Assessing the Proportion of Conformance of a Process from a Bayesian Perspective

Michael Perakis\textsuperscript{a*†} and Evdokia Xekalaki\textsuperscript{b}

A Bayesian technique for assessing the posterior probability that the proportion of conformance of a normally distributed process exceeds a particular value is suggested. Both exact and approximate formulae for the assessment of this probability are derived. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: statistical process control; proportion of conformance; posterior probability; process capability index

1. Introduction

The question of assessing the capability of a process to produce according to some specifications connected with a measurable characteristic (X) of the items produced through it is of great importance in statistical process control. These specifications consist of the lower specification limit (L), which is the minimum acceptable process value; the upper specification limit (U), which is the maximum acceptable process value; and the target value (T), which coincides with the most desirable process value.

Process capability indices constitute the most commonly used tool for measuring process capability. Every process capability index combines the values of process specifications, with those of the parameters of the process, that is, the mean (μ) and the standard deviation (σ), takes non-negative values, and its large values provide evidence of a capable process.

The development and the estimation of a process capability index simplify substantially when the studied process is assumed to be normally distributed and most of the work that has appeared in the literature on this issue has been based on this assumption. For more details on various aspects of process capability indices, such as properties, estimation techniques, and interrelations, the reader is referred to Kotz and Johnson\textsuperscript{1,2} Kotz and Lovelace\textsuperscript{3} Pearn and Kotz\textsuperscript{4} and Wu et al\textsuperscript{5}. Furthermore, Spiring et al\textsuperscript{6} and Yum and Kim\textsuperscript{7} provide an extensive coverage of the bibliography on process capability indices.

It should be noted that after the publication of the paper by Kane\textsuperscript{8}, which undoubtedly motivated many authors to deal with process capability indices, several papers on this issue have been published. The appearance of new papers concerning process capability indices is continuous even in the last 3 years; almost 30 years after the publication of the paper by Kane\textsuperscript{8}, many new papers have been published, introducing new process capability indices or new findings on existing ones (see, e.g., Negrin et al\textsuperscript{9} Pearn et al\textsuperscript{10} Perakis and Xekalaki\textsuperscript{11} Jalili et al\textsuperscript{12} Jose and Luke\textsuperscript{13} and Shiau et al\textsuperscript{14}).

Another measure that provides information on the capability of a process is the proportion of conformance (yield), which is defined as the probability of producing within the so-called specification area, that is, the interval that consists of all the acceptable process values. Under the assumption that the process is normally distributed, it is defined as

\[ P = P(L < X < U) = \Phi \left( \frac{U - \mu}{\sigma} \right) - \Phi \left( \frac{L - \mu}{\sigma} \right) \]  

where \( \Phi(.) \) denotes the cumulative distribution function of the standard normal distribution. The proportion of conformance of a process is associated with the values of the most commonly used capability indices, as several authors, such as Boyles\textsuperscript{15} and Kotz and Lovelace\textsuperscript{3} have shown. Carr\textsuperscript{16} suggested its direct use for measuring process capability, while Perakis and Xekalaki\textsuperscript{17} suggested a process capability index whose value is based on that of the yield of the studied process. Hence, the proportion of conformance plays an important role in statistical process control, and the development of new methods for its estimation is a rather interesting issue.

\textsuperscript{a}General Accounting Office, Athens, Greece
\textsuperscript{b}Department of Statistics, Athens University of Economics and Business, Athens, Greece
\textsuperscript{*Correspondence to: Michael Perakis, General Accounting Office, 57 Panepistimiou Str., 105 64 Athens, Greece.
\textsuperscript{†}E-mail: mperakis@gmail.com

Copyright © 2013 John Wiley & Sons, Ltd.
In most of the cases, the parameters of the examined process are unknown, and an estimate of the actual value of the yield of a process is required. If $X_1, \ldots, X_n$ are the elements of a random sample taken from the process, and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2}$ are the corresponding sample mean and standard deviation, respectively; a point estimator of $p$ can be obtained by substituting $\bar{X}$ and $S$ for $\mu$ and $\sigma$ in (1), respectively. The resulting estimator of $p$ is thus given by

$$\hat{p} = \Phi\left(\frac{U - \bar{X}}{S}\right) - \Phi\left(\frac{L - \bar{X}}{S}\right)$$

