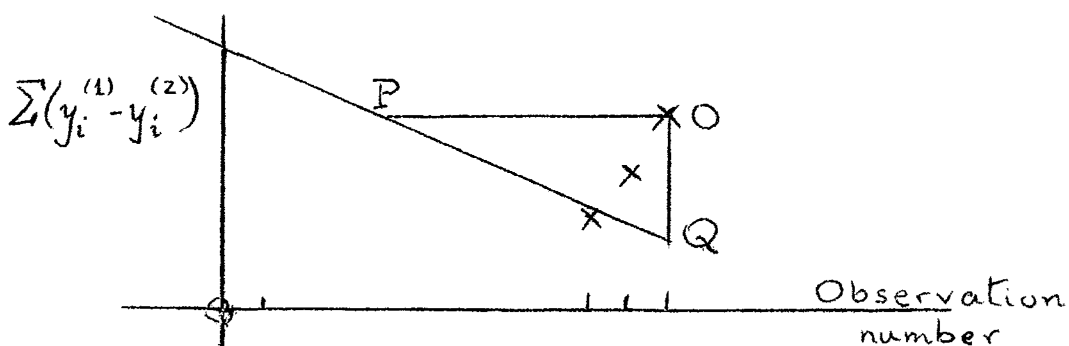

Figure 2.

In the special case where we can assume that the mean (ξ) is controlled at $\xi = 0$, the problem reduces to that of constructing a CSCC for the variance of a normal population with known mean (in this case zero). This case can be covered by the methods similar to those described by Johnson and Leone [5].

It should be noted, however, that the methods described in [5] which use sample range cannot be employed when the sign of each observation is lost. There is a technically correct way of introducing range, even in this case, by assigning positive and negative signs at random (with probability 1/2 each) to the observations. However, apart from the labor involved, the element of arbitrariness militates against acceptance of this method as a practical procedure.

6. Average Sample Number under More General Conditions.

The average number of observations needed for the control of mean procedure (described in section 3) indicates a change in ξ when the true value of ξ is ξ' (a general value) is $(-\ln \alpha_o)E'^{-1}$ where

$$(12) \qquad E' = -\frac{1}{2}\theta_1^2 + \sqrt{(2/\pi)}\, e^{-\frac{1}{2}\theta'^2} \int_0^\infty e^{-\frac{1}{2}y^2} \cosh \theta' y \ln \cosh \theta_1 y \, dy$$

and $\theta' = \xi'/\sigma$.

The integral in this expression can be evaluated in a similar way to that in (5), though the result is a little more complicated.

Numerical evaluation of this quantity, and of similar quantities needed to assess the performance of the procedure described in section 5, will be discussed elsewhere. The present paper is intended to show how cumulative sum control charts can be constructed for certain types of control based on signless observations.

REFERENCES

[1] Elandt, Regina C. The folded normal distribution: Two methods of estimating parameters from moments", Technometrics, Vol. 3 (1961), pp. 551-562.

[2] Johnson, N. L. "A simple theoretical approach to cumulative sum control charts", J. Amer. Statist. Ass., Vol. 56 (1961), pp. 835-840.

[3] Johnson, N. L. "The folded normal distribution: Accuracy of estimation by maximum likelihood", Technometrics, Vol. 4 (1962), pp. 249-256.

[4] Johnson, N. L. and Leone, F. C. "Cumulative sum control charts: Mathematical principles applied to their construction and use. I", Indust. Qual. Control, Vol. 18(12) (1962), pp. 15-21.

[5] __________ " do , II", do , Vol. 19(1) (1962), pp. 29-36.

[6] __________ " do , III", d0 , Vol. 19(2) (1962), pp. 22-28.

[7] Leone, F. C., Nelson and Nottingham (1961) "The folded normal distribution", Technometrics, Vol. 3 (1961), pp. 543-550.