Sociedad de Estadística e Investigación Operativa
Test (1998) *Vol.* 7, *No.* 1, pp. 1–74

The stochastic control of process capability indices

Nozer D. Singpurwalla
The George Washington University
Washington, DC 20052

Abstract

In manufacturing science, process capability indices play a role analogous to economic indices in government statistics. The existing capability indices are passive devices whose main role is to retroactively monitor process capability. They have been developed under the restrictive assumption of process stability, and the procedures for using them are based on ad hoc rules. Using the normative point of view for decision making, it can be shown that some of the indices are, at best, convoluted special cases of a more general strategy; they can be justified only under special assumptions, and the manner in which they are currently used could lead to incoherent actions. The available process capability indices should therefore be abandoned and replaced by procedures that are normative, and also proactive with respect to both, prediction and control. An approach towards achieving this goal is proposed.

Key Words: Process capability indices, quality control, manufacturing science, tolerances, control variables.
AMS subject classification: 62N10, 62A15, 90B30, 93E20.

1 Introduction

A *process capability index* is a dimensionless measure that is used to quantify the performance of a manufacturing process vis-a-vis the process parameters and the product specifications. Process capability indices – henceforth PCI – have recently attracted much attention in engineering design, manufacturing science, and quality assurance. Indeed, in the manufacturing sciences they hold a status akin to that held by economic indicators, like the consumer price index, in government statistics. These indices have been used to *assess*, and to *monitor*, the quality of units produced by a manufacturing process, and thus serve as indicators of the overall capability of a manufacturing system. Their purpose is to ensure (but only

Research Sponsored by The National Institute of Standards and Technology Gaithersburg, Maryland 20899-0001 (Under Purchase Order No. 43NANB610868), The U.S. Army Research Office Grant DAAG-55-97-1-0323, and The Air Force Office of Scientific Research Grant AFOSR-F-49620-95-0107

retrospectively) that the number of nonconforming items in a batch are below a specified limit.

One of the simplest, and perhaps the most commonly used index, is what is known as the "six sigma" index, denoted by C_p [cf. Harry and Lawson (1992)]. This index has been used to estimate the proportion of parts that do not meet engineering specifications. The C_p index, originally proposed by Juran (1974), has its foundation in a fundamental result of probability, namely Tchebychev's inequality; see for example Billingsley (1986), p. 75. The essence of this inequality is the result that the probability of any random variable deviating from its mean by more than three times its standard deviation is small – at most 0.1. This inequality, though sharp, is too broad and too general to be of much practical value; it demands of the user only a knowledge of the variance. Consequently, despite its common use by industry, enhancements and refinements of C_p have been proposed, most of these in the literature on statistical quality control. Each enhancement requires added knowledge of some aspect of the distribution of the random variable. That is, the enhancements assume more than just a knowledge of the variance. An overview of the various PCI's, and a brief discussion of their properties, is given in Section 2. The overview is not intended to be an exhaustive survey; its purpose is to give the reader a flavor of the developments. A more complete perspective is in the recent monograph by Kotz and Lovelace (1997).

It is important to note that the PCI's have, to date, played only a passive role in the manufacturing sciences. The functions of assessing and monitoring are not predictive nor are they proactive, and thus the available PCI's mainly serve as policing devices. Whereas this by itself is a needed activity, a much more useful role can be served by the PCI's if they can also be used to predict and to control the quality of future output. The theme of this paper is to advocate and to develop the idea that process capability indices be made to serve the dual role of assessing and monitoring current quality, as well as predicting and controlling the quality of future output. This theme is developed in Section 3. Section 4 pertains to a demonstration of how the various indices behave when confronted with some real and simulated data.

2 Process capability indices: an overview

2.1 Preliminaries

Let X be some characteristic of interest of a manufactured product, such as say the diameter of a pin, or the length of a shaft. The engineering or design specifications for X are generally stated in terms of a "nominal" or a "target value", say T. That is, T is that value of X which will satisfy the design engineer's criteria for the optimum performance of the product. Now manufacturing the product so that X exactly equals T is prohibitively expensive, and so it is common practice to specify upper and lower "specification" limits, USL and LSL, respectively, and to require that X be within these limits. We do not concern ourselves here with the question of how USL and LSL ought to be specified, though this too is a matter of fundamental importance and should be addressed by the dictates of coherent decision making [cf. Singpurwalla (1992)].

The physical processes that manufacture the part are generally subject to many sources of variation, starting from the quality of the raw material to the aging and wear-out of the manufacturing equipment. Consequently, X is a random quantity (or random variable), whose distribution is often assumed to be a Gaussian with a mean, say μ, and a variance, say σ^2. In manufacturing parlance, the variance σ^2 is referred to as the *natural tolerance* of X. When working with the PCI's it is also a common practice to assume that both μ and σ^2 do not change with time; i.e. the process is *stable*, or what is known in quality control, as *in statistical control.* The assumption of a Gaussian distribution for X is perhaps reasonable, because often X is a quantity that can be measured, and measurements being subjected to symmetric errors are, by tradition – since the time of Gauss – assumed to be Gaussian. However, Bernardo and Irony (1996) raise issue with this assumption.

The question which arises is as to whether the design engineer's compromise in going from the ideal T to the upper and lower specification limits (the USL and the LSL), is matched by the manufacturer's ability to meet such a compromise vis-a-vis the assumed μ and σ^2 mentioned above. The PCI's were introduced to address this matter. But before describing the PCI's some additional notation that is useful in the subsequent text needs to be introduced. The quantity (USL-LSL) is known as the *specification* *interval*; it is denoted by $2d$, where d is the half length of the specification interval. The mid point of the specification interval, denoted by M, is (USL+LSL)/2.

2.2 The traditional indices: their chronological development

It was stated before that the earliest index, namely C_p, was introduced by Juran in 1974, and that C_p has its foundation in Tchebychev's inequality which requires a knowledge of only σ^2. This index, defined as

$$C_p = \frac{\text{USL-LSL}}{6\sigma} = \frac{d}{3\sigma}, \tag{2.1}$$

does not involve the *process mean* μ. Suppose that the target value T equals M, the mid point of the specification interval, and that X has a Gaussian distribution. Observe that C_p is the ratio of the allowable spread of the process to its actual spread, and that if μ coincides with T, then the value $C_p = 1$ will imply that 99.7% of the produced items will fall within the specification limits. To see this, observe that

$$\begin{aligned} P(\text{LSL} < X < \text{USL}) &= P\left(\frac{\text{LSL} - \mu}{\sigma} \le \frac{X - \mu}{\sigma} \le \frac{\text{USL} - \mu}{\sigma}\right) \\ &= \Phi(3C_p) - \Phi(-3C_p) = \Phi(3) - \Phi(-3) = .997, \end{aligned}$$

where $C_p = 1$, and $\Phi(x)$ is the distribution function of the unit Gaussian; i.e.

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du.$$

As long as μ coincides with T, any value of C_p greater than one will increase the above probability, making the manufacturing process more efficient. Since σ^2 is unknown, it has to be inferred from the data, and to compensate for the uncertainties of estimation, industrial practice follows the dictum that C_p must be a minimum of 1.33 (instead of the aforementioned 1). The choice 1.33 is completely *ad hoc*; indeed for pilot (or qualification) runs, C_p is sometimes required to be in excess of 1.66. A possible explanation for this choice of C_p is that 1.66 corresponds, approximately, to a reject rate of one unit per million; see, for example, Juran and Gryna (1980). Since large values of C_p would increase the cost of manufacturing, the critical value of C_p should be based on a formal decision theoretic development; see Section 2.3. The data needed to estimate σ^2, mentioned above, is taken at certain specified points in time, called *rating periods.* The specification of the rating periods also appears to be based on arbitrary considerations.

The numerical claim made above is valid only when μ and T coincide. Should μ not coincide with T, then $C_p = 1$ will not yield the 99.7% figure cited before; it will be smaller. To see why, suppose that μ is at (USL+T)/2; i.e. μ is between the target value and the USL. Then

$$P(\text{LSL} < X < \text{USL}) = \Phi(1.5) - \Phi(-4.5) = .993,$$

so that the proportion of nonconforming parts has more than doubled.

To incorporate the effect of the process mean on the capability index, C_p was refined by introducing the index C_{pk}, where

$$C_{pk} = \text{Minimum}\left\{\frac{\text{USL} - \mu}{3\sigma}, \frac{\mu - \text{LSL}}{3\sigma}\right\},$$

and it is assumed that $\mu \in$ (LSL, USL). This refinement of C_p, presumably originated in Japan, has often been attributed to Kane (1986). An appreciation of this choice, namely that of C_{pk}, can be obtained via an examination of Figure 1.

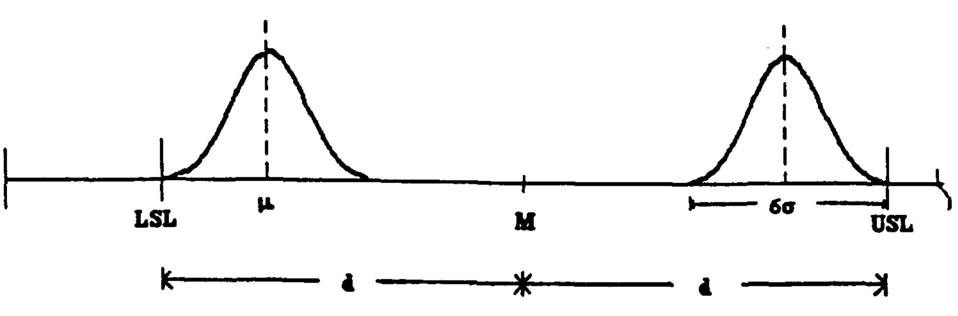

Figure 1: Considerations in defining C_{pk}.

Figure 1 illustrates the fact that the largest value that μ can take, and the allowable spread be still greater than the natural spread, is USL-3σ; similarly, the smallest value that μ can take is LSL+3σ. These considerations suggest that $(\text{USL} - \mu)/3\sigma \geq 1$ and that $(\mu - \text{LSL})/3\sigma \geq 1$. Furthermore, the closer that μ gets to the specification limit, the bigger is the likelihood of producing nonconforming parts; hence the appearance of a minimum in the formula for C_{pk}. Thus, when using C_{pk}, the process is judged capable if C_{pk} is one or greater, and since both μ and σ^2 are to be estimated from the data, industrial practice follows the rule that C_{pk} be greater than 1.33. There is a recent move in industry, especially that which supplies products for the government, to require that C_{pk} be much bigger than 1.33, but once again such decisions should involve a trade-off between costs of manufacture and the consequences of being out of specifications, not an arbitrary specification of some cut-off point.

Verify, see Figure 1, that C_{pk} may also be written as

$$C_{pk} = \frac{d - |\mu - M|}{3\sigma}, \tag{2.2}$$

and since

$$C_{pk} = \frac{d - |\mu - M|}{3\sigma} = C_p - \frac{|\mu - M|}{3\sigma}, \tag{2.3}$$

C_{pk} is smaller than C_p whenever $\mu \neq M$, and $C_{pk} = C_p$ when $\mu = M$. Furthermore $C_{pk} \geq 1$ implies that $C_p \geq 1$, because the last term of (2.3) cannot be negative. The comparative behavior of C_p and C_{pk} has been illustrated by Fries and Richter (1993) who make the point that $C_p \geq 1$ does not imply that $C_{pk} \geq 1$.

It is interesting to note that both C_p and C_{pk} do not involve the target value T, unless it is tacitly assumed that $T = M$. Hsiang and Taguchi (1985) [and also Chan, Cheng, and Spiring (1988)], rectify this omission by replacing the natural tolerance of X, namely, $\sigma^2 \stackrel{\text{def}}{=} E(X - \mu)^2$, with the quantity $E(X - T)^2$, in the denominator of C_p; in so doing, they introduce a new index C_{pm}. Since $E(X - T)^2 = E(X - \mu + \mu - T)^2$, the index C_p now takes the form

$$C_{pm} = \frac{\text{USL-LSL}}{6\sqrt{E(X - T)^2}} = \frac{d}{3\sqrt{\sigma^2 + (\mu - T)^2}}. \tag{2.4}$$

Verify that since C_{pm} can also be written as

$$C_{pm} = \frac{C_p}{\sqrt{1 + \frac{(\mu - T)^2}{\sigma^2}}},$$

an approximation to C_{pm} can be obtained by expanding a function of the form $1/\sqrt{1+x^2}$, and taking the first two terms to yield a new index

$$C_{pq} = C_p \left[1 - \frac{1}{2} \left(\frac{\mu - T}{\sigma} \right)^2 \right];$$

this index was introduced by Gupta and Kotz (1997).

2.2.1 Interpretation of C_{pm}

Whereas (2.4) does have the advantage that it incorporates all the relevant parameters, namely d, μ, T, and σ^2, the geometric intuition which guided the construction of C_p and C_{pk} is now missing. Clearly, when $\mu = T$, $C_{pm} = C_p$, and as μ deviates from T, the denominator of (2.4) gets quadratically inflated; consequently, $C_{pm} \leq C_p$. This suggests that a deviation of the process mean μ from the target T manifests as a de facto *penalty* whereby in the formula for C_p, the natural tolerance (i.e. the variance) of the process gets inflated. The penalty increases quadratically in $|\mu - T|$, but it bears no relationship to the USL and the LSL. With the above interpretation, one would want C_{pm} to be at least one, because with $\mu = T$, $C_{pm} = C_p$, and the process is judged capable when $C_p \geq 1$. Recall that C_p was interpreted as the ratio of the allowable spread to the actual spread of the process. With this in mind, C_{pm} could be interpreted as the ratio of the allowable spread to the actual spread, where the latter is the natural tolerance of the process plus an inflation factor that depends on the deviation of μ from T.

There is another perspective on C_{pm}. Conceptually, the difference between C_p and C_{pm} is that in constructing the former the denominator is $E(X - \mu)^2$, whereas in the latter it is $E(X - T)^2$. The intuition behind C_p is straightforward – a comparison between the allowable and the natural (manufacturing) tolerance. How does one extend this intuition when the divisor is $E(X - T)^2$? One possibility is to view $(X - T)^2$ as a *loss incurred* when X deviates from T, so that $E(X - T)^2$ is the (weighted) average loss incurred with the weights determined by the Gaussian distribution of X with mean μ and variance σ^2. Under this interpretation, the requirement that $C_{pm} \geq 1$, implies that the maximum value that the average loss can take is $d^2/9$. It is not necessary that the loss be a quadratic function of $|X - T|$; it could be any function of $|X - T|$, such as say $|X - T|^a$, for $-\infty < a < +\infty$. A traditional loss function is the indicator function, taking the value 0 when LCL $\leq X \leq$ UCL, and a constant, say 1, otherwise. When such is the case the expected loss is $2\Phi\left(\frac{\text{LSL}-\mu}{\sigma}\right)$, so that

$$C_{pm} \geq 1 \quad \text{iff} \quad 3\sqrt{2}\left[\Phi\left(\frac{\text{LSL}-\mu}{\sigma}\right)\right]^{1/2} \leq d.$$

One may inquire about the necessity of choosing the number six in the denominator term of (2.4) when $E(X - T)^2$ is interpreted as a loss. The only rationale for choosing the six is the equivalence between C_{pm} and C_p when $\mu = T$, and the fact that C_p carries with it a natural interpretation. Similarly, choosing the square root of $E(X - T)^2$ not only ensures equivalence with C_p, but like C_p, it makes C_{pm} a dimensionless quantity.

2.2.2 Limiting behavior of C_p, C_{pk}, and C_{pm}

Observe that both C_p and C_{pk} become arbitrarily large as $\sigma^2 \downarrow 0$, irrespective of where the process is centered; i.e. $\mu = T$ or $\mu \neq T$. By contrast C_{pm} is bounded by the quantity $d/3(|\mu - T|)$ as $\sigma^2 \downarrow 0$, and this is attractive. Since $|\mu - T| < d/3C_{pm}$, the value $C_{pm} = 1$ implies that the process mean μ lies within the middle third of the specification range d.

Finally, whereas C_p is able to provide a probability of nonconformance (recall that $P(\text{LSL} < X < \text{USL}) = \Phi(3C_p) - \Phi(-3C_p)$), C_{pk} and C_{pm} do not. However, both C_{pk} and C_{pm} provide an upper bound on the probability of non-conformance. These bounds turn out to be $2\Phi(-3C_{pk})$ and $2\Phi(-3C_{pm})$, respectively; see Pearn et al (1992).

2.2.3 Sensitivity to departures from target value

Because C_p is independent of the target value T, it is robust against departures of the process mean μ from T. This of course is a drawback of C_p. To display this lack of sensitivity, Kushler and Hurley (1992), also Wallgren (1996 - personal communication), plot $\sigma = d/3C_p$ versus $(\mu - T)$; see Figure 2. Specifically, if $C_p = c$, then a plot of $d/3c$ versus $(\mu - T)$ will be a constant at $d/3c$. By contrast, if $C_{pk} = c$, and $T = M$, then a plot of

$\sigma = \frac{d}{3c} \pm \frac{\mu - T}{3c}$ will display the feature that in order to keep C_{pk} a constant at c, σ will have to decrease linearly as $|\mu - T|$ increases (see Figure 2), and it will be zero when $|\mu - T| = d$; σ is $d/3c$ when $|\mu - T| = 0$, as is to be expected. To investigate the sensitivity of C_{pm} to departures of μ from T, suppose that $C_{pm} = c$; then (2.4) defines the equation to a semi-circle (σ cannot take negative values) with radius $d/3c$, and origin at $(\mu - T) = 0$:

$$\sigma^2 + (\mu - T)^2 = (d/3c)^2;$$

see Figure 2.

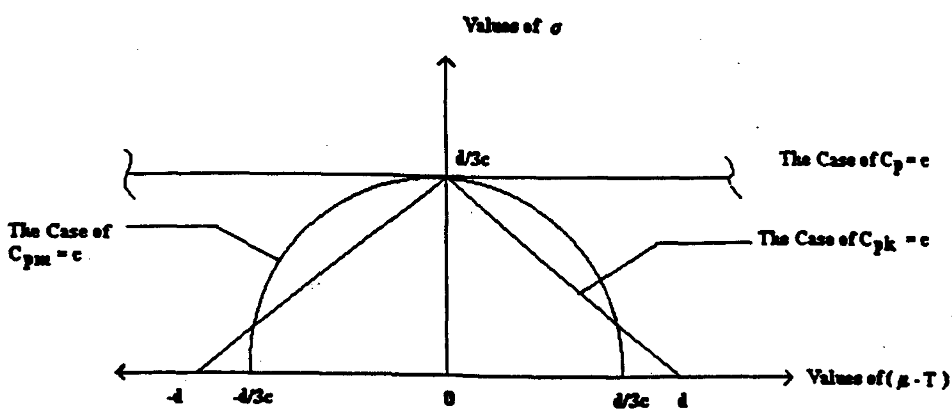

Figure 2: Sensitivity of C_p, C_{pk}, and C_{pm}.

