Journal of Statistical Computation and Simulation
Vol. 75, No. 3, March 2005, 175–187

A process capability index for discrete processes

MICHAEL PERAKIS and EVDOKIA XEKALAKI*

Department of Statistics, Athens University of Economics and Business,
76 Patision St, 104 34 Athens, Greece

(Revised 27 November 2003; in final form 9 February 2004)

Perakis and Xekalaki 2002, A process capability index that is based on the proportion of conformance. *Journal of Statistical Computation and Simulation*, **72**(9), 707–718. introduced a process capability index that is based on the proportion of conformance of the process under study and has several appealing features. One of its advantages is that it can be used not only for continuous processes, as is the case with the majority of the indices considered in the literature, but also for discrete processes as well. In this article, the use of this index is investigated for discrete data under two alternative models, which are frequently considered in statistical process control. In particular, distributional properties and estimation of the index are considered for Poisson processes and for processes resulting in modeling attribute data. The performance of the suggested estimators and confidence limits is tested via simulation.

Keywords: Process capability indices; Proportion of conformance; Approximate confidence limits; Simulation study; Poisson distribution; Attribute data

1. Introduction

Measuring the capability of a process to produce according to some specifications connected to a measurable characteristic X of the produced items, being of great importance to industrial research, has motivated much work. Following the paper by Kane [1], more articles have appeared introducing new indices or studying the properties of existing ones. Excellent reviews on them are given by Kotz and Johnson [2, 3] and Kotz and Lovelace [4]. In addition, Spiring *et al.* [5] provide an extensive bibliography on process capability indices.

The vast majority of the process capability indices that have been considered are associated only with processes that can be described through some continuous and, in particular, normally distributed characteristics. The most widely used such indices are C_{p}, C_{pk}, C_{pm}, and C_{pmk} or their generalizations for non-normal processes, suggested by Clements [6], Pearn and Kotz [7], and Pearn and Chen [8]. Often, however, one is faced with processes described by a characteristic whose values are discrete. Therefore, in such cases none of these indices can

*Corresponding author. Email: exek@aueb.gr

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online © 2005 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00949650410001687244

be used. To our knowledge, the only indices suggested so far whose assessment is meaningful regardless of whether the studied process is discrete or continuous are those suggested by Yeh and Bhattacharya [9], Borges and Ho [10], and Perakis and Xekalaki [11].

A very common example of a process that can be described through a discrete-valued characteristic is the number of defects per produced unit by an industry. In this case, it is obvious that small process values are desirable. In some other cases, however, large values may be desirable. In what follows, we consider only discrete processes with unilateral tolerances (*i.e.* processes connected with just one specification limit, lower (L) or upper (U)), since in practice, discrete processes with bilateral tolerances are only rarely considered. Of course, extending the results obtained below to the case of bilateral tolerances is possible but tedious and requires a much more complicated analysis.

As already mentioned, Perakis and Xekalaki [11] proposed a new index, which can be used regardless of whether the examined process is discrete or continuous. This index is defined as

$$C_{\mathrm{pc}} = \frac{1 - p_0}{1 - p},$$

where p and p_0 denote the proportion of conformance (yield) and the minimum allowable proportion of conformance of the examined process, respectively. As is well known, the term proportion of conformance refers to the probability of producing within the so-called specification area, *i.e.* the interval determined by L and U. If the tolerances are unilateral, then the value of p is given by $P(X > L)$, if only L has been set, and by $P(X < U)$, if only U has been assigned. According to Perakis and Xekalaki [11], the value 0.9973 is a plausible choice for p_0, since this probability is usually regarded as sufficiently large in statistical process control. However, different choices of p_0 can be made, according to the nature of the process examined. In the sequel, the value 0.9973 is chosen. Nevertheless, the analysis given in the sequel can be readily modified for any other choice of p_0. The properties of C_{pc} in the case where the distribution of the process studied is normal or exponential are investigated thoroughly by Perakis and Xekalaki [11].

In this article, the properties of the index C_{pc} are studied, for two distributional assumptions that can be considered in the context of discrete processes. At first, we deal with the case where the underlying distribution of the examined process is the Poisson distribution. Then, we consider the case of processes whose produced items are not measured with respect to one of their characteristics, but are instead classified in two categories. The first category consists of the conforming items labeled by the value 1, while the second consists of the non-conforming items labeled by the value 0. Data of this type are known as attribute data (some examples of such processes can be found in refs. [12, 13]).