The estimator of $p$ given in (2) differs from its minimum variance unbiased estimator given by (see, e.g., Kotz and Johnson$^1$)

$$\hat{p} = \Phi(i_{\theta_1}(\bar{X}, S)(0.5(n-1), 0.5(n-1))) - \Phi(i_{\theta_2}(\bar{X}, S)(0.5(n-1), 0.5(n-1)))$$

where the functions $\theta_1(\bar{X}, S)$ and $\theta_2(\bar{X}, S)$ are defined as

$$\theta_1(\bar{X}, S) = 0.5 \left[ 1 + \frac{U - \bar{X}}{S} \left\{ \frac{(n-1)^2}{n} + \left( \frac{U - \bar{X}}{S} \right)^2 \right\}^{0.5} \right]$$

and

$$\theta_2(\bar{X}, S) = 0.5 \left[ 1 + \frac{L - \bar{X}}{S} \left\{ \frac{(n-1)^2}{n} + \left( \frac{L - \bar{X}}{S} \right)^2 \right\}^{0.5} \right]$$

respectively, and $i_{\theta}(a,b)$ denotes the incomplete beta function ratio defined as

$$i_{\theta}(a,b) = \frac{B_{\theta}(a,b)}{B(a,b)}$$

with $B(a,b)$ denoting the beta function, given by

$$B(a,b) = \int_{0}^{1} y^{a-1}(1-y)^{b-1} dy, \ a, b > 0$$

and $B_{\theta}(a,b)$ denoting the incomplete beta function, given by

$$B_{\theta}(a,b) = \int_{c}^{1} y^{a-1}(1-y)^{b-1} dy, \ a, b > 0, 0 < c < 1$$

However, the estimator $\hat{p}$ is much more convenient than the minimum variance unbiased estimator of $p$, and as pointed out by Kotz and Johnson,$^1$ it is quite adequate for most occasions. For more details on the minimum variance unbiased estimator of the proportion of conformance, the interested reader is referred to Folks et al.$^{18}$ Wheeler,$^{19}$ and Kotz and Johnson.$^1$

In this paper, we focus on the estimation of process yield, and we suggest a Bayesian technique that enables one to assess the posterior probability that the actual value of the proportion of conformance of a normally distributed process exceeds a certain value. Both exact and approximate formulae for the assessment of this probability are derived. Specifically, in Section 2, a brief presentation of the existing techniques for obtaining lower confidence limits for the true value of the proportion of conformance is made. In Section 3, the suggested technique is presented. Plots that clarify various aspects on it are also provided. In Section 4, an illustrative example is given, while in Section 5, a discussion of the advantages of the suggested technique over the existing ones is provided. Further, the use of the proposed technique is considered in connection with process capability index $C_{pc}$, suggested by Perakis and Xekalaki.$^{17}$

2. Some methods for obtaining confidence limits for the yield of a process

In this section, a concise presentation of some techniques suitable for obtaining a lower confidence limit for the actual value of the yield of a process is given. More details on these techniques are provided by Owen and Hua,$^{20}$ Chou and Owen,$^{21}$ Wang and Lam,$^{22}$ and Perakis and Xekalaki.$^{17}$

Owen and Hua$^{20}$ sought confidence limits for the probabilities $p_1 = P(X < L)$ and $p_2 = P(X > U)$, and Chou and Owen$^{21}$ constructed one-sided simultaneous confidence regions on the lower ($p_1$) and the upper ($p_2$) tail areas of the normal distribution. Extensive tables containing the values of the confidence limits obtained from these two methods are included in the corresponding articles. Denoting
the upper confidence limits for \( p_1 \) and \( p_2 \) obtained by any of these two methods by \( p_1^* \) and \( p_2^* \), respectively; Wang and Lam\(^{22} \) argue that a 100(1 - \( \alpha \))% lower confidence limit for \( p \) is given by

\[
1 - p_1^* - p_2^* \quad (3)
\]

An obvious drawback of the approaches described earlier is that their implementation requires knowledge of the entries of the tables provided in the corresponding articles. This restricts their use, and thus Wang and Lam\(^{22} \) introduced another method for constructing an approximate lower confidence limit for \( p \), which is much simpler and does not require any special tables. In particular, Wang and Lam\(^{22} \) sought a value \( p^* \) such that \( P(p \geq p^*) = 1 - \alpha \) and concluded that the desired 100(1 - \( \alpha \))% approximate lower confidence limit for \( p \) is given by

\[
p^* = \Phi \left( \frac{1}{\sqrt{n}} + C_1 \right) \cdot \Phi \left( \frac{1}{\sqrt{n}} \cdot C_2 \right) \quad (4)
\]

where

\[
C_1 = \max(K_1, K_2) \sqrt{\frac{\chi^2_{n-1, \alpha}}{n-1}}
\]

\[
C_1C_2 = \min(K_1, K_2) \sqrt{\frac{\chi^2_{n-1, \alpha}}{n-1}}
\]

\( K_1 = (\bar{X} - L)/S, \ K_2 = (U - \bar{X})/S, \) and \( \chi^2_{n-1, \alpha} \) denotes the \( \alpha \) quantile of the chi-square distribution with \( n - 1 \) DOF.