Figure 2 shows, that for the case $C_p = C_{pk} = C_{pm} = c$, a small deviation from the target value calls for a large reduction in σ for the case of C_{pk} as compared to that of C_{pm}. That is, for small deviations of μ from T, C_{pk} is more sensitive than C_{pm}. However, the situation reverses when the deviations become large. Specifically, as soon as $|\mu - T| \geq 2d/(9c^2 + 1)$, C_{pm} becomes more sensitive than C_{pk}. Thus, if a choice between C_{pk} and C_{pm} is to be based on sensitivity to departures of μ from T, then one must have a priori information about the extent of the disparity between μ and T, and also about the values of C_{pk} and C_{pm}. For example, if $d = 1$, and if $C_{pk} = C_{pm} = 1$, then C_{pk} will be preferred to C_{pm} if it is expected that $|\mu - T| < 1/5$.

2.2.4 The index C_{pmk} and its generalization $C_p(u,v)$

To devise an index that is more sensitive to departures of μ from T, Pearn, Kotz, and Johnson (1992) introduced a new index, the index C_{pmk}, which takes its numerator (denominator) the numerator (denominator) of C_{pk} (C_{pm}); that is, C_{pmk} is hybrid:

$$C_{pmk} = \frac{d - |\mu - M|}{3\sqrt{\sigma^2 + (\mu - T)^2}}. \tag{2.5}$$

An interpretation of the denominator of (2.4) in terms of a weighted loss was given in Section 2.2.1. It was stated there, that in order to have $C_{pm} \geq 1$, the maximum value for the average loss was limited to $d^2/9$ or equivalently to $(\text{USL-LSL})^2/36$. The numerator of (2.5) differs from that of (2.4) by an amount $|\mu - M|$. Thus, were one to interpret the denominator of (2.5) as the weighted loss, then the numerator of C_{pmk} restricts this loss to $(d - |\mu - M|)^2/9$, which when μ deviates from M is smaller than $d^2/9$, the maximum allowable average loss under C_{pm}. Thus in C_{pmk}, there is a de facto penalty for μ deviating from M, and the penalty manifests itself by restricting the loss incurred when μ deviates from T by a quadratic function of $(d - |\mu - M|)$. When $\mu = M$, this penalty vanishes. Because of these considerations, the index C_{pmk} is judged to have attractive features.

Like C_{pm} and C_{pk}, C_{pmk} limits the probability of nonconformance to $2\Phi(-3C_{pmk})$, and when $T = M$, C_{pmk} is bounded above, as $\sigma^2 \downarrow 0$, as $C_{pmk} < d/(3|\mu - T|) - \frac{1}{3}$. Vännaman (1995) shows that among all the indices presented thus far, C_{pmk} is the most sensitive to departures of μ from T.

Whereas the index C_{pm} had the attractive feature that it incorporated the parameters d, μ, T, and σ^2, it did have a glaring omission; namely, the parameter M. The index C_{pmk} rectifies this deficiency, and in doing so makes the indices C_p, C_{pk}, and C_{pm}, its special cases. This feature was exploited by Vännaman, who introduced the index

$$C_p(u,v) = \frac{d - u|\mu - M|}{3\sqrt{\sigma^2 + v(\mu - T)^2}}, \tag{2.6}$$

which with $u = 0$, $v = 0$, reduces to C_p, with $u = 1$, $v = 0$, reduces to C_{pk}, and with $u = 0$, $v = 1$, reduces to C_{pmk}.

The parameters u and v of (2.6) are not just indexing devices whose main purpose is to act as binary "switches". They need not be restricted to the values 0 and 1. Indeed larger values of u and v serve to make $C_p(u, v)$ more and more sensitive to departures of μ from T.

2.2.5 Subsequent developments: generalizations and refinements

Much of the recent statistical literature on PCI's focuses on issues such as sample sizes needed to estimate μ and σ^2, tests of hypotheses about the indices, the effects of skewness [cf. Pearn et al (1992), Wright (1995)] and the effects of correlations between the observed values of X, when estimating the indices. Also considered, are issues such as asymmetric and one sided tolerances [cf. Vännman (1996 a, b)]; asymmetric tolerances occur when $T \neq M$. Other suggestions involve the use of Bayesian ideas to reduce the sample sizes required for estimation [cf. Cheng and Spiring (1989)], and multivariate versions of the PCI's [cf. Taam, Subbaiah, and Liddy (1993), Wierda (1993), and Chen (1994)]. Bayesian ideas have also been used by Bissell (1990) and Pearn and Chen (1996) to produce estimators of C_{pk} that have good frequentist properties.

2.3 Discussion and critique of the PCI's

In all the work that has been referenced thus far, there is a strong underlying assumption, namely that the process under study is *stable* (or *in statistical control*). This implies that within any two rating periods, the underlying distribution which generates the indices does not change (over time). Thus gradual drifts in the process mean and/or the process variance are not allowed. [However, in going from one rating period to the other, the process could be out of control, and the PCI's are designed to detect such limits]. Furthermore, one of the key issues, namely how the PCI should be actually put to use appears to be short changed. The dictum that C_p and C_{pk} should be 1.33 for production runs, and bigger that 1.66 for qualification runs, is completely *ad hoc*. More disturbing is the current move in industry which wants to set the lower limits on the indices to be much greater than 1.33 (or 1.66). What would be the consequences of such a move, especially if it is arbitrary, on the costs of manufacturing? Will the consumer end up paying for this ad hoc move which can be achieved by reducing σ^2 or by increasing d? The passive role played by the PCI's as pure monitoring devices was mentioned before, but needs to be repeated, because it is indeed a drawback.

The normative approach for the control of quality is based on decision theoretic considerations. It provides a vehicle for accomplishing both, the retroactive function of assessment and monitoring, and the proactive functions of prediction and control. Furthermore, the normative approach is able to integrate the three tasks of assessment, prediction, and control within an interactive and unifying framework. Here, one monitors the observable X (rather than the unobservable μ), and makes a decision to continue production, to modulate it, or to stop it, based on the consequences of the deviation of X from T. The decision is proactive and is dictated by the predictive distribution of X and the utilities associated with the deviation of X from T, and also the utilities associated with a control of the process.

To the best of knowledge, the work of Bernardo and Irony (1996) appears to be first to have introduced the normative approach in the context of PCI's. Their work considers the two decisions, to continue production or to stop it; it is overviewed in the next section. The incorporation of the third decision, namely control is introduced here; it is discussed in Section 3.

2.4 The normative approach to process capability

The normative approach to process capability, pioneered by Bernardo and Irony (1996), also starts with the assumption of process stability (i.e. statistical control); however, as will be pointed out later, this requirement can be eased. It departs from the traditional approaches by emphasizing decision making based on predictive values, and in so doing paves the path towards proactiveness in manufacturing based on process capability analysis; however, it falls short by not incorporating the issue of control. Since some of the traditional PCI's turn out to be special cases of the normative set-up, the work of Bernardo and Irony is to be classified a signal contribution to the literature on process capability analysis. This accolade is deserving, because it essentially says that when the available CPI's are viewed in the broader context of decision making, they become at best, convoluted special cases of a more encompassing strategy. Given below is an overview of Bernardo and Irony's development supplemented with other features and attractive properties that have not been remarked upon by the above authors.

Let $0 < T_1 < T_2 < \ldots < T_{i-1} < T_i < \ldots$, denote the rating periods, and suppose that at T_i measurements $x_{i1}, \ldots, x_{in}$, are taken on some attribute of interest, on any n items that were produced during the time period $[T_{i-1}, T_i)$; see Figure 3.

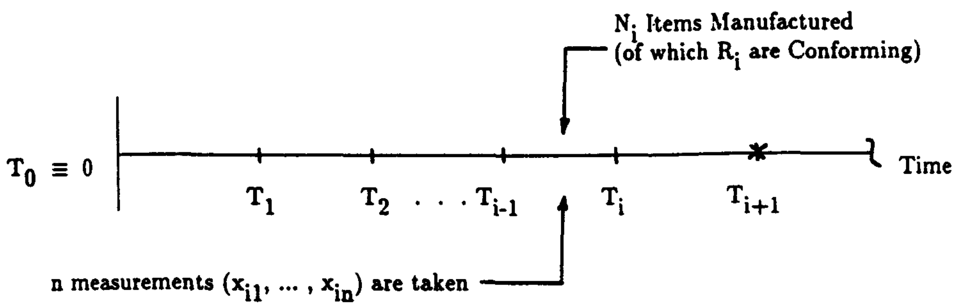

Figure 3: Illustration of rating periods and sample measurements.

At T_i two tasks have to be undertaken: i) an assessment has to be made if a specified proportion of the N_i items that were manufactured during the period $[T_{i-1}, T_i)$ has met the engineering tolerances or not; and ii) a decision has to be made whether to continue production *as is*, during the coming period $[T_i, T_{i+1})$, or to *intervene* in the production process and make the needed changes. If it is assumed that the process is stable, both within and between the rating periods, then the above two tasks boil down to a common calculation. The "continue production as is" decision is denoted d_1, and the "intervene" decision, by d_2. The T_i's, N_i's, $i = 1, 2 \ldots$, and the n are assumed to be fixed and known; an investigation about the optimal choice for T_i and n would be worthwhile. Let

$$\underset{\sim}{x}_i(n) = (x_{i1}, x_{i2}, \ldots, x_{in}), \quad \text{and}$$

$$\underset{\sim}{X}_{i+1}(N_{i+1}) = (X_{i+1,1}, X_{i+1,2}, \ldots, X_{i+1,N_{i+1}});$$

the former represents observed data, and the latter, the unobserved values of the characteristic of interest for the future production.

Coherent decision making [cf. Bernardo and Smith (1994), Chapter 1], requires that the utilities $\mathcal{U}(d_1, \underset{\sim}{x}_{i+1}(N_{i+1})) \stackrel{\text{def}}{=} \mathcal{U}(d_1)$, and $\mathcal{U}(d_2, \underset{\sim}{x}_{i+1}(N_{i+1})) \stackrel{\text{def}}{=} \mathcal{U}(d_2)$, be specified, where $\mathcal{U}(d_j, \underset{\sim}{x}_{i+1}(N_{i+1}))$ is the consequence of choosing decision d_j, $j = 1, 2$, when $\underset{\sim}{X}_{i+1}(N_{i+1})$ takes the value $\underset{\sim}{x}_{i+1}(N_{i+1})$. Let $\pi(\underset{\sim}{x}_{i+1}(N_{i+1})|\underset{\sim}{x}_i(n))$ be the *predictive distribution* of $\underset{\sim}{X}_{i+1}(N_{i+1})$, the future observables, given $\underset{\sim}{x}_i(n)$, the observed data from the latest rating period. This predictive distribution is relatively straightforward to compute if the process is in control. Then, the expected utility of decision d_j, $j = 1, 2$, is

$$\overline{\mathcal{U}}(d_j) = \int \mathcal{U}(d_j, \underset{\sim}{x}_{i+1}(N_{i+1}))\pi(\underset{\sim}{x}_{i+1}(N_{i+1})|\underset{\sim}{x}_i(n))\, d\underset{\sim}{x}_{i+1}(N_{i+1}). \qquad (2.7)$$

Decision d_1 is chosen if $\overline{\mathcal{U}}(d_1) \geq \overline{\mathcal{U}}(d_2)$; otherwise d_2 is chosen. The above rule is a consequence of the principle of *maximization of expected utility*. Whereas enunciating this principle is straightforward, implementing it tends to be demanding, and this could be the only excuse for justifying a use of the traditional PCI's.

To simplify matters suppose that $N_i = N$, $i = 1, 2, \ldots$, and suppose that R_i of the N_i produced items were to conform to specifications. Of course at time T_i, R_{i+1} is unknown, and at time T_{i-1}, R_i can only be known if $n = N_i$. A simple form of the utility of decision d_1 (continue as is) could be

$$\mathcal{U}(d_1) = gR_{i+1} - C(N - R_{i+1}) - lN,$$

where g is the profit from delivering a good item, C is the penalty incurred by delivering a defective item, and l is the cost of manufacturing an item. The utility of d_2, the action to intervene, could take different forms, depending on the nature of intervention.

If the intervention involves inspecting every manufactured item and repairing all nonconforming ones, then

$$\mathcal{U}(d_2) = gN - hN - r(N - R_{i+1}) - lN,$$

where h is the cost of inspecting each item and r is the cost of repairing a failed item. If the intervention involves freezing the production to overhaul the entire manufacturing system, then a fixed cost Q is incurred, plus generally there is a cost due to opportunity lost in supplying the N items. If

q denotes the unit cost of lost opportunity, then $\mathcal{U}(d_2) = -Q - qN$. If Q is amortized over the next k rating periods then, $\mathcal{U}(d_2)$ could be of the form $\mathcal{U}(d_2) = -Q/k - qN$. In any case, it turns out to be so that $\mathcal{U}(d_1) - \mathcal{U}(d_2)$ is of the form $aR_{i+1} - bN + Q^*$, and Q^* a constant greater than or equal to 0. With the above simplification, decision d_1 is chosen if [see (2.7)]

$$\sum_{R_{i+1}=0}^{N} (aR_{i+1} - bN + Q^*)\pi(R_{i+1}|N, \underset{\sim}{x}_i(n)) \geq 0,$$

or that

$$aE(R_{i+1}|N, \underset{\sim}{x}_i(n)) - bN + Q^* = a \cdot N \cdot P(X \in A|\underset{\sim}{x}_i(n)) - bN + Q^* \geq 0,$$

if the $X_{i+1,1}, X_{i+1,2}, \ldots, X_{i+1,N_{i+1}}$ are judged to be (conditionally) independent and identically distributed, and where A is the specification interval [LSL, USL]. Equivalently, decision d_1 is chosen iff

$$P(X \in A|\underset{\sim}{x}_i(n)) \geq \frac{bN - Q^*}{aN} = \frac{b}{a} - \frac{Q^*}{aN} \tag{2.8}$$

In words, the essence of (2.8) is that production for the rating period $[T_i, T_{i+1})$ should be continued as is, only if the expected proportion of conforming items in the rating periods $[T_{i-1}, T_i)$ and $[T_i, T_{i+1})$ exceeds a threshold. The left hand side of (2.8) is the predictive distribution of X, and the threshold, which is the right hand side of (2.8), is related to the underlying costs. When $\mathcal{U}(d_1) = gh_{i+1} - C(N - R_{i+1}) - lN$, $Q^* = \widehat{0}$ and the threshold is simply b/a, where $a = g + C - r$ and $b = a - h$, which for $h \geq 0$ is always less than or equal to a. With $\mathcal{U}(d_2) = -Q/k - qN$, $Q^* = Q/k$, $a = g + C$, and $b = C + l - q$.

Since the left hand side of (2.8) is always positive, and bounded between 0 and 1, a must be greater than or equal to b, irrespective of the form of $\mathcal{U}(d_2)$. Since $a \geq b$, the first term on the right hand side of (2.8) is always less than or equal to one, and if N is made large so that $\frac{Q^*}{aN}$ is very small, the right hand side of (2.8) can also be bounded between 0 and 1. If Q^*/N is larger than b, then the right hand side of (2.8) is negative, in which case the decision d_1, to continue production as is, is always taken.

How does one relate the condition (2.8) to that required of the traditional PCI's? Recall that the PCI's are often required to take values greater than 1. If $P(X \in A|\underset{\sim}{x}_i(n))$ is set equal to u, then one strategy would be to take a monotonic transformation of the terms on both sides of (2.8). A possibility, and the one proposed by Bernardo and Irony, is the *probit transform* which results in the criterion:

$$\text{choose} \quad d_1 \quad \text{iff} \quad \frac{1}{v}\Phi^{-1}(u) \geq \frac{1}{v}\Phi^{-1}\left(\frac{bN - Q^*}{aN}\right) \tag{2.9}$$

where $\Phi(x)$ was defined in Section 2.2, and v is some positive constant, whose role will become clear later. The quantity

$$\frac{1}{v}\Phi^{-1}(u) = \frac{1}{v}\Phi^{-1}(P(X \in A|\underset{\sim}{x}_i(n))), \tag{2.10}$$

has been defined by Bernardo and Irony as the *Bayes capability index*, denoted $C_B(v)$. Instead of the probit transform one may also consider the log-odds transform of the type $\log \frac{u}{1-u}$.

If the predictive distribution of X (given $\underset{\sim}{x}_i(n)$), is a Gaussian with mean μ and variance σ^2 - and this can be made to happen under some very general conditions involving n large - then

$$P(X \in A|\underset{\sim}{x}_i(n)) = \Phi\left(\frac{\text{USL} - \mu}{\sigma}\right) - \Phi\left(\frac{\text{LSL} - \mu}{\sigma}\right), \tag{2.11}$$

and now (2.9), the case of the probit transform reduces to the statement:

$$\text{choose } d_1 \quad \text{iff} \quad \frac{1}{v}\Phi^{-1}\left(\Phi\left(\frac{\text{USL} - \mu}{\sigma}\right) - \Phi\left(\frac{\text{LSL} - \mu}{\sigma}\right)\right) \geq \frac{1}{v}\Phi^{-1}\left(\frac{bN - Q^*}{aN}\right),$$

which with $v = 6$ boils down to the condition:

$$\text{choose} \quad d_1 \quad \text{iff} \quad \left(\frac{\text{USL-LSL}}{6\sigma}\right) \geq \frac{1}{6}\Phi^{-1}\left(\frac{bN - Q^*}{aN}\right). \tag{2.12}$$

Note that the μ and the σ^2 of the predictive distribution of X are based on the prior distribution of X and the sample mean and sample standard deviation of the collection $\underset{\sim}{x}_i(n)$. Incorporation of the prior distribution makes the predictivity feature of $C_B(\nu)$ more realistic vis-a-vis known changes in the manufacturing process. Thus, in using $C_B(\nu)$, the premise of process stability is not an essential one.

The left hand side of (2.12) is precisely C_p, the PCI introduced by Juran. The right hand side of (2.9) with $\nu = 6$ is interesting to analyze. For example, if $\frac{1}{6}\Phi^{-1}(\frac{bN-Q^*}{aN})$ is required to be 1 (or greater), then $\Phi^{-1}(\frac{bN-Q^*}{aN})$ must be at least 6, in which case $\frac{bN-Q^*}{aN} = \Phi(6) \approx 1$, or that $(a-b) = -Q^*/N$, which, since $Q^* \geq 0$ can only happen if $b \geq a$. Thus, requiring C_p to be at least one leads to a contradiction, compelling one to the claim that the traditional use of the process capability index, C_P, leads to incoherence!

Suppose now that μ is not centered at M, but is in the vicinity of LCL in such a way that $(\text{USL}-\mu) \gg (\mu-\text{LSL})$; also suppose that $(\text{USL}-\mu) \gg \sigma$. That is, the process is said to be *non-centered but potentially capable.* Then $\Phi(\frac{\text{USL}-\mu}{\sigma}) \approx 1$, and with $\nu = 1/3$, the left hand side of (2.9) becomes

$$\frac{1}{3}\Phi^{-1}\left(\Phi\left(\frac{\text{USL}-\mu}{\sigma}\right) - \Phi\left(\frac{\text{LSL}-\mu}{\sigma}\right)\right) = \frac{1}{3}\Phi^{-1}\left(1 - \Phi\left(\frac{\text{LSL}-\mu}{\sigma}\right)\right)$$
$$= \frac{1}{3}\Phi^{-1}\left(\Phi\left(\frac{\mu-\text{LSL}}{\sigma}\right)\right) = \frac{\mu-\text{LSL}}{3\sigma} = C_{pk},$$

since $(\mu - \text{LSL}) \gg (\text{USL} - \mu)$.