In particular, in section 2, the index C_{pc} is defined for a Poisson-distributed process. Section 3 deals with the estimation of the index and the construction of confidence limits for its true value under this assumption, while section 4 presents the results of a simulation study that was conducted to examine the performance of the estimators and the confidence limits defined in section 3. Section 5 is devoted to the use of the index C_{pc} in connection with attribute data. Finally, in section 6, some conclusions and points that may become the issue of further research are given.

2. The index C_{pc} for Poisson processes

As already mentioned, one of the advantages of the index C_{pc} is that its assessment is possible even if the examined process is discrete. In this section, the properties of C_{pc} are examined in the case where the studied process is described by a Poisson-distributed characteristic with some parameter $\lambda > 0$. In order to avoid confusion, the following notation is adopted: if only the value of U has been set, the index is denoted by C_{pcu}, while if only the value of L has been set, the index is denoted by $C_{\mathrm{pc}\ell}$. Thus, the index C_{pcu} is defined as

$$C_{\mathrm{pcu}} = \frac{0.0027}{1-p},$$

where $p = P(X < U)$ and the index $C_{\mathrm{pc}\ell}$ is defined as

$$C_{\mathrm{pc}\ell} = \frac{0.0027}{1-p},$$

with p given by $p = P(X > L)$. Note that the values U and L are assumed to lie outside the specification area.

The probability p that is involved in the denominator of the index C_{pcu} is equal to the sum

$$\sum_{x=0}^{U-1} \frac{\mathrm{e}^{-\lambda}\lambda^{x}}{x!}.$$

Note that p is the cumulative distribution function of the Poisson distribution with parameter $\lambda > 0$ evaluated at $U - 1$. As is well known [see ref. 14]

$$\begin{aligned} p &= P(X \le U-1) \\ &= P(\chi^2_{2U} > 2\lambda), \end{aligned}$$

where χ^2_{2U} denotes the chi-square distribution with $2U$ degrees of freedom. Using this property, the index C_{pcu} can be written as

$$C_{\mathrm{pcu}} = \frac{0.0027}{P(\chi^2_{2U} < 2\lambda)}.$$

Similarly, one may observe that the value of p that appears in the denominator of the index $C_{\mathrm{pc}\ell}$ can be written in the form

$$\begin{aligned} p &= 1 - P(X \le L) \\ &= 1 - P(\chi^2_{2(L+1)} > 2\lambda) \\ &= P(\chi^2_{2(L+1)} < 2\lambda) \end{aligned}$$

and thus

$$C_{\mathrm{pc}\ell} = \frac{0.0027}{P(\chi^2_{2(L+1)} > 2\lambda)}.$$

It would be interesting to remark that by their definition, the indices C_{pcu} and $C_{\mathrm{pc}\ell}$ are related directly to the proportion of conformance of the process. This rather interesting property constitutes, undoubtedly, an appealing feature lacked by all the most frequently used process capability indices, such as C_{p}, C_{pk}, C_{pm}, and C_{pmk}.

3. Estimation

If the value of λ is unknown (as is often the case), it has to be replaced by an estimate of it. As is well known, the maximum likelihood estimator of the parameter λ, provided that n independent observations $(X_1, X_2, \ldots, X_n)$ from the Poisson distribution with parameter λ are available, is the sample mean, *i.e.* $\hat{\lambda} = \bar{X}$. Hence, point estimators of the indices C_{pcu} and $C_{\mathrm{pc}\ell}$ can be obtained replacing λ by $\hat{\lambda}$. The resulting estimators are given by

$$\hat{C}_{\mathrm{pcu}} = \frac{0.0027}{P(\chi^2_{2U} < 2\hat{\lambda})} \tag{1}$$

and

$$\hat{C}_{\mathrm{pc}\ell} = \frac{0.0027}{P(\chi^2_{2(L+1)} > 2\hat{\lambda})}, \tag{2}$$

respectively.