Finally, Perakis and Xekalaki\(^{17} \) argue that the coverage of lower confidence limit (4) can be improved using a variant of it given by

\[
p^* = \Phi \left( \frac{1}{\sqrt{n}} + \left( 1 + \frac{1}{n} \right) C_1 \right) \cdot \Phi \left( \frac{1}{\sqrt{n}} \cdot \left( 1 + \frac{1}{n} \right) C_2 \right) \quad (5)
\]

Extensive tables that compare the performances of lower confidence limits (3), (4), and (5) for several combinations of process parameters and sample sizes are provided by Wang and Lam\(^{22} \) and Perakis and Xekalaki\(^{17} \).

### 3. The suggested technique

The techniques described in the previous section result in a bound that the actual value of the proportion of conformance exceeds with a certain confidence level. However, sometimes one may be interested (from a Bayesian perspective) in obtaining the probability that the actual value of the proportion of conformance exceeds a certain value. In such cases, the techniques described in Section 2 cannot be used and, hence, arises the need for the development of a new technique.

In the sequel, a Bayesian technique that enables one to assess the posterior probability that the proportion of conformance of a normally distributed process exceeds a particular value is suggested. The implementation of this technique is meaningful provided that the mean of the process coincides with the midpoint of the specification area, that is, \( \mu = M = (L + U)/2 \). Actually, under this assumption, the proportion of conformance of a normally distributed process, given in (1), simplifies to

\[
p = 2\Phi(d/\sigma) - 1 \quad (6)
\]

where \( d \) denotes the half-length of the specification area, that is, \( d = (U - L)/2 \).

Let \( X = (X_1, \ldots, X_n) \) be a random sample of size \( n \) taken from the studied process. Then, the posterior probability that the true value of the proportion of conformance exceeds a particular value \( p^* \) (0 < \( p^* < 1 \)) is given by

\[
G(p^*) = P(p > p^* | X = x) \quad (7)
\]

Using (6), this probability can be written as

\[
G(p^*) = P(2\Phi(d/\sigma) - 1 > p^* | X = x)
\]

or, equivalently

\[
G(p^*) = P\left( \sigma < \frac{d}{\Phi^{-1}((p^* + 1)/2)} \mid X = x \right) \quad (8)
\]

where \( \Phi^{-1}(.) \) denotes the inverse of the cumulative distribution function of the standard normal distribution.
The probability given in (8) can be evaluated following an approach similar to that followed by Chan et al.\textsuperscript{23} for the assessment of the posterior probability that the actual value of the process capability index \(C_{pm}\) exceeds a particular value. Specifically, under the assumption that \(\mu = M\), the only unknown parameter involved is \(\sigma\), and thus, the likelihood function is given by

\[
 l(\sigma | X = x) = (2\pi)^{-n/2} \sigma^{-n} \exp \left( -\frac{\sum_{i=1}^{n}(X_i - M)^2}{2\sigma^2} \right)
\]

Box and Tiao\textsuperscript{24} have shown that in the case where one is interested in making inference on the standard deviation of a normal distribution with known mean, a non-informative prior distribution should be locally uniform in log \(\sigma\) and thus have a density function proportional to \(\sigma / C_0\), that is,

\[
 f(\sigma) \propto \sigma / C_0
\]

Consequently, assuming a non-informative prior distribution for \(\sigma\) with density function of the previous form, the posterior distribution of \(\sigma\) has a density function given by

\[
 f(\sigma | X = x) = \left( \frac{2}{\Gamma(n/2)} \right) \left( \frac{(n-1)\hat{\sigma}^2}{2} \right)^{n/2} \sigma^{-(n+1)} \exp \left( -\frac{(n-1)\hat{\sigma}^2}{2\sigma^2} \right), \quad 0 < \sigma < \infty
\]

where

\[
 \hat{\sigma}^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - M)^2 = S^2 + \frac{n}{n-1} (\bar{X} - M)^2
\]