Conversely, if $(\mu - \text{LSL}) \gg (\text{USL} - \mu)$, and again if $(\mu - \text{LSL}) \gg \sigma$, then the left hand side of (2.9) would be $(\frac{\text{USL}-\mu}{3\sigma}) = C_{pk}$.

With $v = 3$, the right hand side of (2.9) would be 1 if $\frac{bN-Q^*}{aN} = \Phi(3) = .9985 \approx 1$, so that here again requiring that C_{pk} be greater than 1 leads to incoherence.

To summarize, the normative approach for process capability analysis is able to produce the indices C_p and C_{pk} but only under the assumptions of specific forms of the utility function – namely of the indicator type discussed in Section 2.2.1 – and under the probit transform Φ. Furthermore, the indices are produced when n tends to be large, and when the process is non-centered, but is potentially capable. Finally, the indices as traditionally used, lead to incoherent actions.

What has been discussed so far pertains to decision making at T_i. As stated before, it assumes that the process is stable from one rating period to the next. The question of assessment at T_i is rather straightforward, since the proportion of the N_i (or $N_i = N$) items produced during $[T_{i-1}, T_i)$ that conform to specifications is simply $P(X \in A|\underset{\sim}{x}_i(n))$, which for large n is given as

$$\Phi\left(\frac{\mathrm{USL}-\mu}{\sigma}\right)-\Phi\left(\frac{\mathrm{LSL}-\mu}{\sigma}\right).$$

All of the above calculations are based on the premise that the process is stable within and between rating periods. As discussed in Section 3, appropriate adjustments can be made if the process experiences a drift either within or between rating periods, or if the X_i's cannot be judged independent. This facility is not available with the traditional indices.

To conclude, the normative approach-as presented above-provides a vehicle for assessment and action, the former being retroactive and the latter being predictive, but not proactive. Can this approach be enhanced so that it is also proactive in the sense of being able to influence the quality of future production? The answer to this question is in the affirmative; it is explored next.

3 The control of process capability

The normative approach to process capability, discussed in Section 2.3, assumed that the process is stable or in statistical control, both between any two rating periods, and also when going from one period to the next. It is because of this assumption that a common predictive distribution, namely $P(X \in A|\underset{\sim}{x}_i(n))$, was able to serve two roles: i) to obtain $E(R_{i+1}|N, \underset{\sim}{x}_i(n))$ for making decisions about the future production (i.e. the production during the interval $[T_i, T_{i+1})$),and ii) for assessing the proportion of conforming items manufactured during the period $[T_{i-1}, T_i)$. Clearly, under the assumption of process stability, the decision to "continue production as is" is equivalent to the judgment that the proportion of units that meet specifications is a desirable one. But process stability excludes drifts and sudden shifts that characterize the realities of manufacturing, and thus this assumption is fundamentally idealistic.

As a starting step towards weakening the assumption of process stability, assume that the process is stable between any two rating periods, but that it can gradually drift, or experience drastic changes, in going from one period to the next. This premise is also implicit in the material of Section 2.2 on the traditional indices, but what is different here is that a model which describes how the process can drift is explicitly specified. In order to develop a mechanism for controlling the process, such explicit specifications are a requirement.

Following the notation of Section 2.3, suppose that

$$X_{i+1} = \theta_{i+1} + r_{i+1} \tag{3.1}$$

where θ_{i+1} is some parameter, called the *state of nature*, and r_{i+1} is a disturbance term (a random error), assumed Gaussian, with mean 0 and variance δ^2 – assumed known; $i = 1, 2 \ldots$. Under the set-up of (3.1), θ_{i+1} can be viewed as a proxy for X_{i+1}. To describe the drift of the process, assumed to be generally smooth, suppose that

$$\theta_{i+1} = \theta_i + w_{i+1}, \tag{3.2}$$

where like r_{i+1}, w_{i+1} is also a disturbance term, assumed Gaussian with mean 0 and variance λ^2. The r_i's and the w_i's are assumed to be independent of each other and also independent among each other. For purposes of control, a new variable Y_i, called the *controller*, needs to be introduced, and associated with Y_i is a coefficient H_{i+1} which in some sense describes the leverage that the controller exerts on the state of nature, θ_i. The role of Y_i is to keep θ_{i+1}, a proxy for the observables of the future production at the target value T; this is so because $X_{i+1} = \theta_{i+1} + r_i$. Thus the equation

$$\theta_{i+1} = \theta_i + H_{i+1} Y_i + w_{i+1}, \tag{3.3}$$

describes the evolution of the state of nature from one rating period to the other, as dictated by the controller Y_i. Accordingly, controlling the process boils down to the optimal specification of Y_i in the light of the costs associated with setting Y_i and the deviation of θ_{i+1} from the target T. Observe that when controlling the process, the decision is no more to continue as is or to stop the production. Rather, the aim is to adjust the process so that production will not have to be stopped and the product quality is the desired quality. This is in contract to the decision problems of Section 2. Also, Equations (3.1) and (3.3) are prototypes. They can be generalized to suit situations more complex than the one considered here. Indeed very often, (3.3) is a differential equation whose nature is determined by the physics of the manufacturing process.

Under the set-up of Equations (3.1) and (3.3), it can be shown (see the Appendix), that given Y_i and $\underset{\sim}{x}_i(n)$, the predictive distribution of X_{i+1} (i.e.

the value of X in the interval $[T_i, T_{i+1})$) is a Gaussian with a mean μ_{i+1} and variance $\sigma_{i+1}^2 = \Sigma_i$, say, where:

$$\begin{aligned} \mu_{i+1} &= H_{i+1}Y_i + \widehat{\theta}_i, \\ \widehat{\theta}_i &= H_iY_{i-1} + \widehat{\theta}_{i-1} + R_i(R_i + \delta^2)^{-1}[\overline{x}_i(n) - (H_iY_{i-1} + \widehat{\theta}_{i-1})], \\ R_i &= \Sigma_{i-1} + \lambda^2, \quad \text{and} \\ \Sigma_i &= R_i - R_i^2(R_i + \delta^2)^{-1}. \end{aligned} \tag{3.4}$$

To evaluate the above iteration, starting values $\widehat{\theta}_0$ and Σ_0 have to be specified, and $\overline{x}_i(n) = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$. The starting values $\widehat{\theta}_0$ and Σ_0 go to determine the prior distribution.

The issue that now remains to be addressed pertains to the choice of Y_i. Recall, that the role of the controller Y_i is to ensure that θ_{i+1} is as close as is possible to the target T. This is to be achieved by trading-off the costs of control, that is, the cost of adjusting Y_i, versus the loss due to θ_{i+1}'s deviation from T.

Suppose that $\mathcal{C}_1(\theta_{i+1}, Y_i)$ is the total cost incurred when the state of nature is θ_{i+1} and the controller is set at Y_i. Suppose that $\mathcal{C}_1(\theta_{i+1}, Y_i)$ can be broken up as

$$\mathcal{C}_1(\theta_{i+1}, Y_i) = (Y_i - m_i)^2\mathcal{C}_{1i} + (\theta_{i+1} - T)^2\mathcal{C}_{2i}, \tag{3.5}$$

where m_i is that setting of Y_i which results in the smallest cost; for example, $m_i = 0$ or $m_i = Y_{i-1}$, and $\mathcal{C}_{1i}$ is the unit cost incurred when Y_i deviates from m_i by a unit amount. Similarly, $\mathcal{C}_{2i}$. Note that $(\theta_{i+1} - T)^2$ reflects the loss incurred when θ_{i+1} deviates from the target T, and this factor is analogous to considerations which lead to the index C_{pm} – see Section 2.2.1.

Using the standard arguments for normative decision making – see the Appendix for details – it can be shown that the value of Y_i which minimizes the total expected cost is given by

$$Y_i = (\mathcal{C}_{1i} + H_{i+1}^2\mathcal{C}_{2(i+1)})^{-1}(\mathcal{C}_{1i}m_i + H_{i+1}\mathcal{C}_{2(i+1)}(T - \widehat{\theta}_i)).$$

Under the above set-up, the decision d_1 of Section 2.4 becomes "set the controller to Y_i and continue production during the interval $[T_i, T_{i+1})$", and decision d_2 of Section 2.4 becomes "set the controller to Y_i, inspect every manufactured item and repair all nonconforming ones" or it becomes "overhaul the entire manufacturing system, set the controller to Y_1, and resume manufacturing". If the utilities $\mathcal{U}(d_1)$ and $\mathcal{U}(d_2)$ are of the forms specified in Section 2.4, then decision d_1 will be chosen iff

$$\Phi\left(\frac{\mathrm{USL}-\mu_{i+1}}{\sigma_{i+1}}\right)-\Phi\left(\frac{\mathrm{LSL}-\mu_{i+1}}{\sigma_{i+1}}\right)\geq\frac{b}{a}-\frac{Q^*}{aN}. \tag{3.6}$$

A final issue that remains to be addressed pertains to measuring process capability for a process that is under the control of a controller. For this, what is needed is an assessment of $P(X_i \in A|\underset{\sim}{x}_i(n), Y_{i-1})$ given that Y_{i-1} was chosen in accordance with (3.6).

Once again, it can be argued, using the prior to posterior iterative steps described in the Appendix, that given $\underset{\sim}{x}_i(n)$ and Y_{i-1}, the entity X_i (i.e. the value of X in the interval $[T_{i-1}, T_i)$ has a Gaussian distribution with mean $\widehat{\mu}_i$ and variance $\widehat{\sigma}_i^2$, where:

$$\widehat{\mu}_i=\widehat{\theta}_i, \qquad \text{and} \qquad \widehat{\sigma}_i^2=\Sigma_i+\lambda^2, \tag{3.7}$$

where $\widehat{\theta}_i$ and Σ_i were defined in (3.4).

Thus $P(X_i \in A|\underset{\sim}{x}_i(n), Y_{i-1})$ is given by

$$\Phi\left(\frac{\mathrm{USL}-\widehat{\mu}_i}{\widehat{\sigma}_i}\right)-\Phi\left(\frac{\mathrm{LSL}-\widehat{\mu}_i}{\widehat{\sigma}_i}\right), \tag{3.8}$$

and (3.8) may be used to characterize process capability under stochastic control. Equivalently, one may proceed along the lines suggested by Bernardo and Irony, and follow the logic subsequent to (2.9) to define $C_S(\nu)$ as

$$\frac{1}{\nu}\Phi^{-1}\left(P(X_i \in A|\underset{\sim}{x}_i(n), Y_{i-1})\right), \tag{3.9}$$

as the *stochastically controlled process capability index*, abbreviated as SCPCI.

4 Comparative performance of indices: an empirical analysis

To explore the comparative performance of the indices described before, an empirical analysis involving simulated, as well as real data are considered.

One set of simulated data consists of 1000 independent observations, X_i, $i = 1, \ldots, 1000$, generated from each of the following four distributions (for notation, see the Appendix):

a) $X_i \sim \mathcal{N}(1, 1)$;

b) $X_i \sim \mathcal{N}(1 + (.01)i, 1)$;

c) $X_i \sim \mathcal{N}(1, 1 + (.001)i)$;

d) $X_i \sim \mathcal{N}(1 + (.01)i, 1 + (.001)i)$.

Also considered were 1000 correlated observations generated via the autoregressive process:

e) $X_i = (0.1)X_{i-1} + \sigma_i^2$, with $\sigma_i^2 \sim \mathcal{N}(0, 1)$; σ_i^2's being mutually independent.

Each rating period was assumed to comprise of 100 observations; this would ensure that the calculated sample means and variances are of credible quality. The upper (lower) specification limit is taken to be 4.10 (-1.90), and the target value is 1.10; these choices are arbitrary.

For each rating period – and there are ten in all – the indices C_p, C_{pk}, C_{pm}, and C_{pmk} are computed, using the formulae given in Section 2, with $\mu(\sigma)$ replaced by the sample mean (standard deviation), of each rating period. The values of the indices are shown in Table 1 for the data sets generated via a), b), c), d), and e). Also shown in Table 1 is the probability that any observation within a rating period belongs to the interval [-1.90, 4.10], given the sample mean and the sample standard deviation for that period; see Equation (2.11). The sample means and the sample standard deviations of all the 1000 observations from each of the five generating mechanisms a), b), c), d), and e) turned out to be (.974, 1.004), (5.979, 3.066), (.969, 1.234), (5.974, 3.146), and (-0.029, 1.008), respectively.

An examination of the entries in Table 1 shows that for the data generated via a), none of the traditional indices are consistently greater than one. For example, the index C_p is as low as .876 (for the rating period six) and C_{pmk} as low as .805. Consequently, each of the traditional indices would call for an unwarranted stopping of the process at one rating period or the other. By contrast, the Bayes capability index would call for stopping the process at the 6-th rating period if (with the Q^* of Equation (2.8) equal to zero), $b/a > .990$, where $b = a - h$. Effectively, the Bayes capability index would stop the process if h, the cost of inspection was a miniscule (say about .01) fraction of a. Recall that $a = g + C - r$, where g is the profit for delivering a good unit, C, the penalty for delivering a defective unit, and r the cost of repairing a defective unit prior to its delivery. Clearly, in the

Data set	Index C_p									
a	0.976	1.036	1.032	1.042	1.056	0.876	0.985	1.076	0.935	0.977
b	0.952	0.995	0.954	1.008	1.023	0.880	0.952	1.050	0.898	0.995
c	0.950	0.970	0.923	0.898	0.877	0.703	0.767	0.812	0.688	0.699
d	0.928	0.937	0.863	0.876	0.860	0.708	0.752	0.804	0.673	0.713
e	0.965	1.033	1.032	1.051	1.040	0.879	0.977	1.086	0.934	0.962
Data set	Index C_{pk}									
a	0.923	0.969	0.976	1.032	1.052	0.818	0.920	1.037	0.884	0.960
b	0.843	0.561	0.209	-0.160	-0.517	-0.675	-1.050	-1.539	-1.599	-2.141
c	0.897	0.904	0.871	0.892	0.867	0.649	0.710	0.781	0.646	0.691
d	0.823	0.530	0.191	-0.142	-0.441	-0.538	-0.822	-1.177	-1.193	-1.539
e	0.588	0.617	0.630	0.693	0.702	0.525	0.583	0.683	0.570	0.627
Data set	Index C_{pm}									
a	0.964	1.016	1.018	1.042	1.056	0.863	0.967	1.069	0.924	0.976
b	0.905	0.606	0.390	0.277	0.216	0.184	0.156	0.134	0.119	0.105
c	0.938	0.952	0.912	0.897	0.877	0.694	0.756	0.809	0.683	0.698
d	0.885	0.594	0.384	0.273	0.213	0.183	0.156	0.133	0.118	0.104
e	0.639	0.646	0.658	0.716	0.730	0.603	0.631	0.693	0.631	0.678
Data set	Index C_{pmk}									
a	0.911	0.949	0.962	1.032	1.051	0.805	0.904	1.029	0.874	0.959
b	0.802	0.342	0.086	-0.044	-0.109	-0.142	-0.172	-0.196	-0.212	-0.226
c	0.887	0.887	0.860	0.892	0.867	0.641	0.699	0.778	0.641	0.690
d	0.785	0.336	0.085	-0.044	-0.109	-0.139	-0.170	-0.195	-0.210	-0.225
e	0.389	0.385	0.402	0.472	0.493	0.360	0.376	0.436	0.385	0.442
Data set	$P(\text{LSL} \leq X \leq \text{USL} \mid \underline{x}_i(1000))$									
a	0.996	0.998	0.998	0.998	0.998	0.990	0.996	0.999	0.994	0.997
b	0.994	0.954	0.735	0.316	0.060	0.021	0.001	0.000	0.000	0.000
c	0.995	0.996	0.994	0.993	0.991	0.963	0.977	0.985	0.960	0.964
d	0.992	0.944	0.717	0.335	0.093	0.053	0.007	0.000	0.000	0.000
e	0.961	0.968	0.971	0.981	0.982	0.942	0.960	0.980	0.956	0.970

Table 1: Comparative performance of indices based on simulated data.

case of the data set generated by a), the behavior of the Bayes capability index is more in keeping with reasonable action as compared to the traditional indices. But can this claim be extended to the case of the data sets generated by the other mechanisms?

For data generated via b) and d), both of which incorporate an upward trend, the traditional indices do indeed reflect reasonable behavior (save an occasional lapse by C_p), with the index C_{pmk} having a distinct upper hand – it is unforgiving from the very start! For data generated by b), the Bayes capability index will stop the process at the 1-st rating period only if $h < .006a$, and will allow the process to continue production at the 5-th rating period if the cost of inspection $h > .94a$ – an unrealistic situation. The basic import of the above is that index C_{pmk} appears to be much more responsive to an upward trend in the mean than the Bayes capability index which tends to be lethargic. However, it should be borne in mind that, as computed, the Bayes capability index has not incorporated, via its prior distribution, an upward trend in the mean, and has assumed that the process is stable within a rating period. In actuality trends in the process means are generally hard to foretell; thus the need for modifying the Bayes capability index in the direction of a stochastically controlled version is germane. Finally, a comparison of the performance of the Bayes capability index with regards to the data generated by b) and by d) suggests that for the rating periods four and above, the index is more supportive of the decision to continue production in the case of d) than in the case of b). Such partial behavior of the Bayes capability index in the presence of the heteroschedasticity in d) is contrary to expectation. One solution would be to enhance the Bayes capability index (and for that matter also the stochastically controlled process capability index) by its *robustification*; see for example Meinhold and Singpurwalla (1989), and the references therein. A similar thought arises when one compares the behavior of the Bayes capability index for the data sets generated by a) and by c). The weak sensitivity of the Bayes capability index to the heteroschedasticity in c) could be made pronounced by an appropriate robustification of the index.

An examination of the entries in Table 1 also shows that for the data set generated by c), the index C_{pmk} appears to have an edge over the other traditional indices. But this feature may also raise a concern as to whether the index C_{pmk} is overly sensitive, and that its cautious behavior could be economically unwarranted. This matter can be explored by comparing the behavior of the index C_{pmk} and the Bayes capability index, vis-a-vis the data set generated by e).

Recall that e) describes an autoregressive process of order one, whose mean value is zero, and whose variance is $1/(1-(.1)^2) \approx 1$; see p. 58, of Box and Jenkins (1976). Furthermore, $\rho(k)$, the autocorrelation function of this process at lag k, is of the form $\rho(k) = (.1)^k$. Thus for all intents and purposes, the X_i's generated via e) may be regarded as being the realizations of independent and identically normally distributed random variables with mean 0 and variance ≈ 1. For these data, the index C_{pmk} dictates that the process be stopped at all the ten rating periods, whereas the Bayes capability index allows the process to continue production at all the rating periods (unless of course $h < .058a$, in which case the process will be stopped at the 6-th rating period). The indices C_{pk} and C_{pm} would lead to decisions similar to those of C_{pmk}, but the decisions based on the index C_p would be closer in tune to those based on the Bayes capability index. Why this disparity of decisions based on the index C_{pmk} and the Bayes capability index? The answer lies in the fact that the index C_{pmk} inflicts penalties whenever the sample mean of a rating period deviates from the target T, and also from the midpoint M, whereas the Bayes capability index penalizes whenever a unit fails to belong to its specification limits. In the case of data generated via e), the sample means are in the vicinity of 0, whereas the target value is 1.1 . Thus, the sensitivity of C_{pmk} can only be judged in the light of the realism of its associated penalties. To obtain compatibility between C_{pmk} and the Bayes capability index, the $\mathcal{U}(d_1)$ and $\mathcal{U}(d_2)$ of Section 2.4 should be modified so that there are costs associated with deviations of X from T and M.