Alternative estimators of C_{pcu} and $C_{\mathrm{pc}\ell}$ can be obtained using the minimum variance unbiased estimator of the proportion of conformance. Actually, it can be shown [see ref. 15] that the minimum variance unbiased estimator of $p = P(X < U)$, provided that the distribution of the process is the Poisson, is given by

$$\tilde{p}_1 = \begin{cases} 1, & \text{if } U - 1 > Y = \sum_{i=1}^{n} X_i, \\ \sum_{x=0}^{U-1} \binom{Y}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{Y-x}, & \text{if } 0 \le U - 1 \le Y. \end{cases}$$

Hence, an alternative estimator of the index C_{pcu} is given by

$$\tilde{C}_{\mathrm{pcu}} = \frac{0.0027}{1 - \tilde{p}_1}. \tag{3}$$

It should be noted that the computational formula of $\tilde{p}_1$ in the case where $0 \le U - 1 \le Y$ coincides with the cumulative distribution function of the binomial distribution with parameters Y and n^{-1}, evaluated at the point $U - 1$ and its assessment may be simplified considerably by resorting to the relationship between the binomial and the F distribution. Indeed, it is known [see ref. 14] that

$$\sum_{x=r}^{n} \binom{n}{x} p^x (1-p)^{n-x} = P\left(F < \frac{\nu_2 p}{\nu_1 (1-p)}\right),$$

where F denotes a random variable that follows the F distribution with $\nu_1 = 2r$ and $\nu_2 = 2(n - r + 1)$ degrees of freedom. Using this result, one obtains that

$$\begin{aligned} \sum_{x=0}^{U-1} \binom{Y}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{Y-x} &= 1 - \sum_{x=U}^{Y} \binom{Y}{X} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{Y-x} \\ &= 1 - P\left(F < \frac{\nu_2/n}{\nu_1(1 - 1/n)}\right) \\ &= P\left(F > \frac{\nu_2}{\nu_1(n-1)}\right), \end{aligned}$$

where $\nu_1 = 2U$ and $\nu_2 = 2(Y - U + 1)$.

Proceeding as in the case of the index C_{pcu}, one may find that the value of the index $C_{\text{pc}\ell}$ can be alternatively estimated by

$$\tilde{C}_{\text{pc}\ell} = \frac{0.0027}{\tilde{p}_2}, \tag{4}$$

where $\tilde{p}_2$ is defined as

$$\tilde{p}_2 = \begin{cases} 1, & \text{if } L > Y = \sum_{i=1}^{n} X_i, \\ \sum_{x=0}^{L} \begin{pmatrix} Y \\ X \end{pmatrix} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{Y-x}, & \text{if } 0 \leq L \leq Y. \end{cases}.$$

In the analysis given so far, we have dealt only with the definition of some point estimators of the true value of the indices C_{pcu} and $C_{\text{pc}\ell}$. However, mere knowledge of their values does not suffice for measuring the capability of a process. It would therefore be useful to proceed to the construction of lower confidence limits for their true values as well.

If a random sample of n observations, $X_1, X_2, \ldots, X_n$, from the Poisson distribution with parameter λ is available, it is known that the sum

$$Y = \sum_{i=1}^{n} X_i$$

is also Poisson distributed with parameter $n\lambda$. Thus, a $100(1-\alpha)\%$ approximate upper confidence limit for $n\lambda$ is given by

$$0.5\chi^2_{2(Y+1),1-\alpha},$$

where $\chi^2_{2(Y+1),1-\alpha}$ denotes the $1-\alpha$ quantile of the chi-square distribution with $2(Y+1)$ degrees of freedom [see ref. 14]. Likewise, a $100(1-\alpha)\%$ approximate upper confidence limit for 2λ can be obtained using

$$P(n\lambda < 0.5\chi^2_{2(Y+1),1-\alpha}) = 1 - \alpha$$

or, equivalently,

$$P\left(2\lambda < \frac{\chi^2_{2(Y+1),1-\alpha}}{n}\right) = 1 - \alpha. \tag{5}$$

It is known [see ref. 16, p. 378] that if a $100(1-\alpha)\%$ confidence interval for a parameter θ is given by (a, b), then a $100(1-\alpha)\%$ confidence interval for $g(\theta)$, where $g(\cdot)$ is a strictly monotone function, is given by $(g(a), g(b))$. Considering that $P(\chi^2_{2U} < x)$, being a cumulative distribution function, is a strictly monotone function, we deduce that equation (5) can be rewritten as

$$P\left\{P(\chi^2_{2U} < 2\lambda) < P\left(\chi^2_{2U} < \frac{\chi^2_{2(Y+1),1-\alpha}}{n}\right)\right\} = 1 - \alpha$$

or, equivalently, as

$$P\left(\frac{00027}{P(\chi^2_{2U} < 2\lambda)} > \frac{0.0027}{P(\chi^2_{2U} < \chi^2_{2(Y+1),1-\alpha}/n)}\right) = 1-\alpha.$$