Therefore, the probability given in (8) can be written as

\[
 G(p^*) = \int_{0}^{c_1} f(\sigma | X = x) d\sigma
\]

where

\[
 c_1 = \frac{d}{\Phi^{-1}((p^* + 1)/2)}
\]

Using the transformation

\[
 \gamma = \frac{(n - 1)\hat{\sigma}^2}{2\sigma^2}
\]

relationship (10) reduces to

\[
 G(p^*) = \int_{c_2}^{\gamma} \frac{1}{\Gamma(n/2)} y^{n/2 - 1} e^{-y} dy
\]

where

\[
 c_2 = \frac{n - 1}{2} \left( \Phi^{-1}((p^* + 1)/2) \right)^2 \hat{\sigma}^2
\]

Looking at equation (11), one may notice that the function involved in the integral is the Gamma distribution with parameters \(\alpha = n/2\) and \(\beta = 1\). Therefore, the required probability can also be expressed in terms of the cumulative distribution function of the chi-square distribution with \(n\) DOF given by (see, e.g., Johnson et al.\textsuperscript{25})

\[
 F_{\chi^2}(x) = \frac{\Gamma_x(n/2)}{\Gamma(n/2)}, \quad x > 0,
\]

where \(\Gamma_x(a)\) is the incomplete gamma function defined as

\[
 \Gamma_x(a) = \int_{0}^{x} t^{a-1} e^{-t} dt, \quad x > 0
\]

In particular,

\[
 G(p^*) = 1 - \frac{1}{\Gamma(n/2)} \int_{c_2}^{\gamma} y^{n/2 - 1} e^{-y} dy = 1 - F_{\chi^2}(2c_2)
\]
Note that using Wilson and Hilferty’s approximation of $F_{\chi^2}(\cdot)$, the posterior probability that the actual value of $p$ exceeds $p^*$ can be approximated by

$$G(p^*) \approx 1 - \Phi\left(\frac{\sqrt{9n}}{2}\left\{\sqrt{\Phi^{-1}\left((p^* + 1)/2\right)^2 (n - 1) \hat{\sigma}^2 / nd^2 - 1} + \frac{2}{9n}\right\}\right)$$

Figures 1 and 2 depict the values of the ratio $\hat{\sigma}^2 / d^2$ for which $G(p^*) = P\left(p > p^* \left| \left(\hat{\sigma}^2 / d^2\right)\right\right)$ equals 0.90, 0.95, and 0.99 versus the sample size and for $p^*$ equal to 0.95 and 0.99, respectively. These figures provide some insight into the behavior of the suggested technique as the sample size and the values of $G(p^*)$ and $p^*$ vary. In particular, from these figures, one may notice the following:

- For fixed sample size and $p^*$, the values of the ratio $\hat{\sigma}^2 / d^2$ decrease as the value of $G(p^*)$ increases.
- For fixed values of $G(p^*)$ and $p^*$, the values of the ratio $\hat{\sigma}^2 / d^2$ increase as the sample size increases.
- For fixed sample size and $G(p^*)$, the values of the ratio $\hat{\sigma}^2 / d^2$ decrease as the value of $p^*$ increases.

4. An illustrative example

In order to illustrate the assessment of (12) and (13), we use an artificial example taken from Wang and Lam. A sample of size $n = 30$ has been collected from a process with specifications $L = 68$ and $U = 78$. For this sample, the mean is equal to $\bar{x} = 72.8$, and the standard deviation $s = 2$. Using (3), Wang and Lam obtained an estimate of the 95% lower confidence limit for $p$ equal to 0.93798 using the approach by Owen and Hua and equal to 0.9214 using the approach by Chou and Owen. Further, using (4), the estimate of the corresponding lower confidence limit obtained by them was 0.9414. It can be readily seen that the modified form of (4) given in (5) results in $p'' = 0.94915$. 

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig1.png}
\caption{The values of the ratio $R = \hat{\sigma}^2 / d^2$ for which $G(0.95) = P\left(p > 0.95 \left| \left(\hat{\sigma}^2 / d^2\right)\right\right)$ equals 0.90, 0.95, and 0.99 versus the sample size.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig2.png}
\caption{The values of the ratio $R = \hat{\sigma}^2 / d^2$ for which $G(0.99) = P\left(p > 0.99 \left| \left(\hat{\sigma}^2 / d^2\right)\right\right)$ equals 0.90, 0.95, and 0.99 versus the sample size.}
\end{figure}
These methods are applicable in cases where one considers a fixed level of confidence (e.g., 95%) and looks for a value that the actual proportion of conformance exceeds at this level of confidence. The new technique suggested in Section 3 estimates the probability that the proportion of conformance will exceed a certain value. Obviously, the assumption that is required for the implementation of this technique seems to be realistic because the p-value for the \( t \)-test of the hypothesis \( H_0: \mu = M = 73 \) is equal to 0.59. Using (9), we obtain