4.1 Comparative performance against real data

In Table 2 , some data on tool wear abstracted from Grant and Leavenworth (1974) is presented. The data gives the sample means and the ranges of 13 groups of observations, each group consisting of five observations; the ranges are viewed as a proxy for the sample standard deviations. The upper and lower specification limits for these data were 0.6480 and .6400 respectively, and the target value T is 0.6440 [cf. Spiring (1991)].

Also shown in Table 2 are the values of C_{pm}, C_{pmk}, and the probability that an observation within a rating period (in this case a group) belongs to the interval [.6400, .6480], given the sample statistics for that interval. The values of the indices C_p and C_{pk} for this data set turn out to be substantially greater than one, for all the rating periods; they are not shown in Table 2 .

An examination of the entries in Table 2 shows that whereas the indices C_p and C_{pk} would allow the process to continue as is, throughout the life-cycle of the tools, the indices C_{pm} and C_{pmk} will act differently, except say at rating periods six through ten. According to Spiring (1991) the large fluctuations in the indices C_{pm} and C_{pmk} are the manifestations of

Group Number	Sample Mean	Sample Range	C_{pm}	C_{pmk}	$P(\text{LSL} \leq X \leq \text{USL} \mid \underset{\sim}{x}(5))$
1	.6417	.0011	.511	.217	.939
2	.6418	.0016	.524	.236	.870
3	.6424	.0010	.727	.436	.992
4	.6431	.0015	1.136	.881	.980
5	.6433	.0009	1.546	1.276	1.000
6	.6437	.0010	2.546	2.355	1.000
7	.6433	.0014	1.383	1.141	.990
8	.6436	.0004	2.805	2.525	1.000
9	.6441	.0006	5.013	4.887	1.000
10	.6444	.0011	2.119	1.907	.999
11	.6456	.0009	.731	.438	.996
12	.6457	.0007	.694	.399	.999
13	.6454	.0009	.830	.539	.998

Table 2: Comparative performance of indices for tool wear data.

not only the process capability but also the presence of other assignable causes of tool wear. That is, the process capability indices have also served as aids for identifying other causes. In contrast to the behavior of C_{pm} and C_{pmk}, the Bayes capability index leads to actions that mimic those dictated by the indices C_p and C_{pk}. Here again, the cause for this disparity is penalty for deviation from target which the indices C_{pm} and C_{pmk} levy; the Bayes capability index penalizes only when there is a failure to belong to a specification interval. Here again, compatibility between the Bayes capability index and the indices C_{pm} and C_{pmk} can be achieved if the utility functions of Section 2.4 can be justifiably modified.

Like the fluctuations of C_{pm} and C_{pmk}, the fluctuations of $P(LSL \leq X \leq USL \mid \underset{\sim}{x}(5))$ may be used to detect assignable causes. However, the entries of Table 2 suggest that the fluctuations of the latter appear to be less pronounced than those of the former, so that $P(LSL \leq X \leq USL \mid \underset{\sim}{x}(5))$ may not be a very revealing indicator of assignable causes.

4.2 Performance of indices for a process under stochastic control

To illustrate the comparative behavior of the indices for process with and without a controller, 50 observations were generated via the mechanisms given by Equations (3.1) and (3.2), and the Equations (3.1) and (3.3). The starting value θ_0 was described as $\theta_0 \sim \mathcal{N}(5,1)$, and the parameters δ^2 and λ^2 were taken to be 2 and 1, respectively. The coefficient H_{i+1} of the controller Y_i – see Equation (3.3) – was taken to be 2, and T was set at 5. To incorporate the scenario of an unconstrained or "best case" control, the coefficients C_{1i} and C_{2i} of Equation (3.5) were set at zero. The number of rating periods was chosen to be 10 so that each rating period has five observations. The upper (lower) specification limit was taken to be 6.73 (3.27).

Figure 4 shows plots of the data generated via Equations (3.1) and (3.2), and via Equations (3.1) and (3.3). The effect of the controller Y_i in centering the data around its target $T = 5$ is apparent. Tables 3 and 4 show the comparative behavior of each of the indices for the data generated as described, with and without the controller.

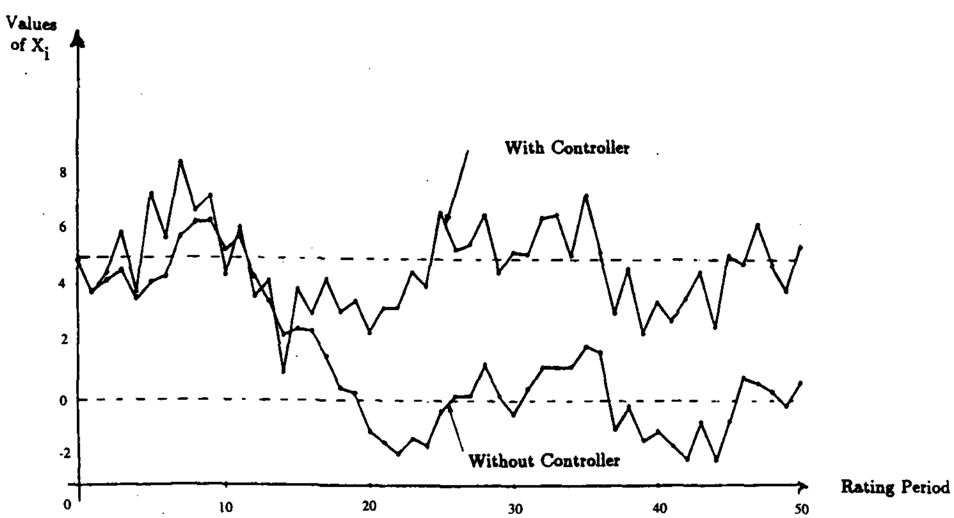

Figure 4: Plots of date generated with and without a controller.

	C_p		C_{pk}		C_{pm}	
Rating Period i	Without Control	With Control	Without Control	With Control	Without Control	With Control
1	.544	.475	.151	.219	.203	.217
2	.648	.610	.576	.316	.365	.264
3	.530	.418	.409	.355	.286	.237
4	.976	.992	-.522	.416	.122	.286
5	.791	.941	-.953	.346	.086	.265
6	.878	.611	-.516	.453	.118	.318
7	.684	.641	-.310	.584	.125	.364
8	.749	.593	-.354	.530	.125	.336
9	.646	.696	-.490	.533	.105	.361
10	.782	.576	-.334	.470	.129	.317

Table 3: Comparative performance of indices for processes with and without a controller. First part

An examination of the entries in Tables 3 and 4 show that after an initial period of adjustment, the controller comes into action and tends to increase the probabilities of coverage (see the last two columns of Table 4), and also the values of the all the indices, save the index C_p. The reason behind C_p's lack of responsiveness to the controller is that there is no provision in C_p to reflect target values which is what the controller works towards.

To summarize, the results of the empirical investigations suggest that the Bayes capability index and its stochastically controlled version are meaningful devices that perform in consort with the other traditional indices. However, they possess the advantage of having an interpretive value. These normative indices can stand further improvements via a consideration of alternate utility functions and their robustification.

Acknowledgements

I am grateful to Dr. Michael McGrath for introducing me to this topic, and to Professors José Miguel Bernardo and Telba Irony whose paper inspired my interest in it. Comments by Professor Bernardo also helped sharpen the material; Section 4 is due to his urging. Professor Samuel Kotz made comments on an earlier version of this paper and was kind enough to let me have a pre-publication copy of his manuscript. The material in Section 4.2 is based on unpublished work with Dr. Richard Meinhold. Finally,

	C_{pmk}		$P(\text{LSL} \leq X \leq \text{USL}\mid \underset{\sim}{x}_i(5))$	
Rating Period i	Without Control	With Control	Without Control	With Control
1	-.051	.015	.389	.466
2	.295	.042	.726	.542
3	.176	.175	.607	.523
4	-.203	-.002	.002	.495
5	-.242	-.026	.000	.437
6	-.206	.175	.004	.657
7	-.191	.308	.036	.725
8	-.194	.274	.022	.687
9	-.215	.215	.011	.718
10	-.190	.216	.023	.657

Table 4: Comparative performance of indices for processes with and without a controller. Second part.

I thank Yefim Vladimirsky for computational assistance, and Ms. Yuling Cui for assistance in putting the material together.

Appendix

The development here starts with the (a priori) assumption that given $\underset{\sim}{x}_i(n)$, the unknown parameter θ_i has a Gaussian distribution with mean $\widehat{\theta}_i$ and variance Σ_i, for $i = 0, 1, 2, \ldots$. The starting values $\widehat{\theta}_0$ and Σ_0 have to be specified by the user. The above assumption will be denoted by the notation "$(\theta_i \mid \underset{\sim}{x}_i(n)) \sim \mathcal{N}(\widehat{\theta}_i, \Sigma_i)$"; this notation will also be used if θ_i is a vector quantity, in which case $\widehat{\theta}_i$ will also be a vector and Σ_i a matrix.

Since the r_i's and the w_i's of Equations (3.1) and (3.3) are serially and contemporaneously independent, and since by construction θ_{j-1} is independent of w_j, $j = 1, 2, \ldots$, it follows that given $\underset{\sim}{x}_i(n)$, the vector displayed below has the distributional form:

$$\left(\begin{array}{c|c} \theta_{i-1} & \\ w_i & \underset{\sim}{x}_{i-1}(n) \\ r_i & \end{array} \right) \sim \mathcal{N}\left(\begin{pmatrix} \widehat{\theta}_{i-1} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{i-1} & 0 & 0 \\ 0 & \lambda^2 & 0 \\ 0 & 0 & \delta^2 \end{pmatrix} \right). \quad \text{(A.1)}$$

Furthermore, (3.1), (3.2), and their underlying assumptions imply that given Y_{i-1}, the vector below may be written as:

$$\left(\begin{array}{c|c} \theta_i & \\ & Y_{i-1} \\ X_i & \end{array} \right) = \begin{pmatrix} H_i Y_{i-1} \\ \\ H_i Y i - 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 0 \\ \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_{i-1} \\ w_i \\ r_i \end{pmatrix}. \quad \text{(A.2)}$$

Consequently, it follows from (A.1) that:

$$\left(\begin{array}{c|c} \theta_i & \\ & Y_{i-1}, \underset{\sim}{x}_{i-1}(n) \\ X_i & \end{array} \right) \sim \mathcal{N}\left(\begin{pmatrix} H_i Y_{i-1} + \widehat{\theta}_{i-1} \\ \\ H_i Y_{i-1} + \widehat{\theta}_{i-1} \end{pmatrix}, \begin{pmatrix} R_i & R_i \\ \\ R_i & R_i + \delta^2 \end{pmatrix} \right), \quad \text{(A.3)}$$

where R_i has been defined in (3.4). The result preceding (3.4) and (3.7) now follows from the conditional distribution properties of the multivariate state of nature θ_{i+1}, and the historical data until and including the time interval $[T_{i-1}, T_i)$; this data is the collection of vectors $\underset{\sim}{x}_1(n), \dots, \underset{\sim}{x}_i(n)$.

At node $\mathcal{D}_2$ an estimator of θ_{i+1}, say $\widehat{\theta}_{i+1}$, will be selected – based on the predictive distribution of $\underset{\sim}{X}_{i+1}(N_{i+1})$, and the subsequent to this nature will yield θ_{i+1}. The terminal utility is of the form given by (3.5) and the aim is to choose Y_i and $\widehat{\theta}_{i+1}$ that minimize (3.5).

According to standard results of normative decision making [cf. Bernardo and Smith (1994)], $\widehat{\theta}_{i+1}$ is the expected value of the posterior distribution of θ_{i+1}, were one to know $\underset{\sim}{X}_{i+1}(N_{i+1})$. Averaging over the predictive distribution of $\underset{\sim}{X}_{i+1}(N_{i+1})$, one has, at node $\mathcal{R}_1$, the expected utility in selecting $\widehat{\theta}_{i+1}$, were the controller to be set to Y_i. At $\mathcal{D}_1$, that value of Y_i which maximizes normal distribution (cf. Meinhold and Singpurwalla (1983)), and the relationships (3.1)–(3.3).

For the optimal choice of Y_i, the decision tree given in Figure A1 is helpful. In the mentioned decision tree, at decision node $\mathcal{D}_1$, the controller is set to its optimal value Y_1. This results in the potentially observable sequence $\underset{\sim}{X}_{i+1}(N_{i+1}) = (X_{i+1,1}, \cdots, X_{i+1,N_{i+1}})$. At $\mathcal{D}_1$, one has knowledge of the target T for the above expected utility is chosen. The prescribed scheme boils down to maximizing the quantity

$$C_{1i}(Y_i - m_i)^2 + (H_{i+1}Y_i + \widehat{\theta}_i - T)^2 C_{2i}$$

yielding the result given in Section 3.

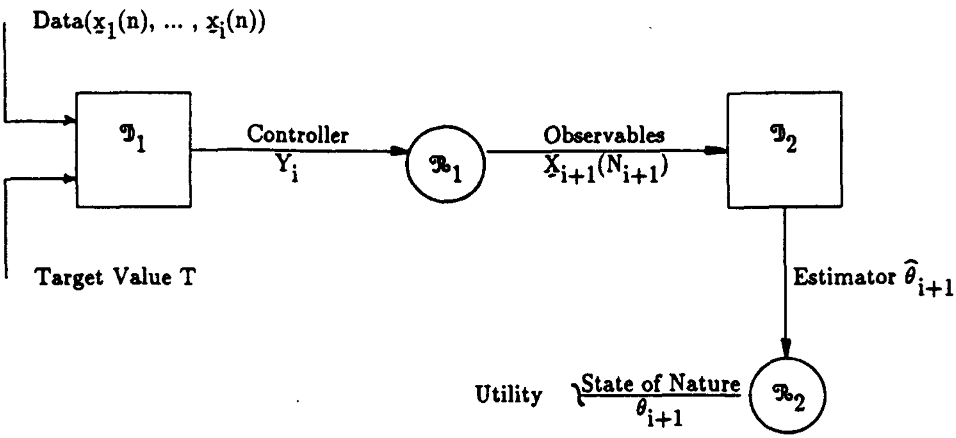

Figure A1: Schemata for optimal control of process indices.

References

Bernardo, J.M. and T.Z. Irony (1996). A general multivariate Bayesian process capability index. *The Statistician*, **45**, 3, 487–502.

Bernardo, J. M. and A. F. M. Smith (1994). *Bayesian theory.* John Wiley, Chichester.

Billingsley, P. (1986). *Probability and measure.* John Wiley and Sons, Inc.

Bissell, A. F. (1990). How reliable is your capability index? *Applied Statistics*, **39**, 331–340.

Box, G.E.P. and G. M. Jenkins (1976). *Time series analysis, forecasting and control.* Holden-Day, San Francisco.

Chan, L. K., S. W. Cheng and F. A. Spiring (1988a). A new measure of process capability: C_{pm}. *Journal of Quality Technology*, **20**, 162–175.

Chen, H. (1994). A multivariate process capability index over a rectangular solid tolerance zone. *Statistica Sinica*, **4**, 749–758.

Cheng, S. W. and F. A. Spiring (1989). Assessing process capability: a Bayesian approach. *IIE Transactions*, **21**, 97–98.

Fries, A. and K. J. Richter (1993) (Appendix E). Process capability and performance indices. Defense Science Board Task Force Report, Engineering in the Manufacturing Process.

Grant, E. L. and R. S. Leavenworth (1974). *Statistical quality control*, 5th Edition. McGraw Hill, New York, NY.

Gupta, A. K. and S. Kotz (1997). A new process capability index. *Metrika*, **45**, 213–224.

Harry, M. J. and J. R. Lawson (1992). *Six sigma producibility analysis and process characterization.* Addison-Wesley Publishing Co. Inc.

Hsiang, T. C. and G. Taguchi (1985). A tutorial on quality control and assurance-The Taguchi method. *ASA Annual Meeting*, Las Vegas.

Juran, J. M. (1974). *Jurans quality control handbook*, 3rd edition. McGraw-Hill, New York.

Juran, J. M. and F. Gryna (1980). *Quality planning and analysis.* McGraw-Hill, New York.

Kane, V. E. (1986). Process capability indices. *Journal of Quality Technology*, **18**, 41–52.

Kotz, S. and C. R. Lovelace (1997). *Process capability indices in theory and practice.* Arnold, London. To be published.

Kushler, R. H. and P. Hurley (1992). Confidence bounds for capability indices. *Journal of Quality Technology*, **24**, 4, 188–195.

Meinhold, R. J. and N. D. Singpurwalla (1983). Understanding the Kalman filter. *The American Statistician*, **37**, 2, 123–127.

Meinhold, R. J. and N. D. Singpurwalla (1989). Robustification of Kalman filter models. *Journal of the American Statistical Association*, **84**, 406, 479–488

Pearn, W. L., S. Kotz and N. L. Johnson (1992). Distributional and inferential properties of process capability indices. *Journal of Quality Technology*, **24**, 216–231. (Abbreviated as PKJ).

Pearn, W. L. and K. S. Chen (1996). A Bayesian-like estimator of C_{pk}. *Communications in Statistics.*

Singpurwalla, N.D. (1992). A Bayesian perspective on Taguchi's approach to quality engineering and tolerance design (with discussion). *IIE Transactions*, **24**, 5, 18-32.

Spiring, F. A. (1991). Assessing process capability in the presence of systematic assignable cause. *Journal of Quality Technology*, **23**, 2, 125–134.

Taam. W., P. Subbaiah and J. W. Liddy (1993). A note on multivariate capability indices. *Journal of Applied Statistics*, **20**, 339–351.

Vännman, K. (1995). A Unified approach to capability indices. *Statistica Sinica*, **5**, 2, 805–822.

Vännman, K. (1996a). Capability indices when tolerances are asymmetric. Paper to be presented at the Cochin-ISI/QISM Conference, Cochin, India.

Vännman, K. (1996b). Families of capability indices for one-sided specification limits. Paper to be presented at the Cochin-ISI/QISM Conference, Cochin, India.

Wierda, S. J. (1993). A multivariate process capability index. *Transactions, ASQC Quality Congress*, 342–348.

Wright, P. A. (1995). A process capability index sensitive to skewness. *Journal of Statistical Computation and Simulation*, **52**, 195–203.

DISCUSSION

G. Box
University of Wisconsin, USA

I agree with the author that the value of capability indices is highly dependent on long term stability of the process and I would add that, the second law of thermodynamics ensures that no process is ever in such a state of control. This does not mean, of course, that the ideas of Shewhart and Deming concerning process monitoring based on the NIID stationarity approximation are useless. This approximation has over the years resulted in the detection and elimination of thousands of problems and so produced improvement of a host industrial processes, demonstrating once more that all models are wrong, but some models are useful.