Hence, an approximate $100(1-\alpha)\%$ lower confidence limit for the true value of the index $C_{\mathrm{pc}\ell}$ is given by

$$\frac{0.0027}{P(\chi^2_{2U} < \chi^2_{2(Y+1),1-\alpha}/n)}. \tag{6}$$

Similarly, in order to find a $100(1-\alpha)\%$ lower confidence limit for the true value of the index C_{pcu}, one may be based on the result that the quantity

$$0.5\chi^2_{2Y,\alpha}$$

constitutes a $100(1-\alpha)\%$ approximate lower confidence limit for $n\lambda$ [see ref. 14]. Therefore,

$$P\left(2\lambda > \frac{\chi^2_{2Y,\alpha}}{n}\right) = 1-\alpha,$$

or, equivalently

$$P\left\{P(\chi^2_{2(L+1)} < 2\lambda) > P\left(\chi^2_{2(L+1)} < \frac{\chi^2_{2Y,\alpha}}{n}\right)\right\} = 1-\alpha.$$

This in turn leads to

$$P\left(\frac{0.0027}{1-P(\chi^2_{2(L+1)} < 2\lambda)} > \frac{0.0027}{1-P\left(\chi^2_{2(L+1)} < \frac{\chi^2_{2Y,\alpha}}{n}\right)}\right) = 1-\alpha.$$

Thus, an approximate $100(1-\alpha)\%$ lower confidence limit for the true value of the index $C_{\mathrm{pc}\ell}$ is given by

$$\frac{0.0027}{1-P(\chi^2_{2(L+1)} < \chi^2_{2Y,\alpha}/n)}. \tag{7}$$

4. A simulation study

In order to test the performance of estimators (1) and (3) and the observed coverage of lower confidence limit (6) for the index C_{pcu}, a simulation study was conducted. It should be noted that by their definition estimators (2) and (4) and lower confidence limit (7) are expected to have similar performances and thus the conclusions stated below can also be extended to them.

In the conducted simulation study 50,000 random samples were generated from the Poisson distribution for various values of the parameter λ, five different sample sizes (25, 50, 100, 200, and 400) and two alternative values of U (a relatively small value (5) and a larger one (20)) so as to detect the influence of all these factors on the behavior of the two estimators.

Tables 1 and 2 summarize the obtained results. In particular, table 1 corresponds to the case where $U = 5$, while table 2 corresponds to the case where $U = 20$. Their entries are as follows.

Table 1. The results of the simulation study for $U = 5$.