\[
\hat{\sigma}^2 = S^2 + \frac{n}{n-1} (\bar{X} - M)^2 = 2^2 + \frac{30}{29} (72.8 - 73)^2 = 4.04138
\]

and, thus, the probability that the actual value of the proportion of conformance exceeds 0.95 is given by

\[
G(0.95) = \int_{c_2}^{\infty} \frac{1}{\Gamma(30/2)} \sqrt{\frac{30}{2}} e^{-y^2/2} dy
\]

where

\[
c_2 = \frac{30-1}{2} \left( \Phi^{-1}(0.95 + 1/2) \right)^2 \frac{4.04138}{5^2} = 9.00471
\]

Hence, the value of the required probability is \( 1 - F_{30,25} (30.0942) = 0.958381 \). Alternatively, one may use the approximation given in (13), which results in

\[
G(0.95) \approx 1 - \Phi \left( \sqrt{\frac{9 \times 30}{2} \left( \frac{1}{\Phi^{-1}(0.95 + 1/2)^2} \frac{4.04138}{30 \times 5^2} + \frac{2}{9 \times 30} \right) } \right) = 0.958319
\]

Figure 3 illustrates the behavior of \( G(p^*) = P(p > p^*|X=x) \) as a function of \( p^* \) for this particular example. From this plot, one may observe that \( G(p^*) \) is equal to unity for all the values of \( p^* \) that are smaller than 0.9. For values of \( p^* \) that exceed 0.9, the value of \( G(p^*) \) decreases rapidly and tends to zero as \( p^* \) approaches unity.

### 5. Discussion

A new technique has been suggested that enables the assessment of the posterior probability that the proportion of conformance of a normally distributed process exceeds a particular value. As already pointed out, this technique can be a useful tool in the statistical process control and can be used supplementarily to the existing techniques for assessing lower confidence limits for \( p \) according to the value that one seeks for. Thus, one should use the existing techniques when seeking for a lower confidence limit with a given confidence level and the suggested technique whenever seeking for the probability that \( p \) exceeds a given value.

It should be noted that the suggested technique may also be used in connection with the index \( C_{pc} \) suggested by Perakis and Xekalaki. This index is defined as

\[
C_{pc} = \frac{1-p_0}{1-p^*}
\]

where \( p_0 \) denotes the minimum allowable proportion of conformance.

---

Figure 3. The values of \( G(p^*) = P(p > p^*|X=x) \) versus \( p^* \) for the data of the example considered in Section 4.
In particular, the suggested technique enables one to assess the probability

\[ G'(c) = P(C_{pc} > c | X = x) \]  

where \( c \) is a positive constant. Indeed, it can be verified that relationship (14) reduces to

\[ G(p^*) = P\left(p > 1 - \frac{1 - p_h}{c} | X = x \right) \]

Hence, assessing the probability in (14) is equivalent to assessing the probability in (7), with

\[ p^* = 1 - \frac{1 - p_h}{c} \]

References


Authors’ biographies

Dr. Michael Perakis is currently working as a special scientist/statistician for the General Accounting Office of the Ministry of Finance of Greece and as a tutor/advisor at the postgraduate program on “Quality Assurance” of the Hellenic Open University. He has worked as an adjunct lecturer at the Athens University of Economics and Business (AUEB) and the University of Piraeus. His research interests include Statistical Process Control, Statistical Indices, Analysis of Categorical Data, and Sampling Techniques.

Evdokia Xekalaki is a professor of Statistics, Department of Statistics, AUEB. Before joining AUEB, she taught at the Department of Statistics, Trinity College, University of Dublin; at the Department of Statistics, University of Missouri-Columbia; at the Department of Statistics and Actuarial Science, University of Iowa; and at the Department of Mathematics, University of Crete. She is an associate editor of the Journal of Applied Stochastic Models in Business and Industry and of the Journal of Statistical Distributions and Applications and has acted as chief editor of the journal of Quality Technology and Quantitative Management. Her research interests include Statistical Indices, Model Evaluation and Selection, Inventory Decision Models, Statistical aspects of ARCH and stochastic volatility, and Over-dispersion and under-dispersion modeling.