However, while for the detection of assignable causes we need only to approximate fairly short time behavior, capability indices are concerned with long term process behavior, for which the stationarity approximation is likely to be inadequate. As the author implies, stationarity of the process output can then only be induced by appropriate process adjustment. Such adjustment is possible by feedback control or by feedforward control, or both. If, by such control, the output from the process is rendered stationary and so varies about a fixed mean then standard capability indices can be used and the problem posed by the author will be solved. Furthermore, if given the particular production circumstances, smallest process variation about the target can be achieved then maximum possible process capability will be obtained.

To devise suitable feedback control procedures, we must first abandon the IID approximation and I agree with the author that to then represent non- stationarity process disturbances, that would occur if no control were instituted, the "noisy random walk" he postulates in his equation 3.1 et seq. is known to provide a very valuable model. Some of the considerations that point to its central importance are as follows.

1. Consider the standardized variogram $V_k = \mathrm{Var}(X_{t+k} - X_t)/\mathrm{Var}(X_{t+1} - X_t)$. For the IID model, $V_k = 1$ for all k (implying, for example, that observations made a hundred hours apart will be no more discrepant than observations taken one hour apart). For the noisy random walk however, V_k increases linearly with k - a result that makes sense (see for example Box and Kramer (1992)). The rate at which V_k increases depends on an important parameter which I will here call π. In the author's notation it satisfies the equation $\lambda^2/\delta^2 = (1-\pi)^2/\pi$.

2. For the noisy random walk the value X_{t+1} is such that

$$X_{t+1} = \widetilde{X}_t + a_{t+1} \tag{1}$$

where $\widetilde{X}_t$ is an exponentially weighted moving average (EWMA) with smoothing constant π such that

$$X_t = (1-\pi)(X_t + \pi X_{t-1} + \pi^2 X_{t-2} + \dots), \quad 0 \le \pi \le 1, \tag{2}$$

and $\{a_t\}$ is a sequence of IID random variable (which we will call white noise) having mean zero and $\sigma_a = \delta/\sqrt{\pi}$ (see e.g. Box and Luceño (1997)). Thus given data $\{X_t\}$ up to time t the function that provides a minimum mean square error (MMSE) forecast/estimate of X_{t+1} is the EWMA $\widehat{X}_t$.

2. It is easy to see that the model (1) may be written as a stochastic difference equation in either of the two forms

$$\text{(a) } X_{t+1} - 2X_t = a_{t+1} - \pi a_t \quad \text{or} \quad \text{(b) } \widehat{X}_t - \pi \widehat{X}_{t-1} = (1-\pi)X_t. \qquad (3)$$

This model is called an integrated moving average (IMA) a special case of the "ARIMA" models of Box and Jenkins. Model building using (reconstructed) industrial process disturbances data has led, with surprising frequency, to the choice of this IMA model.

At this point however, I part company with the author, because to achieve efficient feedback in a quality control context there are a number of further considerations that must be taken into account.

Process Dynamics: There may be inertia and possibly pure delay in the control system so that the effect of a change in the input may not be fully experienced at the output in one time interval. Frequently the relation between the input to the control system Y_t and its effect $\widehat{Y}_t$ at the output can be approximated by a first order difference equation of the same form as 3(b) and such that

$$\widehat{Y}_t - \varphi \widehat{Y}_{t-1} = (1-\varphi) Y_{t-1-f}. \qquad (4)$$

For a so-called "responsive system" where an adjustment is fully effective at the output in one time interval and $\varphi = 0$ and $f = 0$.

Minimum Mean Square Error Control: In the unusual circumstance when there is no limitation on the extent or the frequency of the compensating manipulation using (3) and (4) assuming a responsive system, MMSE control may be obtained by continually making an adjustment of the simple form

$$Y_t - Y_{t-1} = y_t = c_1 e_t + c_2 e_{t-1}, \qquad (5)$$

where et is the error (deviation from target data) at time t and the constants where c_1 and c_2 are simple functions of π and φ (given, for example, by Box and Jenkins (1968)).

Equation (5) then corresponds to a discrete form of the control engineer's PI proportional plus integral) control since then

$$Y_t = k_1 e_t + k_2 \sum_{i=1}^{t} e_i, \qquad (6)$$

where $k_1 = -c_2$ and $k_2 = c_1 + c_2$. Furthermore, for such control $e_t = a_t$ so the controlled process is stationary white noise. It can be further shown that such control is equivalent to continually arranging that an EWMA of past adjustments $\widehat{Y}_t$ with smoothing constant (just cancels an EWMA of past disturbances X_t with smoothing constant π (Box and Luceño (1997)).

Constrained Linear Control: Unfortunately this MMSE of control may not be much use because frequently it requires excessively large adjustments $\{Y_t\}$ at input. Constrained linear adjustment schemes which produce an uncontrolled minimum of

$$\sigma_y^2 + \alpha\sigma_e^2 \tag{7}$$

have therefore been studied by, among others, Whittle (1963), Box and Jenkins (1968), MacGregor (1972), and Astrm and Wittenmark (1984). These schemes can give remarkably large reductions in σ_y^2 for very small increases in σ_e^2, but they tend to be complicated and not easily derived or put into practice.

Constrained PI Control: It has recently been shown (Box and Luceño (1995, 1997)) that simple PI control in which σ_y^2 is constrained by minimization of (7) can produce control schemes which are very nearly as good as optimal linear control. Furthermore, if desired, such schemes may be put into practice manually using a chart no more difficult to apply than a Shewhart chart (see e.g. Box and Jenkins (1970)).

Minimizing Frequency of Adjustment and of Sampling: The above schemes allow for adjustment each time the process is observed. Such frequent adjustment is often inconvenient and costly. Assuming a responsive system with a disturbance given by 3(a) it can be shown that for a given increase in the output standard deviation (ISD) the average interval between adjustments (AAI) is minimized by using an EWMA chart with the position of the limit lines determined not by probability considerations, but by the desired values of the AAI and ISD. Equivalently, this system is optimal for a fixed cost of adjustment and a quadratic off-target loss function (Box and Jenkins (1963)). More recently (Box and Kramer (1992), Box and Luceño (1994)) simultaneous minimizing of frequency of the adjustment has been included.

In summary then, I agree with the author that to obtain a meaningful

CP index requires allowance for nonstationarity of the disturbance. This may best be done, however, by first using an appropriate system of feedback control which taking account of circumstances of the problem produces smallest (stationary) variation about the target value. Standard CP indices may then be applied and maximum process capability obtained. Although such feedback systems for quality control have been under development for more than thirty years, it is disturbing to find there is no overlap between the author's references and those contained in this discussion. Finally, development of improved quality control can take place only if quality control practitioners themselves can understand and use the methods presented. With this in mind the material described above has been brought together at a suitably accessible level in a recent book (Box and Luceño (1997)).

References

Aström, K.J, and B. Wittenmark (1984). *Computer Controlled Systems: Theory and Design.* Prentice Hall, Englewood Cliffs, NJ.

Box, G.E.P. and G.M. Jenkins (1968). Discrete Models for Feedback and Feed-Forward Control. *The Future of Statistics* (D.G. Watts ed.). Academic Press, New York, pp. 201-240.

Box, G.E.P. and G.M. Jenkins (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

Box, G.E.P., G.M. Jenkins and G.C. Reinsel(1994). *Time Series Analysis, Forecasting and Control,* 3rd ed. Prentice Hall, Englewood Cliffs, NJ.

Box, G.E.P. and T. Kramer (1992). Statistical process monitoring and feedback adjustment. A discussion. *Technometrics,* **34**, 251-285.

Box, G.E.P. and A. Luceño (1994). Selection of sampling interval and action limits for discrete feedback adjustment. *Technometrics,* **36**, 369-378.

Box, G.E.P. and A. Luceño (1995). Discrete proportional-integral control with constrained adjustment. *Journal of the Royal Statistical Society, Series D - The Statistician,* **44**, 479-495.

Box, G.E.P. and A. Luceño (1997). The anatomy and robustness of discrete proportional-integral adjustment and its application to statistical process control. *Journal of Quality Technology,* **29**, 248-260.

Box, G.E.P. and A. Luceño (1997). *Statistical Control by Monitoring and Feedback Adjustment.* J. Wiley, NY.

MacGregor, J.F. (1972). *Topics in the Control of Linear Processes Subject to Stochastic Disturbances.* Unpublished Ph.D. thesis, University of Wisconsin-Madison.

Whittle, P. (1963). *Prediction and Regulation by Linear Least-Squares Methods.* English University Press, London.

D.R. Cox
University of Oxford, United Kingdom

The need to reduce dimensionality arises in many fields and by far the most widely used method of achieving such is via the construction of one or more indices on more or less intuitive grounds; formal methods of dimension reduction do of course exist but are not that commonly employed. Note, however, the extensive economic literature on the construction and properties of index numbers.

In specific applications a more explicit formulation of objectives will nearly always be better in some sense and Professor Singpurwalla's very interesting analysis of a quality control situation exemplifies this.

One important aspect of standardized indices is that they facilitate comparison of different situations. A biomedical application is the use of body mass index, weight divided by height squared; while a different combination of weight and height may well lead to a more sensitive analysis in the context of any one specific application, the standard index has some value when used across different studies.

Professor Singpurwalla emphasizes the normative aspects of the formal decision analysis and I have no wish to dissent from this. At the same time the normative analysis is only as good as the numbers put into the analysis for specific features. In such discussions of industrial quality there may well be a temptation to undervalue such aspects as long-term goodwill of customers and thus to underrate long term high quality; for a brief historical discussion, see Cox (1990).

I was glad to see that Professor Singpurwalla does not assume a process in statistical control. There is an interesting historical and philosophical issue connected with this. The pioneers of statistical quality control, especially W.A. Shewhart, emphasized the importance of control because it implied absence of assignable causes, i.e. total randomness was identified with a process essentially incapable of improvement. Modern thinking has, for a number of different reasons moved somewhat away from that.

References

Cox, D.R. (1990). Quality and reliability: some recent developments and a historical perspective. *Journal of the Operational Research Society*, **41**, 95-101.

D. K. Dey
University of Connecticut, USA

1 Introduction

It is a pleasure to congratulate the author on an interesting and important paper that points out the difficulty of the use of existing process capability indices, which are typically in industry. Viewing as an engineering decision-making problem, the author attacks the core of the problem such as whether to intervene in a production process or not, to accept or reject a production batch or whether to review the managerial action need or not. The author demonstrates that some of the standard process capability indices can evolve in special cases under artificial assumptions from a very general methodology. Using the normative approach to process capability, the author develops a unified method with respect to both, prediction and control. Finally, the author relaxes the assumption of process stability and develops a nice dynamic model which describes how the process can drift.

This is a very nice expository paper and I hope it will have significant impact on the use of process capability indices in practice. I will comment on three issues which are not considered in this paper. In Section 2, I will discuss certain small sample features, in the sense that, often it is unreasonable to assume that the predictive distribution is normal. I will consider the situation when the predictive distribution is a scale mixtures of normals which includes in particular a Student t-distribution. Section 3 is devoted to modelling situations when there is an evidence that the process distribution is not symmetric. I develop a new skew-normal distribution to model situation. Finally in Section 4, I mentioned about multivariate extension.

2 Predictive distribution as scale mixtures of normals

Suppose $x_1, \ldots, x_n$ are the actual measurements which is assumed to be normal with mean μ and variance σ^2, then it is well known that (see, for example, Geisser (1993)) under the standard "noninformative" prior $\pi(\mu, \sigma) \propto \sigma^{-1}$, the predictive distribution is Student t with $n-1$ degrees of freedom. Thus, for small sample size n, it is unreasonable to assume normality of the predictive distribution. In fact, often for small n, it is reasonable to assume that the predictive distribution belongs to a large class of scale mixtures of normals, which is defined as follows. Suppose we assume the i^{th} measurement

$$x_i \sim N(\mu, k(\lambda)\sigma^2) \tag{2.1}$$

and

$$\lambda \sim \pi(\lambda) \tag{2.2}$$

where $k(\lambda)$ is a positive function of a one-dimensional positive-valued scale mixing variable λ and $\pi(\lambda)$ is a mixing distribution which is either discrete or continuous. The class of scale mixtures of normals is quite rich and includes Student t, logistic, symmetric stable and exponential power family distributions. Taking $k(\lambda) = 1$ and the mixing distribution $\pi(\{1\}) = 1$, the scale mixture reduces to the usual normal distribution. Student t-distribution with ν degrees of freedom is generated by taking $k(\lambda) = 1/\lambda$ and $\lambda \sim \mathcal{G}(\nu/2, \nu/2)$, i.e., $\pi(\lambda) \propto \lambda^{\nu/2-1} exp\{-\frac{\nu}{2}\lambda\}$. A logistic distribution is obtained by taking $k(\lambda) = 4\lambda^2$ where λ follows an asymptotic Kolmogorov distribution with density $\pi(\lambda) = 8\sum_{k=1}^{\infty}(-1)^{k+1}k^2\lambda \exp\{-2k^2\lambda^2\}$.

It follows immediately, that using (2.1) and (2.2), one can easily simulate the predictive distribution, for example, using Theorem 1 of Bernardo

& Irony (1996). In general, a Monte Carlo estimate of

$$P(x \in A|D) = \int_A (2\pi k(\lambda))^{1/2}\sigma^{-1/2}\exp\left\{-\frac{k^{-1}(\lambda)}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}\pi(\lambda)d\lambda dx_i, \tag{2.3}$$

where $\mathcal{D} = \{x_i(n)\}$ is the current data, can be obtained using sampling based approach e.g., Gibbs sampler (Gelfand & Smith (1990)). Therefore, Bayesian capability index can be easily computed.

3 Skewed predictive distribution

Often in a production process, there is a presence of trend, i.e., the process mean can shift upwards or downwards. Assuming a symmetric distribution in such scenario is not justified. Even when sample size is large, the predictive distribution may not be symmetric. To model such scenario, suppose $x_1, \ldots, x_n$ are the actual values, i.e., a random sample from some fixed underlying distribution and the observed measurements $y_1, \ldots, y_n$ are related to $x_1, \ldots, x_n$ as

$$y_i = x_i + \delta z_i \tag{3.1}$$

where z_i is (the measurement error random effect) and $\delta(-\infty < \delta < \infty)$ is an unknown parameter, indicating skewness. Under the assumption that $x_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$ and $z_i's$ are exchangeable with a pdf $g(z)$, equation (3.1) defines a general class of skewed normal distribution. In particular, when $z_i's$ are exchangeable with half normal pdf, i.e.,

$$g(z) = \frac{2}{\sqrt{2\pi}}e^{-x^2/2}, \ z > 0 \tag{3.2}$$

then it can be show that the pdf of y_i is

$$f(y_i) = \frac{2}{\sqrt{\sigma^2+\delta^2}}\varphi\left(\frac{y_i-\mu}{\sqrt{\sigma^2+\delta^2}}\right)\Phi\left(\frac{\delta(y_i-\mu)}{\sigma\sqrt{\sigma^2+\delta^2}}\right), \tag{3.3}$$

which is called a skew-normal distribution with skewness parameter δ. Here φ and Φ denote the standard normal pdf and cdf respectively. Clearly when $\delta = 0$, there is no measurement error. A simpler version of (3.3) is given in in Azzalini & Dalla Valle (1996). Now from (3.3), it can be shown that the mean of Y_i is

$$E(Y_i) = E[E(Y_i|z)] = \mu + \delta\sqrt{\frac{2}{\pi}} \tag{3.4}$$

where $\delta' = \delta\sqrt{\frac{2}{\pi}}$ can be treated as the parameter which detects the unknown shift of the process mean. Now under a hierarchical set up, putting prior on μ, σ^2 and δ, one can fit a hierarchical Bayes model and obtain the posterior predictive distribution and a Monte Carlo estimate of $P(x \in A|D)$ using sampling based approach. The details are omitted here and will be reported elsewhere. The basic idea is to develop a Metropolis-Hastings algorithm along a Gibbs sampler.

4 Multivariate process capability

The use of process capability indices in connection with multivariate measuraments has created more controversy and suffers from more drawbacks than it's univariate counterpart. The univariate measures of process capability indices are not always directly extendable to multivariate problems. See Kotz & Johnson (1993) and Niverthi (1998), for detailed references. However, the normative approach taken by the author is directly extendable to the multivariate problem, which is mentioned in Bernardo & Irony (1993). In the univariate problem, the specification of the tolerance region A is usually the interval (L, U) where L and U are respectively lower and upper specification limit. Thus, for multivariate problem, if the tolerance region is a well defined set e.g., a p-dimensional rectangle, where p is the number of variables, then $P(x \in A|D)$ can be obtained easly. The model described in Section 2, can also be extended easily to the multivariate scenario.

However, the specification of the tolerance region in the multidimensional case is often specified as

$$V = x : h(x - T) \leq r_0 \tag{4.1}$$

where r_0 is prespecified and h is a known positive function, e.g., $h(x - T) = |x - T|^{\nu}$, ν a constant, and T is the target value. Fortunately, a Monte Carlo estimate of $P(\nu|D)$ can be obtained using a sampling based approach. Bayes capability index can be defined using an inverse monotonic cdf tranformation, e.g., $\Phi^{-1}(\cdot)$ on $P(\nu|D)$ where Φ is again the standard normal cdf.

Another advantage of using the sampling based approach is that one can find samples from the posterior predictive distribution of $h(x-T)|D$. Now following Chen (1994), one can define a multivariate process capability based on posterior predictive distribution as

$$MC_b = r/r_0$$

where r is defined as $P(h(x-T) \leq r|D) = \gamma$, where γ is, say the upper 95 percentile of the posterior predictive distribution of $h(x-T)|D$.

Another approach for the multivariate problem is to develop vector valued process capability indices from a Bayesian perspective. Niverthi (1998) considered this in detail. Finally, the author has made a very significant contribution, for which he is to be complimented.

References

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Chen, H. (1994). A multivariate process capability index over rectangular solid tolerance zone. *Statistica Sinica*, **4**, 749-758.

Chen, M-H. & Dey, D.K. (1998). Bayesian modeling of correlated binary response via scale mixture of multivariate normal link functions. *Sankhya, Series A*, to appear.

Geisser, S. (1993). *Predictive Inference: An Introduction.* London: Chapman and Hall.

Niverthi, M. (199). Bayesian methods for quality control and software reliability. *Ph.D. thesis, Departament of Statistics, University of Connecticut.*

Gelfand, A. & Smith, A. (1990). Sampling babased approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

A. Fries
Institute for Defense Analyses, USA

Five years ago I was approached at work by someone from a different division in the midst of supporting an ongoing Defense Science Board (DSB) study and documenting their findings and recommendations. (As I understand it, DSBs are occasionally convened by the U.S. Department of Defense to review current policies and procedures, and to provide independent advice to the Secretary of Defense.) Apparently this particular DSB, focusing on engineering in the manufacturing process, had received several briefings on current industry practice, including the standard use of the C_p and C_{pk} Process capability Indices (PCIs). I was asked to help co-author an appendix on the subject to introduce and define the terms, and, more importantly, to establish minimum sample size requirements (Fries and Richter, 1993). The tone of the request and accompanying dialogue seemed to be that the conclusion had already been reached that C_p and C_{pk} were the way to go, and that the only outstanding question was how small could one make the inspection sample sizes. After some preliminary reading and follow-up reviews of the literature, I pointed out some problems with this presumption, all of which I am gratified to note have also been reported by Professor Singpurwalla in "The Stochastic Control of Process Capability Indices".