λ	C_{pcu}	p		Sample size				
				$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
0.7	3.4371	0.9992	(1)	8.7235	5.1668	4.1568	3.7443	3.5867
			(2)	29.7184	6.9289	4.6916	3.9592	3.6847
			(3)	0.9340	0.9025	0.9186	0.9098	0.9022
			(4)	0.9632	0.9512	0.9506	0.9551	0.9527
			(5)	1.4638	1.7256	2.0075	2.2744	2.5349
			(6)	1.0036	1.3289	1.6700	1.9958	2.3095
0.9	1.1518	0.9977	(1)	2.3040	1.5341	1.3239	1.2306	1.1902
			(2)	4.4669	1.8560	1.4380	1.2794	1.2129
			(3)	0.8991	0.8999	0.9060	0.9064	0.9018
			(4)	0.9612	0.9619	0.9521	0.9551	0.9516
			(5)	0.5437	0.6218	0.7199	0.8081	0.8873
			(6)	0.3960	0.4993	0.6161	0.7232	0.8198
1.1	0.4968	0.9946	(1)	0.8110	0.6184	0.5505	0.5225	0.5106
			(2)	1.1801	0.7074	0.5845	0.5376	0.5177
			(3)	0.9137	0.9023	0.9046	0.9069	0.9053
			(4)	0.9637	0.9590	0.9559	0.9524	0.9513
			(5)	0.2470	0.2860	0.3257	0.3630	0.3957
			(6)	0.1887	0.2366	0.2846	0.3296	0.3693
1.3	0.2532	0.9893	(1)	0.3718	0.2990	0.2744	0.2640	0.2585
			(2)	0.4803	0.3310	0.2869	0.2697	0.2612
			(3)	0.9213	0.9076	0.9071	0.9084	0.9009
			(4)	0.9637	0.9585	0.9518	0.9511	0.9497
			(5)	0.1338	0.1527	0.1731	0.1915	0.2064
			(6)	0.1058	0.1293	0.1536	0.1759	0.1942
1.5	0.1453	0.9814	(1)	0.1969	0.1667	0.1555	0.1501	0.1478
			(2)	0.2375	0.1797	0.1609	0.1526	0.1490
			(3)	0.9064	0.9114	0.9129	0.9056	0.9026
			(4)	0.9593	0.9579	0.9553	0.9521	0.9532
			(5)	0.0806	0.0919	0.1032	0.1128	0.1209
			(6)	0.0655	0.0793	0.0928	0.1045	0.1145
1.7	0.0912	0.9704	(1)	0.1159	0.1022	0.0962	0.0937	0.0925
			(2)	0.1333	0.1083	0.0988	0.0949	0.0931
			(3)	0.9214	0.9154	0.9126	0.9096	0.8993
			(4)	0.9614	0.9578	0.9527	0.9520	0.9528
			(5)	0.0527	0.0600	0.0666	0.0725	0.0772
			(6)	0.0438	0.0526	0.0606	0.0677	0.0735
1.9	0.0613	0.9559	(1)	0.0746	0.0674	0.0641	0.0627	0.0619
			(2)	0.0830	0.0705	0.0655	0.0634	0.0622
			(3)	0.9089	0.9024	0.9113	0.9042	0.8998
			(4)	0.9650	0.9565	0.9503	0.9534	0.9511
			(5)	0.0368	0.0417	0.0460	0.0497	0.0526
			(6)	0.0312	0.0370	0.0422	0.0467	0.0503
2.1	0.0435	0.9379	(1)	0.0514	0.0471	0.0453	0.0443	0.0439
			(2)	0.0558	0.0488	0.0460	0.0446	0.0441
			(3)	0.9209	0.9094	0.8990	0.9026	0.9027
			(4)	0.9562	0.9601	0.9556	0.9490	0.9493
			(5)	0.0272	0.0305	0.0335	0.0358	0.0378
			(6)	0.0234	0.0274	0.0310	0.0339	0.0363

Table 2. The results of the simulation study for $U = 20$.

λ	C_{pcu}	p		Sample size $n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
8	10.67	0.9997	(1)	17.3033	13.4537	11.8845	11.2921	10.9865
			(2)	26.8773	16.3278	13.0195	11.8048	11.2298
			(3)	0.9075	0.9088	0.9048	0.9014	0.9002
			(4)	0.9537	0.9558	0.9534	0.9527	0.9510
			(5)	5.1100	5.8581	6.6689	7.5273	8.2563
			(6)	3.7346	4.6987	5.7032	6.7342	7.6278
9	2.557	0.9989	(1)	3.6990	3.0711	2.7960	2.6742	2.6136
			(2)	5.1091	3.5480	2.9939	2.7647	2.6559
			(3)	0.9061	0.9026	0.8997	0.8996	0.9012
			(4)	0.9517	0.9537	0.9505	0.9494	0.9497
			(5)	1.2887	1.4888	1.6880	1.8760	2.0356
			(6)	0.9828	1.2288	1.4726	1.7021	1.8995
10	0.782	0.9965	(1)	1.0494	0.8949	0.8382	0.8078	0.7940
			(2)	1.3339	0.9962	0.8821	0.8282	0.8039
			(3)	0.9025	0.9091	0.8997	0.9007	0.9025
			(4)	0.9549	0.9505	0.9513	0.9495	0.9526
			(5)	0.4192	0.4762	0.5392	0.5923	0.6379
			(6)	0.3314	0.4030	0.4786	0.5440	0.6004
11	0.291	0.9907	(1)	0.3657	0.3240	0.3066	0.2989	0.2942
			(2)	0.4363	0.3508	0.3184	0.3045	0.2970
			(3)	0.9034	0.9049	0.9018	0.9007	0.9032
			(4)	0.9530	0.9521	0.9509	0.9508	0.9522
			(5)	0.1648	0.1868	0.2084	0.2278	0.2429
			(6)	0.1344	0.1615	0.1879	0.2115	0.2303
12	0.127	0.9787	(1)	0.1517	0.1384	0.1326	0.1296	0.1282
			(2)	0.1726	0.1467	0.1364	0.1314	0.1291
			(3)	0.9054	0.9028	0.8993	0.9039	0.9019
			(4)	0.9526	0.9520	0.9509	0.9518	0.9517
			(5)	0.0759	0.0856	0.0947	0.1022	0.1084
			(6)	0.0637	0.0755	0.0865	0.0958	0.1035
13	0.063	0.9573	(1)	0.0728	0.0677	0.0653	0.0644	0.0638
			(2)	0.0798	0.0706	0.0666	0.0650	0.0641
			(3)	0.9059	0.9084	0.9041	0.9012	0.9019
			(4)	0.9509	0.9534	0.9532	0.9522	0.9501
			(5)	0.0400	0.0446	0.0487	0.0523	0.0551
			(6)	0.0344	0.0400	0.0450	0.0495	0.0530
14	0.035	0.9235	(1)	0.0395	0.0373	0.0363	0.0358	0.0355
			(2)	0.0422	0.0384	0.0368	0.0360	0.0356
			(3)	0.9042	0.9018	0.9043	0.9028	0.9018
			(4)	0.9492	0.9488	0.9514	0.9523	0.9515
			(5)	0.0236	0.0260	0.0281	0.0299	0.0313
			(6)	0.0207	0.0237	0.0263	0.0285	0.0302
15	0.022	0.8752	(1)	0.0236	0.0226	0.0221	0.0219	0.0218
			(2)	0.0247	0.0230	0.0223	0.0220	0.0218
			(3)	0.9068	0.9037	0.9039	0.9008	0.8986
			(4)	0.9557	0.9520	0.9489	0.9511	0.9492
			(5)	0.0152	0.0165	0.0178	0.0187	0.0195
			(6)	0.0136	0.0153	0.0168	0.0180	0.0189