This backdrop takes to me to my first substantive comment. The first two sections of the subject paper provide an expository summary of traditional PCIs, highlighting their interrelationships, what they actually measure, and what they do not. Of primary importance is that there are several common misconceptions entailed in their ordinary interpretation. My only complaint about this material is that it is five years too late to help me with my PCIs "assignment"!

Another key observation made by Professor Singpurwalla, which in my previous exposure to the topic regrettably took me longer to fathom and appreciate than it should have, is that the strong assumption of process stability implicitly accompanies routine applications of PCIs. This realization and its practical consequences did not necessarily please those who desired to minimize the burden of sampling and testing, beginning with the earliest stages of product development and production. I strongly concur with Sinpurwalla's veneration of the normative approach to process capability, pioneered by Bernardo and Irony (1996), but am prompted to raise several questions.

First, the assertion is made that at the end of a rating period the assessment whether a prescribed portion of the items manufactured during that period met specifications is equivalent to deciding whether to continue production as is or to intervene. My sense is that the degree of compliance, e.g., 100 percent or marginal, might influence one's determination of intervening or not. For that matter, a persistent pattern of progressively worsening compliance might also trigger an intervention, even when no individual breach of engineering tolerances has been manifested. A corresponding methodological query is whether the predictive distribution of future observables for the next rating period should be conditioned on all past observed values vice merely those from the last rating period?

Another set of questions relates to the general form of the difference between the utilities of decisions to continue production as is and to intervene, as reported in the discussion preceding Equation (2.8). Does the analytical form given there, expressed as $aR - bN + Q^*$ with $a > b$, hold specifically for the second pair of utility functions presented (here I openly admit my personal ignorance of some of the key cost terminologies) and more generally for other representations? Do there exist any rational utility functions for which the common PCIs are coherent (in the sense of Bernardo and Smith (1994), Chapter 1)?

Third, it is not clear what is truly Bayesian about the "Bayes" capability index, $C_B(v)$, defined in Equation (2.10), in the sense that "subjective probability" does not appear to enter directly into the calculations? More thoughts on Bayesian considerations appear below.

In Section 3 of the subject paper, Professor Singpurwalla directly confronts the "ugly beast" of process instability and seeks not only to tame it but also to "control" it. This is a truly ambitious undertaking and he clearly deserves both our admiration and respect. It should not be surprising, however, that, as with any initial excursion into uncharted territory, progress is incremental and open to debate. Obviously, if the problem were easy, the "optimal" solution would already be available and universally acknowledged. Two classes of comments and questions follow, ones specific to the particular construct pursued in the subject paper, and others that apply generally to the broad concept of process control.

My guess is that many practitioners would find it difficult, if not impossible, to implement the "control" approach outlined in Section 3. One major obstacle I conjecture would be a fundamental lack of understanding of what the terms Yi and Hi physically represent. For instance, even if the cost parameters were known how does one calculate the right-hand-side of Equation (3.4)? For that matter, how does a computed value of Yi translate into a specific process intervention action?

It is certainly fair to demand that any "control" methodology be readily comprehensible and robust to unintended interpretations. Other general issues include:

- Should adjustments be made after each inspection period?; i.e., why "mess" with a process that is not exhibiting any (strong) symptoms of undesired performance?
- Should information be utilized from previous inspection periods, vice just the latest?
- How can the subjective assessments of product developers and engineers be incorporated into determinations, both at the beginning of manufacturing and subsequently after each inspection period and intervention event?

In summary, Professor Singpurwalla's paper is an impressively useful contribution to the theory and application of PCIs. The overview of current methodologies is especially informative, for neophytes and experienced practitioners alike. Further, his solitary intrepid foray into the realm of process "control" should serve to engage the remainder of the statistical, operations research, and industrial engineering communities, to develop alternative approaches, and to discuss and test out (in the real world as well as in simulation studies) their relative merits and weaknesses. That is, after all, how progress is attained.

J.K. Ghosh
Indian Statistical Institute, India, and Purdue University, USA

Professor Singpurwalla provides a very illuminating review of existing PCI's and goes on to discuss their relevance to a Bayesian as well as other more appropriate Bayesian options.

I have seen the PCI's before in my interactions with colleagues in Quality Control but had not been impressed by any but the simplest measure. Professor Singpurwalla's review makes them sensible and attractive by throwing light on their evolution. I find this section very useful.

The Bayesian sections are important for what they promise rather than what they achieve. In particular this may be the best way to handle multivariate problems. The two more fundamental problems are to take these sections a step further where they can be implemented and test the implementations for some robustness against misspecification of prior, utilities, model. The prior should capture the notion of a substantial change, a sort of change point, when the process goes out of control. The utilities should contain a term that measures the damage to a process which is left out of control for too long. A related problem in optimal stopping is posed and solved in a simple case by Shirayev (1978).

The discussion of the role of the PCI's in the Bayesian approach is interesting but incomplete for the following reasons. The deduction of (2.11) from the previous step is incorrect because Φ^{-1} does not operate in so simple a way on the difference of two Φ-terms. The equation (2.10) and its predecessor also seem odd because of the inclusion of a completely arbitrary v. How can an arbitrary v add insight ? A more promising line may be to take a normal model and a conjugate prior. Such a combination should lead to something like a PCI-based decision rule.

References

A.N. Shirayev (1978). *Optimal stopping Rules*, Springer, New York.

M. A. Gómez–Villegas
Universidad Complutense de Madrid, Spain

Let me begin with my sincere congratulations to the author. This is, as far I know, the second time that Professor Singpurwalla uses the Bayesian methodology, to make clear existing aspects of the frequentist approach, the other one is his landmark paper with Menhold(1983) about understanding the Kalman filter.

The paper has two principal parts, the first one is dedicated to contemplate the capability indices in the quality control, from the point of view of the bayesian decision theory and the second one to predict and to control the quality of future output using a dynamic linear model.

I am going to concentrate the discussion on three main points about what I would like to know the author opinion.

Why not to determine a credible interval? Once the possibility of traditional indices to measure the quality of a manufactured product has been seen, why not obtaining a *credible set* C using the *predictive distribution* of $\pi(R_{i+1} \mid N, \underline{x}(n))$ that satisfies $P(R_{i+1} \in C \mid N, \underline{x}(n)) \geq 1 - \alpha$? If, in the notation of the paper, the interval (LCL, MCL) is contained in C the process is in control. In this way the introduction of the utility function gives a meaning to $1 - \alpha$, it would be

$$\frac{b}{a} - \frac{Q^*}{aN} \leq 1 - \alpha,$$

in the simplification supposed in the paper. Are not all the other considered aspects an *ad hoc* behaviour too?

Robustness The considered procedure depends fairly on the ratings periods T_i, where the control is done, and the number n of items observed in each period $[T_{i-1}, T_i)$. May the author give any idea of how is this repercussion or how could it be dealt with?

Closely connected with the above comment and with reference to the control of process capability, a Kalman filter is used. An empirical analysis involving simulated and real data is considered where all the observations are normal. An important application would be the validation of the model with the errors sources distributed as Student distributions, in the line of Meinhold and Singpurwala (1989) or Girón and Rojano (1994), also relevant is Gómez, Gómez–Villegas and Marín (1998) where a multivariate generalization of the *power exponential* distribution is given.

References

Meinhold, R.J. and N.D. Singpurwalla (1983). Understanding the Kalman filter. *The American Statistician*, **37**, 2, 123–127.

Meinhold, R.J. and N.D. Singpurwalla (1989). Robustification of Kalman filter models. *Journal of the American Statistical Association*, **84**, 479-486.

Gómez, E., M.A. Gómez–Villegas and J.M. Marín (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics (Theory and Methods)*, **27**, 3, 589-601.

T. Z. Irony
The George Washington University, USA

1 Introduction

I would like to congratulate Nozer Singpurwalla for a clear, thoughtful and comprehensive overview and critique of Process Capability Indices (PCI). He gave us an excellent historical picture of the development of capability indices and the clear interpretations provided in his article are greatly needed in the industrial engineering and quality assurance communities. In fact, I would like to see his article widely disseminated among those communities, in order to generate some thought and discussion that would facilitate the acceptance of the new Bayesian Capability Indices presented here.

2 Capability analysis: a decision problem

I commend the author for agreeing that PCI's should be used for making decisions and consequently should be conceived under a formal decision theoretical framework. According to his article, "the traditional capability indices have been playing a passive role", which is "to retrospectively ensure that the number of nonconforming units in a batch is bellow a specified limit". By traditional PCI's we mean

$$C_p,\ C_{pk},\ C_{pm},\ C_{pmk} \text{ and } C_p(u,v).$$

I would like to reinforce Singpurwalla's criticism because PCI's have been used to make decisions and I believe that they were conceived with this idea in mind. The fact that C_p should be in excess of 1.66 in pilot (qualification) runs means that the process is not qualified (not accepted) if $C_p < 1.66$. Moreover, PCI's have also been used for making predictions. Deming [(1982 - Chapter 7); (1989 - Chapter 11)] stresses repeatedly the importance of stability or statistical control of a process (or exchangeability in the Bayesian parlance) because "a process out of control has no predictive value". Practitioners perform capability analysis with the idea that past performance should predict future behavior, and Deming warned that this would hold only if the process is stable. For all this reasons I strongly agree that the traditional capability indices have been inadequate for not being constructed according to the normative approach dictated by Decision Theory, and I thank Singpurwalla for recognizing the Bayesian Capability Index as a step in the right direction.

3 Incoherence and the traditional capability indices

I would like to comment on Singpurwalla's analysis of utility functions. According to his notation, decision should be made iff

$$P\left(X \in A \middle| \underline{x}_i(n)\right) \geq \frac{b}{a} - \frac{Q^*}{aN}, \quad \text{where } a = (g+C-r) \text{ and } b = (g+C-r-i).$$

Here, $a > b$, unless i, the cost of inspecting each item is null.

Consequently, he notes, if Q^*/N is larger than b,the right hand side is always negative in which case decision to continue production as is, is always made. I would like to mention that this is an extreme case in which the cost of stopping the production is so high that it would be better to accept the whole batch even if all items are nonconforming. This situation is not admissible and if it happens, the tolerance limits should be revised. Otherwise, there is no point in performing capability analysis, just keep producing because anything will be better than stopping.

Another point raised by the author which I found very interesting is that requiring C_{pk} to be greater than 1 would lead to incoherence when regarded under a decision theoretic framework. That could be avoided if the batch size, N, is very large and the cost of inspecting an item, i, is small, because in fact $\Phi(3) = 0.9985 < 1$, and if N is large enough, one could have

$$b = \frac{Q^*}{N} + 0.9985a \quad \text{and} \quad b < a.$$

The same would hold for C_p since $\Phi(6) < 1$, but to avoid incoherence, the value of N would have to be huge. In capability analysis, the probabilities of producing conforming items are usually extreme (greater than 0.999) and cannot always be approximated by 1.

4 A remark on the Control of Process Capability

Although the idea of Control of Process Capability appears to be natural and straightforward, it will generate major controversy among the quality assurance community, mainly due to Juran's and Deming's criticism of overadjustment. In fact, I also have a problem envisioning how to automatically control a production based in a mathematical model, without a major decision making process involving management and experts in the system.

One of Deming's major contributions was to discuss and to emphasize the importance of the stability of a system, i.e. the importance of maintaining the system in statistical control. According to him, the significance of the concept of a stable process was already highlighted by Shewart. A stable process is a process subjected only to "common (or assignable) causes of variation", which reflect the natural variation of the process. A stable process has a definable identity and capability and its behavior in the future is predictable. On the other hand, a process goes out of statistical control whenever one or more "special causes of variation" arise.

Deming stresses that in order to analyze any process, to use it to make predictions, and to define its capability, it must be in statistical control. Moreover, in order to improve a system, statistical control must be achieved first (Deming and Juran). They explain this idea by stating that there are two distinct ways to improve a system:

1. To achieve stability : this is done by removing the special causes of variation. The discovery and removal of a special cause of variation is usually responsibility of someone who is connected directly with the operations that yield the available data.

2. To improve the whole process: this may be done only by management who can modify the system as whole.

Consequently, the issue of controlling the system is very complex, and control cannot be achieved automatically. Every time there is a drift in the system, a search must be conducted in order to find and to remove the special cause of variation that caused such a drift. The action of adjusting the system every time there is a drift will lead to what Deming called "overadjustment" [Deming (1982) - pp. 116; Deming (1988)- pp. 327] and the system will branch off more and more away from the target. This idea is illustrated by Deming's (and Lloyd Nelson's) celebrated "funnel experiment". It is also important to mention the work of Grubbs (1983) when talking about the problem of optimum convergence to the target.

In summary, Prof. Singpurwalla's idea of using the mathematical model for process control and for obtaining the Control of Process Capability must be utilized with caution. The controller should be used only when a special cause of variation is located and only in order to eliminate such a cause. It would be nice to have an example of how the controller would be defined in practice.

References

Deming, W. E. (1982). *Quality, Productivity and Competitive Position.* M.I.T. Center for Advanced Engineering Study, Cambridge, MA.

Deming, W. E. (1988). *Out of the Crisis.* M.I.T. Center for Advanced Engineering Study, Cambridge, MA.

Grubbs, F. S. (1983). An optimal procedure for setting machines. *Journal of Quality Technology,* **5**, 4, 155-208.

W. Kliemann
Iowa State University, USA

1 Control of Process Capability Indices

A Bayesian approach to process capability indices (PCI's) is a new area, initiated by Bernardo and Irony (1996). The current paper by Singpurwalla suggests the inclusion of control theoretic aspects into this framework. The idea goes roughly like this: Let $\{T_i; i \in \mathbb{N}\}$ be an increasing sequence (of time points) and let $\{X_i; i \in \mathbb{N}\}$ be the process variable of interest at time T_i. This variable is given as $X_i = \theta_i + r_i$, where θ_i is the important process parameter, and r_i is a disturbance term. The control Y_i affects θ_i via $\theta_i = \theta_{i-1} + H_i Y_{i-1} + \omega_i$, where H_i describes the amount of influence of the controller Y_{i-1} o n the parameter θ_i, and ω_i is another disturbance term. Hence control of θ_i has to be understood as *parameter tuning* according to some objective function $\mathcal{C}(\theta_i, Y_{i-1})$. Since the underlying process is not modelled dynamically, this leads to a *static* optimization problem, which can be solved explicitly for quadratic objective functions separating θ_i and Y_{i-1}, as shown by Singpurwalla in his paper.

However, one should think of the Bayesian decision framework as being dynamic in the following sense: At time T_{i+1} a decision is made based on the data D_{i+1} (in the interval $[T_i, T_{i+1})$) and prior information, according to (static) utility function(s) and a decision criterion, whose acceptance depends on the predictive distribution $P(X_{i+1} \in \cdot | D_{i+1})$. Singpurwalla takes this idea one step further by adding a control component and a corresponding objective function $\mathcal{C}(\theta_{i+1}, Y_i)$: In a first step the control parameter Y_i is chosen such that $\mathcal{C}(\theta_{i+1}, Y_i)$ is optimal. Here $Y_i = f(C_{1i}, C_{2i}, m_i, H_i, T, \widehat{\theta}_i)$, where C_{1i}, C_{2i}, m_i are parameters of $\mathcal{C}$, T is the target value of the process, and $\widehat{\theta}_i$ is an estimator of $\widehat{\theta}_i$, depending on $\widehat{\theta}_{i-1}$, model parameters, and the data D_i in $[T_{i-1}, T_i)$. In a second step, at time T_{i+1}, a decision is made based on the data D_{i+1} of the process with optimal control setting Y_i^0, using the predictive distribution $P(X_{i+1} \in \cdot | D_{i+1}, Y_i^0)$.

The goal of the additional control step is to tune Y_i such that $P(X_{i+1} \in A | D_{i+1}, Y_i^0)$, with A the tolerance region of X, becomes large, while penalizing high control cost. Singpurwalla achieves this using the objective function $\mathcal{C}(\theta_{i+1}, Y_i) = (Y_i - m_i)^2 C_{1i} + (\theta_{i+1} - T)^2 C_{2i}$, which involves two separate quadratic terms penalizing cost (deviations from the minimal cost setting m_i) and deviations form the process target value T. Note that separation and quadratic cost lead to the explicit optimal solution Y_i^0.

The second term of the objective function is interesting. Like the PCI C_{pm} it takes into account deviation from T. In the presented setup, the utility function(s) $\mathcal{U}$ do not depend on T, nor does the decision criterion. Hence a training of the process to the midpoint M of the tolerance region A will, in general, lead to larger values of $P(X_{i+1} \in A|D_{i+1}, Y_i^0)$. Therefore, the proposed strategy is consistent with the described decision setup if $T = M$. In this case the quantity (with $A = [L, U]$)

$$P(X_{i+1} \in A|D_{i+1}, Y_i^0) \sim \Phi\left(\frac{U - \widehat{\mu}_{i+1}}{\widehat{\sigma}_{i+1}}\right) - \Phi\left(\frac{L - \widehat{\mu}_{i+1}}{\widehat{\sigma}_{i+1}}\right)$$

may be used for decisions related to process capability. If $T \neq M$, a conflict between the two steps in the decision process may arise. This conflict can be resolved, e.g., via including the target value T in the utility function(s) $\mathcal{U}$, resulting in a different value with which the predictive probability $P(X_{i+1} \in A|D_{i+1}, Y_i^0)$ is compared, or via different weights of the variable X on the intervals $[L, T]$ and $[T, U]$, in analogy to the PCI C_{pm}. In a Bayesian decision theoretic context, the first option is, of course, preferable.

It remains to discuss the properties of the proposed 'stochastically controlled process capability index' $C_{Bi}(V)$. We will do so for the uncontrolled case, since the arguments are similar and easily adapted to the controlled situation.

2 Bayesian Interpretation of Process Capability Indices

In this section we use the following notation based on the paper under discussion: $A = [L, U]$ is the specification interval or tolerance region with midpoint M, X is a random variable, describing the process parameter of interest, $P(X \in \cdot|D_i) =: P(X \in \cdot)$ is the distribution of X given the data D_i, $\Psi : \mathbb{R} \to [0, 1]$ is the *cdf* of P, which we assume for notational simplicity to be strictly monotone with inverse Ψ^{i-1} and density ψ, $\Phi : \mathbb{R} \to [0, 1]$ is the standard Gaussian *cdf* with inverse Φ^{-1} and density φ. The univariate setup is chosen for notational convenience, all arguments are easily adapted to the multivariate case.

A Bayesian decision setup, as developed in Bernardo and Irony (1996) or the current paper, leads to the choice of a strategy d_1 (interpreted as 'continue operation of the process') iff the expected proportion of conforming items exceeds a certain threshold B, i.e. iff

$$P(X \in A) \geq B. \tag{2.1}$$

The specific form of B given in paper is

$$B = \frac{b}{a} - \frac{Q}{aN} \tag{2.2}$$

where a is related to the profit of a conforming item, b to the cost of producing an item, N is the total number of items produced in a time interval $[T_i, T_{i+1})$, and Q represents a fixed cost. Note that the criterion (2.1) is invariant under strictly monotone increasing transformations.

In order to relate the criterion (2.1) to traditional PCI's, we consider the transformation $\Psi_y : \mathbb{R} \to [-\Psi(y), 1 - \Psi(y)]$, $\Psi_y(x) = \int_y^x \psi(u)du$, and we denote by Ψ_y^{-1} its (strictly monotone increasing) inverse function. The transformation Ψ_y can take on negative values, but as we will see, it serves our purposes. We obtain $\Psi_L^{-1}(P(X \in A)) = \Psi_L^{-1}(\Psi_L(U)) = U$ and $\Psi_L^{-1}(0) = \Psi_L^{-1}(\Psi_L(L)) = L$, hence $U - L = \Psi_L^{-1}(P(X \in A)) - \Psi_L^{-1}(0)$. Therefore the criterion (2.1) is equivalent to

$$C_p = \frac{U-L}{6\sigma} = \frac{1}{6\sigma}\left[\Psi_L^{-1}(P(X \in A)) - \Psi_L^{-1}(0)\right] \geq \frac{1}{6\sigma}\left[\Psi_L^{-1}(B) - \Psi_L^{-1}(0)\right]. \tag{2.3}$$

In this formula the variance σ^2 of X is assumed to be known. Similar expressions can be derived using Ψ_U and Ψ_M.

For the interpretation of the PCI C_{pk} we consider the case $\mu > M$, where μ is the mean of X, again assumed to be known – the case $\mu < M$ is completely analogous using Ψ_U. Setting w.l.o.g. $\mu = 0$, we obtain as above: Criterion (2.1) is equivalent to

$$C_{pk} = \frac{U}{3\sigma} = \frac{1}{3\sigma}\Psi_L^{-1}(P(X \in A)) \geq \frac{1}{3\sigma}\Psi_L^{-1}(B). \tag{2.4}$$

If the predictive distribution of X is Gaussian with mean μ and variance σ^2, then the criteria above specialize to

$$C_p = \frac{1}{6}\left[\Phi_L^{-1}\left(\Phi\left(\frac{U-\mu}{\sigma}\right) - \Phi\left(\frac{L-\mu}{\sigma}\right)\right) - L\right] \geq \frac{1}{6}\left[\Phi_L^{-1}(B) - L\right] \tag{2.5}$$

and

$$C_{pk} = \frac{1}{3}\Phi_L^{-1}\left(\Phi\left(\frac{U-\mu}{\sigma}\right) - \Phi\left(\frac{L-\mu}{\sigma}\right)\right) \geq \frac{1}{3}\Phi_L^{-1}(B). \tag{2.6}$$

These criteria should be compared to Singpurwalla, Formulas (2.9), (2.10) and above (2.11), as well as Bernardo and Irony (1996), Formula (17). The Bayes capability index in these papers is defined as

$$C_B = \frac{1}{3}\Phi^{-1}(P(X \in A)). \tag{2.7}$$

Let us consider a modified index, given by (for $\mu \geq M$)

$$C_B^* = \frac{1}{3}\Phi_L^{-1}(P(X \in A)). \tag{2.8}$$

(If there is compelling reason to assume that the predictive distribution $P(X \in \cdot)$ of X is not Gaussian, then Ψ_L^{-1} should be used. The arguments below remain valid for this case, with the obvious modifications.) We obtain for $P(X \in \cdot) \sim \mathcal{N}(\mu, \sigma^2)$ with known μ and σ, which we set w.l.o.g. to be $\mu = 0, \sigma = 1$:

1. The Bayesian decision criterion (2.1) leads to acceptance of the strategy d_1 iff

$$\int_L^U \varphi(u)du \geq B, \tag{2.9}$$

which means in terms of C_B

$$\int_{-\infty}^{3C_B} \varphi(u)du \geq B \tag{2.10}$$

and in terms of C_B^*

$$\int_L^{3C_B^*} \varphi(u)du \geq B, \tag{2.11}$$

which is equivalent to (2.4) with $C_B^* = C_{pk}$, while (2.3) results in

$$\int_{L}^{L+6C_p^*} \varphi(u)du \geq B \tag{2.12}$$

with $C_p^* = C_p$ as in (2.3) and (2.5).

Note that $C_B^* = C_p^*$ for $\mu = M = T$. For known μ and σ, both of these indices are equivalent to the established PCI's, i.e. $C_B^* \sim C_{pk}$ (for $\mu \geq M$) and $C_p^* \sim C_p$. On the other hand, the Bayesian index C_B is asymptotically equivalent to C_{pk}, compare Bernardo and Irony (1996), Formulas (18) and (19), and Singpurwalla.

2. The criteria for the acceptance of the strategy d_1 boil down to the same inequality for the indices C_B, C_B^*, and C_p^*, since different transformations for the r.h.s. B in (2.1) are used. Note that $C_B < C_B^*$ if both describe $P(X \in A)$. On the other hand if the value of the Bayesian PCI's is equal, $C_B = C_B^* = c$, then $\int_{-\infty}^{3c} \varphi(u)du > \int_{L}^{3c} \varphi(u)du$, and C_B leads to acceptance of larger B-values, while C_B^* retains the established interpretation for C_{pk}.

3. The Bayesian criterion (2.9), as well as the derived criteria (2.11) and (2.12) determine an upper limit $\int_{L}^{\infty} \varphi(u)du =: C$ for the acceptance of the strategy d_1 : d_1 will not be accepted if $B \geq C < 1$. The index C_B, on the other hand, increases for $C_B \to \infty$ to 1, resulting in a different upper limit.

4. Let us consider the interpretation of the various PCI's in terms of underlying utilities, represented in B. In economically meaningful situations, i.e. $a > b > 0, Q \geq 0$, we have $B \leq \frac{b}{a} < 1$, and hence an upper limit $C < 1$ on the l.h.s. of (2.1) makes sense. Of course, all four criteria (2.9)–(2.12) can be satisfied for certain parameter combinations with $b > a$, but these combinations will not occur in practice. Infact, non of the criteria requires $b > a$ to hold for the acceptance of strategy d_1, and hence none of them is incoherent, including the traditional PCI's in (2.11) and (2.12).

5. The Bayesian framework leads to another interesting observation about the traditional PCI's. In practice, their values are usually chosen with the 3σ- interpretation of the Gaussian distribution in mind. If the actual distribution of the variable X is non-symmetric, has heavy tails, or if μ

deviates from the midpoint M of the specification interval, then (often arbitrary) 'adjustments' are made. Formulas (2.11) and (2.12) show a way of how to formulate the choice of C_p and C_{pk} as optimization problems in the Bayesian decision context, which reads for C_B^* (with $\mu > M$) as follows: Minimize $c > 0$ such that

$$\int_L^{3c} \psi(u)du \to \min$$

under the constraint

$$\int_L^{3c} \psi(u)du \geq B.$$

The optimal solution, for continuous ψ and $B < \int_L^{\infty} \psi(u)du$, is obviously

$$C^0 = \frac{1}{3}\Psi_L^{-1}(B),$$

and $C_{pk} = C^0$ results in a justifiable choice of the PCI. Reasoning for the index C_p is similar.

6. The expression derived in Theorem 1 of Bernardo and Irony (1996) for the index C_B remains true for the index C_B^* as presented in (2.7), if the transformation Φ_L^{-1} is used instead of Φ^{-1}. A similar expression, with the obvious modifications, *also holds for the index* C_p^* *from (2.11).*

7. We emphasize again that all arguments above hold, mutatis mutandis, for arbitrary *cdf*'s Ψ with continuous density ψ, and if a proper definition of Ψ_L^{-1} is used, for any *cdf*. Thus the Bayesian decision approach to PCI's and stochastically controlled PCI's yields far-reaching justification and guidelines for process capability indices, which retain the traditional interpretation of the established standards if C_B^* is used. Note that analogous observations can be made using Φ_U and Φ_M.

References

Bernardo, J.M. and T.Z. Irony (1996). A general multivariate Bayesian process capability index. *The Statistician*, **45**, 3, 487–502.

S. Kotz
The George Washington University, USA

To paraphrase August de Morgan's assessment of Laplace's contributions as published in *The Dublin Review 3* (1837), p.348:

> "Singpurwalla never attempted the investigation of a subject without leaving upon it the marks of difficulties conquered: sometimes indirectly ... ." Indeed, it should be evident to anyone familiar with Singpurwalla's work that he is a "thinker about probability and statistics" and not merely and indefatigable calculator or an sympathetic dogmatist-although it is posible that these three aspects may have been combined.

The paper under review is no exception and it bears the unmistaken features of his outstanding and often provocative research in the field of reliability theory for over some twenty years.

The first part of the paper (Introduction and Sections 2.1-2.3) provides a lucid and compact assessment of the historical developments and the current state of the theory of process capability indices. There are a few minor inconsequential imprecisions and lacunas inevitable in any short review of a rather vast subject which has been so popular and controversial in the last decade or so. A comparison with a parallel description presented in Kotz and Johnson's (1993) volume and especially in the forthcoming Kotz and Lovelace's (1998) more elementary monograph might be instructive.

It is only in subsection 2.4 and in a brief Section 3 that the author presents his innovative ideas emphasizing the normative approach to process capability and his suggestion for controlling process capability, with a laudable aim to weaken the assumption of process stability inherent in all the "classical" PCI's.

The arguments in Section 2.3 which essentially lead to sequential PCI's are imbued with the well known concepts of probabilistic decision making in which the expected utility plays a pivotal role, are cleverly and gracefully combined. The argument results in the basic equation (2.8) which is no doubt an interesting result, certainly worthy of further investigation and refinement. The author seems to overemphasize the drawbacks involved in the possible lack of coherence of the "initial" (primitive) process capability indices C_p and C_{pk} which are also marked by several other defects.As the article on "Coherence" by Patrizia Berti and Pietro Rigo (Dept. of Statistics, University of Florence)-to appear in the forthcoming Volume 3 of the *Encyclopedia of Statistical Sciencies* (Update)-convincingly shows incoherence is often an *inevitable* property of many useful and appropriate summary indices and occurs in various natural situations. The choice of the utility of decisions d_1 ("continue productions as it is") and that of d_2 ("intervene") are admittedly somewhat simplistic albeit convenient for an easy derivation of the final conclusion -Equation (2.8). Equation (2.11) is, however, more problematic where the Gaussian (predictive) distribution of X is assumed. Under this assumption the problem of the sampling distributions of $\widehat{C}_p$, $\widehat{C}_{pk}$, $\widehat{C}_{pm}$ and even other more refined indices has been quite satisfactorily solved within the classical framework (see, e.g., Kotz and Johnson (1993)) and the conclusion here reached by Singpurwalla may seem (at least for an uninitiated) to be reiterating the obvious.

The last Section of the paper is potentially most promising and illuminating. The models proposed are admittedly only prototypes. Again the (unjustified?) Gaussian asssumption plays an important role and the decomposition of the total cost given by (3.3) resembles too much the classical approach which is justly criticized by Singpurwalla. Further weeding out and refinements are required to render the writer's bold ideas suitable for practical applications, especially due to strong rift existing between the theoreticians and practitioners in this particular area.

At this early stage the author of these short comments can only offer an unqualified endorsement and recommendation to pursue further this promising avenue towards a more meaningful and flexible assessment of process capabilities. The writer is certainly on a right track.

References

Berti, P. and R. Rigo,(1998). Coherence-an Update. To appear in the *Encyclopedia of Statistical Sciences*, **Vol. 3** (Update), J. Wiley, NY.

Gillispie, C.C. (1997). *Pierre-Simon Laplace. A Life in Exact Science*, Princeton University Press, Princeton, NJ, pp.272-273.

Kotz, S. and N.L. Johnson (1993).*Process Capability Indices.* Chapman and Hall, London.

Kotz, S. and C.R. Lovelace (1998).*An Introduction to Process Capability Indices.* Arnold, London. (forthcoming)

D.V. Lindley
United Kingdom

Knowing nothing of the engineering associated with the use of PCI's, it is perhaps foolish to comment. However, I will stick my neck out and offer some general remarks, because there is at least a prima facie case, demonstrated in this paper, that theory can be applied with useful engineering consequences. Hopefully an engineer will be provoked by my folly to think afresh about the problem behind PCI's and, in the worst scenario, will have light shed on the process. The best scenario would have the current procedure fundamentally altered.

Singpurwalla, as Bernardo & Irony (1996), are surely right to point out that the engineer has a decision problem: he has to choose between a number of actions. An index based purely on Tchebychev's inequality may be adequate, but only consideration of potencial acts can demonstrate this; probability cannot be enough. As these authors point out, utility must also be introduced. Let us consider the ingredients in the engineer's problem.

At any time in the production process, there are items already manufactured and potencial items to be made in the future. Decisions concern both types. The manufactured items may be accepted or rejected. The future production may be retained with the same process in the past, or there may be some intervention. Utilities have to be considered for these possibilities. It is somewhat disappointing to see such simple, and maybe unrealistic, forms used at the moment. For example, is it really sensible to impose upper and lower limits? Is it realistic to say that an item just above the upper is hopeless, whereas one just below is fine? A function with a maximum at the target value and decreasing either side, often in an unsymmetric way, may be more realistic than the 0-1 form used here. It is regrettable that statisticians are so casual about utility, failing adequately to relate its value to the practicalities of the problem.

From considerations of utility one has to pass to probability. Here the statistician is on more familiar ground and has provided a range of ideas. A most valuable reference here is the book by West & Harrison (1989) which explores a class of probability models and associated decisions that would appear to apply to the type of data to which PCI's refer. In many respects they generalize the sound ideas that are used in Section 3 of the present paper.

I do not elaborate, since to do so sensibly requires active collaboration between engineer and statistician. It is the engineer's (or perhaps consumer's) utilities that are relevant, not the statistician's, whose role is limited to articulation of the practice. But were I do elaborate, we all know what the engineer would say - it is too complicated; something like $d/3\sigma$ is required. They are surely right, but with modern computing facilities the end product of these complicated deliberations on probability and utility can be simple. The choice of action demands maximization and expectation, two procedures which have been extensively studied by computer scientists and statisticians. It is possible to imagine a program where the engineer inputs the specific data and where the output is the optimum act; even an index if that is what is preferred. All the complexity lies within the computer. The consultation between engineer and statistician and the statistician's writing of the program will be complicated. But once done, the operation of decision is easy. Furthermore, if the program has flexibility in the provision of adjustable parameters that accommodate shifts in utility or probability, the one program can suffice for many applications.

It is my view that statisticians have not woken up to possibilities of practical applications of decision analysis, especially through intelligent use of computers; and, in particular, to the appreciation that engineers are not limited to simple expressions but can employ much better methods without any extra on-line burden. The present paper encouragingly moves us in the direction of the implementation of the ideas here outlined.

References

West, M. & Harrison, J. (1989). *Bayesian Forecasting and Dinamyc Models.* New York: Springer-Verlag.

M.F. McGrath
Sarnoff Corporation, USA

1 Introduction

In the late 1980s and early 1990s, leading manufacturing companies, such as Motorola and Texas Instruments, adopted a "six sigma" philosophy of customer satisfaction through total quality. At the heart of the 6σ philosophy is a belief that nearly perfect quality is essential for a product line or a company to survive in a global competitive marketplace. The term 6σ connotes a goal that products and processes should experience, on average, only 3.4 defects per million opportunities. The goal is admedly ad hoc. It is based on the premise that the old rule of thumb for the mean and variance of a capable process (that $\mu \pm 3\sigma$ be within specification limits) is not good enough. Instead, the goal should be $\mu \pm 6\sigma$. Design engineers and manufacturing operations managers have developed procedures using Process Capability Indices (PCIs) as a way to manage progress toward this goal. The commercial success of these leading companies has inspired widespread adoption of the philosophy and management procedures throughout industry. Today, procedures using PCIs (particularly C_p and C_{pk}) have become common in engineering, manufacturing and quality control practices around the world.

Professor Singpurwalla's paper provides a much needed framework for understanding the implications of alternative PCIs, and for bridging the gap from 6σ philosophy to practice. He has approached this subject primarily from the standpoint of assessing and controlling manufacturing processes. From this standpoint, the paper identifies serious deficiencies in current practice, including the ad hoc nature of setting requirements and the use of PCIs solely as retrospective indicators. Beyond providing an extremely valuable expository framework, Professor Singpurwalla proposes a new PCI and a procedure that can be used both for assessing and controlling processes.

I would like to offer two comments from a different standpoint, that of design engineering. The first comment suggests a resolution for the apparent dilemma of incoherence resulting from ad hoc specification of C_p

(Section 2.4 of the paper). The second suggests a possible direction for a better linkage between design engineering and manufacturing operations, based on Professor Singpurwalla's proposed process control procedure and its associated PCI.

2 Resolving the apparent incoherence caused by ad hoc goals

What matters ultimately to the designer is that $P(X \in A \mid H)$ be close to 1, where X is some key characteristic, A is the interval [LSL,USL], and H represents historical data on the process that generates X. For complex products involving n components and processes, each with critical characteristic X_i, the standard engineering practice is to assume independence and compute the Rolled Throughput Assembly Yield as $Y_{RT} = \prod P(X_i \in A_i \mid H_i)$ for $i = 1, \ldots, n$. When n is large, designers must select components and processes with values of $P(X_i \in A_i \mid H_i)$ very close to 1 in order for Y_{RT} to be acceptably high – hence the dictum that each process should have a goal of 6σ quality.

The attractiveness of focusing on a PCI, rather than directly on $P(X_i \in A_i \mid H_i)$, is that the PCI combines several controllable variables or parameters of product and process design. Engineers often use such indices to characterize designs and structure tradeoff studies – for example, evaluating alternative airfoil shapes using lift/drag ratio as a figure of merit. Section 2.2 of the paper provides an insightful discussion of the evolution of PCIs from C_p to C_{pk}, C_{pm} and other indices that take progressively more control variables into account.

The 6σ goal is equivalent to requiring $C_p > 2$ in steady state production. Since this is difficult to measure or achieve in production start-up, an ad hoc requirement such as $C_p > 1$ may be imposed initially, with an expectation of subsequent improvements to the process. Section 2.4 of the paper, "The Normative Approach to Process Capability", demonstrates that, under reasonable assumptions regarding cost, such ad hoc requirements can lead to incoherence. The problem here may simply be a question of deciding which cost assumptions are reasonable.

From the 6σ philosophical viewpoint, accepting defective items can have disastrous (to management) financial consequences – including loss of market share, declining stock prices and even demise of the company – to say nothing of adverse consequences to the consumer. The effect of requiring $C_p > 1$ is, implicitly, to set a very high value for C, the loss from accepting a defective item. The implied value of C may be several orders of magnitude larger than the other unit costs in the utility functions of Section 2.4 (viz.: g, the unit profit from accepting a good item; ν the unit manufacturing cost; q, the unit opportunity loss when production is stopped; and Q^*/N, the amortized cost of fixing the production system). When C is much larger than the other costs, we can resolve the apparent incoherence discussed following equation (2.11) as follows.