(1) The mean of the values of estimator (1).
(2) The mean of the values of estimator (3).
(3) The observed coverage of 90% lower confidence limit (6).
(4) The observed coverage of 95% lower confidence limit (6).
(5) The mean of the 90% lower confidence limit (6).
(6) The mean of the 95% lower confidence limit (6).

From tables 1 and 2, one can observe that:

- The mean of estimator (1) is always closer to the actual value of the index than the mean of estimator (3).
- The means of both estimators tend to approach the actual value of C_{pcu} as n increases.
- The difference between the means of estimators (1) and (3) diminishes as the sample size increases.
- Both estimators seem to overestimate the actual value of C_{pcu}, especially for small samples.
- The coverage achieved by lower confidence limit (6) seems to be quite close to the nominal for all the studied cases, especially when the sample size becomes large enough.

5. The index C_{pc} for attribute data

This section considers the use, properties, and estimation of the index C_{pc} for measuring the capability of a process in terms of qualitative aspects and, in particular, on the basis of data on the numbers of items produced by it that conform to certain standards (binary data on conforming/non-conforming items). In the sequel, two different cases are considered. In the first case, the specifications are given solely in terms of the minimum acceptable proportion of conformance, while in the second case, the specifications are given in terms of the number of defective items per package.

5.1 *Specifications in terms of the minimum acceptable proportion of conformance*

Let us assume that the examined process produces items, each of which can be regarded either as conforming (1) or as non-conforming (0), and that the specifications of this process consist solely of p_0, the minimum acceptable proportion of conformance. Obviously, for such types of data, the specifications do not have the usual form (*i.e.* the form L, U, T).

Assuming for simplicity as before that the minimum acceptable proportion of conformance is equal to 0.9973, the index C_{pc} is defined as

$$C_{pc} = \frac{0.0027}{1 - P(X = 1)}. \tag{8}$$

From equation (8) one may observe that in this case the proportion of conformance of the process is given by $p = P(X = 1)$.

We now consider the estimation of the index C_{pc}. The only unknown parameter involved in the expression of this index is the proportion of conformance of the process, *i.e.* $p = P(X = 1)$. Let us assume that n independent observations $X_1, \ldots, X_n$ from the examined process are available. Each of these observations constitutes a realization of a Bernoulli random variable with common parameter p (provided that the process is in statistical control so as to be assumed that the value of p remains unchanged). It is known [see ref. 14] that in such cases the maximum likelihood and the moment estimator of p coincide and are given by

$$\hat{p} = \frac{Y}{n}, \tag{9}$$

where $Y = \sum_{i=1}^{n} X_i$. Note that the estimator given in equation (9) is at the same time the minimum variance unbiased estimator of p, since it is unbiased and its variance coincides with the lower bound determined by the Cramer–Rao inequality [see ref. 14].

A point estimator of the index C_{pc} can be obtained by replacing the proportion of conformance in equation (8) by its estimator given in equation (9). Consequently, the resulting estimator is defined as

$$\hat{C}_{\text{pc}} = \frac{0.0027}{1 - \hat{p}}. \tag{10}$$

The existing theory on the estimation of p may also be used as a basis for the construction of confidence limits for the true value of the index C_{pc}. Actually, it is known that a $100(1-\alpha)\%$ confidence interval for the actual value of p is given by

$$(p_L, p_U),$$

where the values of the limits p_L and p_U are determined through the formulae

$$p_L = \frac{\nu_1 F_{\nu_1,\nu_2,\alpha/2}}{\nu_2 + \nu_1 F_{\nu_1,\nu_2,\alpha/2}} \tag{11}$$

and

$$p_U = \frac{\nu_3 F_{\nu_3,\nu_4,1-\alpha/2}}{\nu_4 + \nu_3 F_{\nu_3,\nu_4,1-\alpha/2}}, \tag{12}$$

respectively [see ref. 14]. Note that in equations (11) and (12), $F_{\nu_1,\nu_2,\alpha}$ denotes the $100\alpha\%$ quantile of the F distribution with ν_1 and ν_2 degrees of freedom. Furthermore, the values of ν_1, ν_2, ν_3, and ν_4 that appear in equations (11) and (12) are given by $2Y$, $2(n-Y+1)$, $2(Y+1)$, and $2(n-Y)$, respectively.

The limits given in equations (11) and (12) can also be used in order to obtain a confidence interval for C_{pc}. Actually, considering the fact that p_L and p_U constitute the limits of a $100(1-\alpha)\%$ confidence interval for p, it follows that

$$P(p_L < p < p_U) = 1 - \alpha,$$

or, equivalently,

$$P\left(\frac{0.0027}{1-p_L} < \frac{0.0027}{1-p} < \frac{0.0027}{1-p_U}\right) = 1 - \alpha.$$

Hence, a $100(1-\alpha)\%$ confidence interval for the actual value of the index C_{pc} is given by

$$\left(\frac{0.0027}{1-p_L}, \frac{0.0027}{1-p_U}\right). \tag{13}$$

Similarly, a $100(1-\alpha)\%$ lower confidence limit for the true value of C_{pc} can be obtained from

$$\left(\frac{0.0027}{1-p'_L}\right),$$

where

$$p'_L = \frac{\nu_1 F_{\nu_1,\nu_2,\alpha}}{\nu_2 + \nu_1 F_{\nu_1,\nu_2,\alpha}}. \tag{14}$$

The values of ν_1 and ν_2 in equation (14) are defined as in the case of equation (11).

If the sample size is sufficiently large and the value of p is not very close to 0 or 1, whence the binomial distribution can be approximated by the normal distribution, approximate confidence intervals for p (and, consequently, for C_{pc}) can be constructed using the percentiles of the standard normal distribution. However, since in the analysis of the capability of industrial processes the value of p is often expected to be fairly large, the normal approximation should be avoided, even for large samples.

Finally, it should be remarked that the analysis given above generally requires large samples, as a consequence of the fact that in most of the cases the value of p is large and thus the probability of observing some defective items in a small sample becomes negligible.

5.2 *Specifications in terms of the number of defective items per package*

In this section, the analysis of the capability of processes connected with attribute data in the case where the items produced from the process are packed in similar boxes (packages), each of which contains n items, is considered. In such cases, the specifications of the process are not necessarily assigned in terms of the minimum acceptable proportion of conformance, as previously, but can be assigned in terms of the minimum allowable number of conforming items that each box should contain, as well. Therefore, if each box is supposed to contain more than L conforming items, the proportion of conformance of the process is equal to $p = P(X > L)$, and thus, the index C_{pc} is defined as

$$C_{\text{pc}} = \frac{0.0027}{1 - P(X > L)}.$$

Obviously, if the studied process is in statistical control, the probability of having a non-defective item is fixed. Denoting this probability by q, the number of conforming items contained in each box follows the binomial distribution with parameters n (known) and q (unknown). Hence, the proportion of conformance of the process is given by

$$\begin{aligned} p &= P(X > L) \\ &= \sum_{x=L+1}^{n} \binom{n}{x} q^x (1-q)^{n-x} \\ &= 1 - \sum_{x=0}^{L} \binom{n}{x} q^x (1-q)^{n-x}. \end{aligned} \tag{15}$$

For the estimation of p, one may estimate the value of q, involved in equation (15) taking its maximum likelihood estimator or using the minimum variance unbiased estimator of the cumulative distribution function of the binomial distribution [see ref. 15].