First, in equation (2.9) we redefine the quantity $\Phi^{-1}[(bN - Q^*)/aN]$ as $\Phi^{-1}[\Phi(Z_0) - \Phi(-Z_0)]$, where $\Phi(Z_0) - \Phi(-Z_0) = (bN - Q^*)/aN$. This interpretation is consistent with the definition of C_p, which assumes a normal distribution centered between the upper and lower specification limits. Then, in the discussion following equation (2.11), we find that requiring $C_p = 1$ means that $\Phi^{-1}[\Phi(Z_0) - \Phi(-Z_0)] = 6$, so $Z_0 = 3$, in which case $(bN - Q^*)/aN = .9973$.

When C is much larger than any of the variables g, ν, q or Q^*/N, the quantity takes the form $(C - \delta_1)/(C + \delta_2)$, where $0 \leq \delta_i \ll C$ for $i = 1, 2$. Hence requiring $C_p > 1$ implies that $(C - \delta_1)/(C + \delta_2) > .9973$, a condition which can be met if C is larger than $(\delta_1 + \delta_2)$ by a factor of 400 or more. From management's "big picture" viewpoint, it may be reasonable to believe that accepting a defect on a one dollar item could cost the company \$800. Had we imposed the condition $C_p > 2$ (i.e. the 6σ goal), it turns out C must be larger than the other cost variables by a factor of about 10^9 – an unreasonable penalty for accepting a defective item. Thus, the ad hoc requirements imposed by management, even if they are not incoherent, may lead to actions that cannot be economically justified. This is a main point of the paper, and one which deserves wholehearted agreement.

3 Better linkage between design engineering and manufacturing operations

In Section 3, Professor Singpurwalla proposes a method for active control of process capability, one that removes the restrictive assumption of process stability. This method explicitly addresses total expected costs and uses a minimum cost control function. This process can be characterized with a PCI called $C_{bi}(v)$, the stochastically controlled process capability index. Two pragmatic questions arise: (1) Can this method be implemented in practice, given the difficulty of quantifying the cost of quality? (2) Are design engineers prepared to start using such a new index, given their inertia to date in adopting indices more advanced than C_p and C_{pk}?

Regarding total expected costs, it is clear from the foregoing discussion that assigning a value to the cost of accepting a defective item is highly subjective and highly influential. A healthy debate among the company's business managers, design engineers and process managers may be needed to reach consensus on the cost function, which is a precondition to the stochastic process control procedure. In fact, this debate may turn out to be one of the most positive features of the new approach. Perhaps the best way to get started toward implementation would be to focus on a "problem" process currently in operation that is falling short of its PCI goals and is not stable. The benefits of the new stochastic process control method could be demonstrated in such an example, first through simulation and then in actual practice.

Regarding the question of acceptance by designers, thanks to Professor Singpurwalla's unifying framework, the change to a new PCI need not be traumatic. If necessary, the new index $C_{bi}(v)$ can be translated into an equivalent value for one of the traditional PCIs, but now based on cost and process control considerations. As confidence is gained with the new method, and engineers become familiar on an intuitive level with $C_{bi}(v)$, the translation step will no longer be needed.

In sum, this paper is a singular contribution, both in its unique unifying framework and in its promising step forward from process assessment to process control. If pragmatic implementations can be worked out in process control operations, migration of the new PCI into design engineering (or at least translation into terms design engineers are comfortable with) cannot be far behind. Professor Singpurwalla may indeed have sown the seeds of a revolution – one where product designers and process planners start to structure tradeoffs based on costs and process control considerations rather than traditional rules of thumb.

D. Peña
Universidad Carlos III de Madrid, Spain

This is an interesting paper and I want to congratulate the author and the editors of TEST for the interesting topic they have chosen. Some of the most important ideas about quality of processes and products were introduced by statisticians, and the study of quality problems has generated some key advances in statistical ideas, as the Neyman-Pearson theory of statistical testing, the decision theory by Wald, the idea of sequential tests and many others. Quality problems is an area in which Bayesian decision theory ideas are especially useful, but although this fact has been recognized long ago (see for instance Cox 1960 or Hamburg, 1962) the research on this area has not yet had a clear influence in the way quality control is applied in practice. All these reasons made this paper especially welcome. I will concentrate my discussion in two points: (1) the use of Process Capacity Indexes (PCI) as a tool for process control, (2) the particular time series model advocated in section 3 of the paper.

Starting with the first point, this paper generalizes the approach by Bernardo and Irony (1996) in which better PCI are derived from formal decision theory ideas. The approach is ingenious and elegant, have some nice mathematical properties, and can be useful in the initial application of process control in which the emphasis is on inspection. I agree that, with this objective, the Bayesian capability index is a good contribution, superior to traditional ones. However, I do not believe that PCI are useful in the most important stage in which the objective of the control is process improvement.

The standard way in which process control for measurements is implemented is by using control charts for the mean and the variability (range or standard deviation) of small samples taking during the process operation. The objective of this operation is to keep the process in state of control and identify assignable causes which may affect the process in order to improve it. Note that, in this approach, the information about the specifications limits is never considered. In fact, plotting in the control charts the specification limits instead of the variability of the statistic under consideration is generally considered as a very dangerous practice. In short, the specification limits are what we wish, whereas the variability is what we get and it is always a danger to take one for the other.

Capacity indexes have been introduced as descriptive measures to relate our wishes and our realities and the paper proposes to use it to guide action through a formal decision problem. This proposal is similar to the approach followed in acceptance sampling, are in which Bayesian decision theory has a long tradition (see for instance, Hamburg, 1962). However, this approach takes out of the analysis the key ingredient of process control : looking for assignable causes and reducing process variability. All products coming from the process can be good (if we have a large capability) and still we want to know how to reduce the variability and improve the process. The PCI are not designed for this objective.

For instance, consider the simulated cases in section 4. In case a) the process is in state of control, and if we use the standard approach by taking subgroups of size 100, as done in section 4 (which by all means is a very large size for industrial practice) we did not find any reason to stop the process. This is in contrast with the result using capability indexes, including the BCI. In case b) in which the mean of the process has an upward trend. The 3σ limits for the mean chart will be .3 (if the standard deviation is estimated from the first group in my simulation the limit will be .35). As the mean of the expected mean for the first group will be 1.5 (we get 1.48 in a simulation) the process will be stopped immediately and the problem investigated. A better approach is to monitor the process by using small samples of size 4 or 5 and apply the usual rules to identify trends, in this way the upward trend will be detected before the fist 100 observations are produced and the problem corrected. Similar comments can be extended to the other cases. However, the behavior of the capability indexes depend on the specification limits that are in the example completely arbitrary. If instead of (4.10, -1.90) we take a larger interval, all the capacity indexes can call for continuing the process, failing to identify that a trend has appeared.

The limitations of the PCI for process control has been already recognized in the literature. See for instance Bissell (1994), pag 244.

In summary, I would not recommend capability indexes for process control. They are not suited for identifying changes, keeping the process in control and improving it by reducing the variability. Rather they come from inspection ideas, similar to acceptance sampling, which has been dismissed by leading industries since the quality revolution in the 80's. I do believe that decision theory ideas are useful in quality control and a good example is the approach by Taguchi who has shown how to establish specification limits using this ideas.

My second comment is directed to the time series model considered in Section 3. In the paper, following a suggestion by Bernardo and Irony (1996),a dynamic model tis introduced to allow for autocorrelation among the observations. The model is equivalent to assuming that the measurement follows the IMA(1,1) model

$$X_{i+1} = X_i + a_{i+1} - \lambda a_i$$

where the parameter λ depend on the ratio between the two process noises, r_i, w_i. This model has a long tradition in quality control (Roberts, 1959) and the exponentially weigthed average chart derived from it is an standard tool for quality control (for instance, it is incorporated in the simplest standard statistical software for beginners as Statgraphics). The use of this model as a reference one to approximate the process time series structure has been study by Montgomery and Mastrangelo (1991). The use of adaptive control on this process has also been widely investigated (see Box and Luceño, 1997).

References

Bisell, D. (1994). *Statistical Methods fos SPC and TQM.* Chapman and Hall.

Box, G.E.P. and A. Luceño (1997) *Statistical Control by Monitoring and Feedback Adjustment.* J. Wiley.

Cox, D. R. (1960). Serial Sampling Acceptance Schemes Derived from Bayes Theorem. *Technometrics*, **2**, 3, 353-360.

Hamburg M. (1962). Bayesian Decision Theory and Statistical Quality Control. *Industrial Quality Control*, **19**, 6, 10-14.

Montgomery, D. C. and C.M Mastrangelo (1991). Some Statistical Process Control Methods for Dependent Data. *Journal of Quality Technology*, **23**, 3, 179-203.

Roberts, S.W. (1959). Control Charts based on Geometric Moving Averages. *Technometrics*, **1**, 239-250.

Rejoinder by N.D. Singpurwalla

The contributions of the twelve discussants reinforces my view that often it is the "discussion" that is the crux of a paper. I am grateful to all for taking the time to comment, and for their ideas. The discussions revolve around five themes: the control of indices, the nature of utilities, robustness, incoherence, and practical implementation. With respect to control, Professors Box and Kliemann, who appear to be favorably dispositioned, suggest enhancements, each with his own orientation. I do not view this (to quote Professor Box) as "parting company with the author"; rather, it is leading the way. The absence of an overlap in references is to be attributed to "guilt by ignorance". The Box-Luceño (1997) reference is a valuable lead, and if Box-Jenkins (1970) is any guide, this may revolutionized the practice of statistical quality control. The notion that feedforward/feedback be used to induce stationarity of the process and gain maximum possible process capability, is intriguing. It suggests that the entropy of the process be incorporated in the utility function. It also suggests an accommodation between the classical and the normative approaches for dealing with the process capability indices. Professor Kliemann who has made signal contributions to the mathematics of control theory, has de facto produced a paper of his own. Whereas I am not clear as to what he means by the statement "the underlying process is not modeled dynamically", I am in agreement with the suggestion that if the utility functions in the normative approach were to incorporate the target value T, then the conflict between the two steps in the decision process which arises when $T \neq M$ will be resolved. If by a discussion of the properties of the "stochastically controlled process capability index", Professor Kliemann implies its relationship with the traditional indices, then such an exercise will be futile, since the latter are passive entities. Professors Irony and Peña are not in favor of controlling the process capability indices on grounds that reducing process variability is the aim of control, not ensuring that X belongs to its specification interval. There may be some merit to this concern, especially, if as Professor Cox points out, indices are a vehicle for reducing dimensionality. Incidentally, I find this perspective on indices an imaginative one, worthy of further exploration; for example the statistics of object identification. Professor Irony's claim that the traditional indices are also used for decision making and prediction should be recognized under the understanding that the former is not normative, and that the latter assumes process stability. Professors Gomez-Villegas, Peña, and Dr. Fries draw attention to the important decisions of optimal sample size selection and choice of rating periods. These in principle can be addressed under the normative approach the control theorists have been doing itbut with the added baggage of specifying the appropriate utilities and the optimization over several decision variables. Ironically, it is Dr. Fries, who with Dr. McGrath, while being in favor of the idea of controlling the indices, question if the methodology can be put to actual practice. Both Drs. Fries and McGrath interact with industry and so their concern about implementation is taken seriously. But there may also be an opportunity here, because, by virtue of their positions, they can put into practice the theme epoused by Professor Lindley, who claims that advances in computer technology can put decision theory at the hands of users who need not know its detailed inner workings.

Many of the discussants have, in one form or the other, commented on the underlying utilities; Professor Kotz dismisses them as being too simplistic! Professor Cox suggests an expansion of the utilities to include customer goodwill; this consideration requires thinking about utilities over several future periods rather than just the next period as is done in the paper. This suggestion also meshes with Professor Kliemann's who advocates notion of "dynamic utilities". Dr. Fries' comment that in practice intervention is based on a pattern of progressive worsening, rather that just a violation of th specification interval, also suggests an elaboration of the utility function. However, there is a dilemma here because industrial practice simultaneously requires a consideration of several caveats, and then demands that the proposed procedures be easy to implement. These requirements are conflicting; however, they can now be harnessed with the assistance of modern computing. Professor Gómez-Villegas raises the matter of obtaining a credible set of coverage $(1-\alpha)$, and then investigating if the specification interval is contained in it. Such an approach would be normative only if utilities can be mapped into probabilities; i.e. if we could meaningfully relate $(1-\alpha)$ with our $(b/a - Q/aN)$. I do not see a natural way to make this connection. On the matter of robustness concerns have been expressed by Professors Cox, Dey, Ghosh, Gómez–Villegas, and Kotz. Professor Kotz is wondering if the Gaussian assumption is key to the paper. However, as Professors Dey and Kliemann show, the methodology easily extends to alternatives. Indeed Professor Dey considers several, to include non-symmetric, logistic, symmetric stable, and the exponential power family. Though not directly so, Professor Dey also brings in the role of computer based sampling approaches in process capability analyses. Like control, robustness mandates the involvement of computer technology for process capability assessment. Professor Gomez-Villegas' robustification of the control model is a natural next step. Besides robustness, Professor Ghosh raises several other issues which I will now attempt to address. Change points in the process, if anticipated in advance can be incorporated in the normative approach either via the system equation (3.2), or via the assumption of the joint distribution of the X_i's. The assumption of process stability is *not essential* in the normative approach, and this is one of its virtues. A use of the normal (Gaussian) model and conjugate prior is implicit in the assumption that the posterior distribution of X, (with or without a controller), is a Gaussian. I am in agreement with Professor Ghosh, that the arbitrary r adds no insight. Its purpose however, is to relate the Bayes capability index with the traditional indices; as such it may be viewed as a scaling parameter. I acknowledge, with some embarrassment, the error leading to (2.12) [(2.11) in Professor Ghosh's discussion]. This error was also noted by Professors Lindley (in his cover letter to me), and Kliemann, who has shown see his discussion that

$$\text{choose } d_1 \quad \text{iff} \quad C_p \overset{\text{def}}{=} \frac{\text{USL} - \text{LSL}}{6\sigma} \geq \frac{1}{6}\left[\Phi_L^{-1}\left(\frac{bN - Q^*}{aN}\right) - \text{LSL}\right].$$

A matter with which I take exception is Professor Ghosh's statement about "mis-specification of prior". This, it appears, is a contradiction; a subjective Bayesian endeavors to specify an honest prior. Robustness studies are still germane because different individuals may have different priors, and yet all could arrive at the same decision. Interestingly, Dr. Fries, asks "What is truly Bayesian about the Bayes Capability Index? Subjective probability does not appear to enter ..." This is really a question for Bernardo and Irony to address, but my view is that in making the assumptions of exchangeability and Gaussianity, subjective considerations were involved. Furthermore, it is not so that all Bayesians take a subjective view of probability; personally, I do. In taking the normative view for decision making one subscribes to the Bayesian paradigm, and hence the index proposed by Bernardo and Irony is legitimately Bayesian. Related to the matter of robustness is the question raised by Dr. Fries about conditioning on the posterior distribution. Should we condition on the entire past or just the previous past? This is a good point because conditioning on the entire past will make the procedure lethargic to sudden changes in the process mean;

that is, it will be slow to respond to a shift in the process. Alternatively, conditioning on the previous past will make the procedure overly sensitive to the slightest change. Robustifying the model using heavy tailed distributions has been the traditional approach for addressing this issue. Incorporating the controller could be another. Combining the two may provide the most economical approach, but his matter can only be resolved in the context of a specific application, involving specific utilities. Clearly the point raised by Dr. Fries is one that needs more attention.

Lastly, on the matter of incoherence, Dr. Fries inquires if there exist any (the emphasis is his) utilities for which the traditional process capability indices are coherent, and Professor Irony attempts to argue incoherence away via limiting considerations. I remain to be convinced of the latter, but as far as the former is concerned the incoherence argument rests on the principle that it takes only one counter-example to disprove a theorem. The same argument holds with respect to Dr. McGrath's attempts at resolving the incoherence. Dr. McGrath also invokes the "Rolled Throughput Assembly Yield" practice for justifying the dictum of Six Sigma quality goal. It should be recognized that the engineering practice of assuming independence to obtain $Y_{RT} = \prod_1^n P(X_i \in A \mid H_i)$ is *conservative*, if the X_i's are positively dependent (e.g. exchangeable), so that the Six Sigma quality goal is desirable for the consumer so long as the costs of manufacturing to this level of quality are not passed on to the consumer. Indeed, the practical rationale for using a normative approach for developing process capability indices is an explicit consideration of costs via profits and penalties. It is also useful to bear in mind that the procedures proposed here pertain to the *manufacturer's decisions*. It is the manufacturer's utilities and probabilities that are considered. The adversarial nature of the consumer manufacturer relationship alluded to by Dr. McGrath has been treated elsewhere [cf. Lindley and Singpurwalla (1991, 1993)]. Professor Kliemann denies incoherence of the traditional indices on the grounds that none of the criteria require ba for adopting decision d1. However, the argument of incoherence is based on a reverse of the above logic. For example, if C_p (or C_{pk}) is taken to be one or greater, then the normative criteria lead to the conclusion that ba, which cannot happen because $b = ah$, and h_0. Whereas I am not in agreement with Professor Kliemann's position on incoherence, I do applaud his introduction of the new indices C_B^* and C_p^* and their exact relationship to C_{pk} and C_p. Incidentally, I am wondering if the right hand sides of Kliemann's formulae (3a) and (4a) should involve the terms $\Phi^{-1}_{\frac{L-\mu}{\sigma}}(B)$ instead of the quoted $\Phi^{-1}_{L}(B)$?

Closing Comments

When Professor Bernardo first contacted the discussants, some more than the number who have responded, I felt like being thrown in the lion's den. This was especially so, because many of the discussants have a lion-like stature. The thought of responding to each seemed Herculean. However, it so happened that a pattern of comments evolved. More important, the issues raised by one discussant were addressed by another, reducing my task to that of simply synthesizing the discussion. Much of the discussion pertained to the laudatory comments on Professors Bernardo and Irony's pioneering paper. It is gratifying to have been the agent for providing a platform of visibility to this work.

The discussion has also drawn attention to two conflicting principles that drive industrial practice. The first being the need to account for as many caveats as practitioners can possibly think of, and the second being the desired to have procedures that are easy to implement. The Six Sigma quality goal is in full accord with the simplicity principle, and appears to be the rave of modern manufacturing [cf. Conlin (1998)]. However, it nowhere comes close to the first principle; indeed, the normative argument, which incorporates costs, shows that the Six Sigma goal could lead to incoherence. The solution therefore lies with computer technology and the computational sciences. These not only provide the needed tools for normative decision making, as Professor Lindley asserts, but can also facilitate a consideration of computer intensive statistical techniques as Professor Dey advocates. Finally, I can't resist but say a few words about my friend Professor Sam Kotz. Surely, he is incoherent! Who else would write the first paragraph of his discussion? But then perhaps, as he quotes "incoherence is an inevitable property of many useful and appropriate summary indices, and occurs in various natural situations". Is he implying that even God is incoherent?

References

Conlin, M. (1998). Revealed at last: the secret of Jack Welch's success. *Forbes*, January 26, 1998, p. 44.

Lindley, D. V. and N. D. Singpurwalla (1991). On the evidence needed to reach agreed action between adversaries with application to acceptance sampling. *Journal of the American Statistical Association*, **86**, 416, 933-937.

Lindley, D. V. and N. D. Singpurwalla (1993). Adversarial life testing. *Journal of the Royal Statistical Society, Series B*, **55**, 4, 837-847.