Thus, if a random sample of k boxes of items produced from the process has been collected and X_i denotes the number of conforming items contained in the ith box, $i = 1, \ldots, k$, the maximum likelihood estimator of the parameter q is given by

$$\hat{q} = \frac{T}{kn},$$

where

$$T = \sum_{i=1}^{k} X_i. \tag{16}$$

Thus, a point estimate of C_{pc} can be obtained substituting $\hat{q}$ in equation (15) and using the obtained result instead of p in the formula of the index C_{pc}.

Alternatively, the minimum variance unbiased estimator of p is given by

$$\tilde{p} = \begin{cases} 0, & \text{if } L > T, \\ 1 - \sum_{x=0}^{L} \dfrac{\binom{n}{x}\binom{nk-n}{T-x}}{\binom{nk}{T}}, & \text{if } L \leq T, \end{cases}$$

where T is defined as in equation (16) and leads to the following estimator of the index C_{pc}:

$$\tilde{C}_{\text{pc}} = \frac{0.0027}{1 - \tilde{p}}$$

6. Discussion

In this article, the properties of the index C_{pc}, suggested by Perakis and Xekalaki [11] are examined under two different distributional assumptions that result in discrete valued data (count and binary data). The obtained results offer a useful approach for measuring process capabilities on the basis of qualitative aspects since none of the most broadly used capability indices can be used in connection with this type of data despite the fact that they are quite frequently encountered in process control. The study of the properties of this index on the basis of such data under different distributional assumptions, for either continuous or discrete processes that usually arise in applications, would be an interesting issue for further research.

References

[1] Kane, V.E., 1986, Process capability indices. *Journal of Quality Technology*, **18**(1), 41–52 (Corrigenda, **18**(4), 265).
[2] Kotz, S. and Johnson, N.L., 1993, *Process Capability Indices* (Chapman and Hall).
[3] Kotz, S. and Johnson, N.L., 2002, Process capability indices – a review, 1992–2000. *Journal of Quality Technology*, **34**(1), 2–19.
[4] Kotz, S. and Lovelace, C.R., 1998, *Process Capability Indices in Theory and Practice* (Arnold).
[5] Spiring, F., Leung, B., Cheng, S. and Yeung, A., 2003, A bibliography of process capability papers. *Quality and Reliability Engineering International*, **19**(5), 445–460.
[6] Clements, J.A., 1989, Process capability calculations for non-normal distributions. *Quality Progress*, **22**(9), 95–100.
[7] Pearn, W.L. and Kotz, S., 1994–1995, Application of Clements' method for calculating second- and third-generation process capability indices for non-normal pearsonian populations. *Quality Engineering*, **7**(1), 139–145.

[8] Pearn, W.L. and Chen, K.S., 1995, Estimating process capability indices for non-normal pearsonian populations. *Quality and Reliability Engineering International*, **11**, 389–391.

[9] Yeh, A.B. and Bhattacharya, S., 1998, A robust process capability index. *Communications in Statistics – Simulation Computation*, **27**(2), 565–589.

[10] Borges, W. and Ho, L.L., 2001, A fraction defective based capability index. *Quality and Reliability Engineering International*, **17**, 447–458.

[11] Perakis, M. and Xekalaki, E., 2002, A process capability index that is based on the proportion of conformance. *Journal of Statistical Computation and Simulation*, **72**(9), 707–718.

[12] Bothe, D.R., 1997, *Measuring Process Capability* (McGraw-Hill).

[13] Montgomery, D.C., 1997, *Introduction to Statistical Quality Control* (Wiley).

[14] Johnson, N.L., Kotz, S. and Kemp, A.W., 1993, *Univariate Discrete Distributions* (Wiley).

[15] Folks, J.L., Pierce, D.A. and Stewart, C., 1965, Estimating the fraction of acceptable product. *Technometrics*, **7**(1), 43–50.

[16] Mood, A.M., Graybill, F.A. and Boes, D.C., 1974, *Introduction to the Theory of Statistics* (McGraw-Hill).