Michele Scagliarini

A Sequential Hypothesis Testing Procedure for the Process Capability Index C_{pk}

Quaderni di Dipartimento

Serie Ricerche 2016, n. 5
ISSN 1973-9346

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

A SEQUENTIAL HYPOTHESIS TESTING PROCEDURE FOR THE PROCESS CAPABILITY INDEX C_{pk}

Michele Scagliarini
Department of Statistical Sciences, University of Bologna, Bologna Italy; email: michele.scagliarini@unibo.it

Abstract

In this study we propose a sequential procedure for hypothesis testing on the C_{pk} process capability index. We compare the properties of the sequential test with the performances of non-sequential tests by performing an extensive simulation study. The results indicate that the proposed sequential procedure makes it possible to save a large amount of sample size, which can be translated into reduced costs, time and resources.

Keywords

Average Sample Size; Brownian Motion; Maximal Allowable Sample Size; Power Function; Simulation Studies.

1. Introduction

Process capability indices assess the relationship between the actual process performance and the manufacturing specifications, and are the tools most frequently used for measuring the capability of a manufacturing process. The analytical formulation of these indices is easy to understand and straightforward to apply. The process capability indices most widely used in industry today are (Montgomery, 2009):

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{d}{3\sigma} \tag{1}$$

$$C_{pk} = \frac{d - \left|\mu - \tfrac{1}{2}\left(USL - LSL\right)\right|}{3\sigma} = \frac{d - \left|\mu - m\right|}{3\sigma} \tag{2}$$

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + \left(\mu - T\right)^2}} = \frac{d}{3\sqrt{\sigma^2 + \left(\mu - T\right)^2}} \tag{3}$$

where μ is the process mean, σ is the process standard deviation, LSL and USL are the specification limits, $d = \left(USL - LSL\right)/2$ is the half-length of the specification interval, $m = \left(USL + LSL\right)/2$ is the midpoint of the specification interval and T is the target value of the process.

Process capability indices have received much interest in statistical literature over the last decades. Evidence of this interest is provided by several books and numerous articles. With reference to books, those of Kotz and Johnson (1993), Bothe (1997), Kotz and Lovelace (1998), Wheeler (1999), Polansky and Kirmani (2003), Pearn and Kotz (2006) can be included. Among the articles we quote the complete overview published by Kotz and Johnson (2002), the bibliographies by Spiring et al. (2003) and Yum and Kim (2011).

Often, as a part of contractual agreement, suppliers are required to provide evidence that their processes satisfy a minimum level of capability. Such decision-making problem of demonstrating whether the process capability exceeds a pre-set capability requirement can be approached in terms of hypothesis testing.

Literature concerning process capability hypothesis testing includes numerous interesting researches. Just to mention a few we quote: the pioneering work by Kane (1986); the tests on C_{pk} investigated by Pearn and Chen (1999), Perakis and Xekalaki (2003), Pearn and Lin (2004), Chen and Hsu (2004) and Lin (2006); the Bayesian approach proposed by Fan and Kao (2006); the hypothesis testing studies on C_{pmk} by Pearn and Lin (2002) and Pearn et al. (2005); the model free approach testing procedure proposed by Vännman and Kulachi (2008); the recently unified and comprehensive analysis of hypothesis testing with process capability indices by Lepore and Palumbo (2015); and finally, the sequential procedure for testing the equality of two indices C_{pm} by Hussein et al. (2012).

In general, sequential methods for hypothesis testing (Tartakovsky *et al.* 2014) are appealing since they make it possible to reach decisions much more quickly, on average, than non-sequential procedures with the same discriminating power. This property, in the framework of the manufacturing industry, means that sequential procedures can lead to saving sample size, time and cost with consequent economic benefits and without any loss in quality.

In this study, starting from some of the results obtained by Hussein et al. (2012) which provided a sequential approach for testing the equality of the indices C_{pm} for two processes, we propose a sequential procedure for hypothesis testing on the index C_{pk}.

We compared the sequential test properties with the performances of two non-sequential tests by performing an extensive simulation study. The results indicate that the proposed sequential test makes it possible to save a large amount of sample size, which can be translated into reduced costs, time and resources.

The paper is organized as follows. In Section 2, we review two of the most used tests for assessing whether a process is capable or not based on the C_{pk} process capability index. In Section 3 we present the general sequential test procedure proposed by Hussein et al. (2012) and Hussein (2005). In Section 4 we develop and propose a sequential method for testing hypotheses on C_{pk}. In Section 5 we study the performances of the proposed test by performing a set of simulation studies. Section 6 contains a discussion of the results and finally, our concluding remarks are reported in Section 7.

2. Hypothesis testing on C_{pk}

To demonstrate whether a process meets the capability requirements the hypotheses of interest are

$$H_0 : C_{pk} \le c_{pk,0} \text{ the process is not capable} \tag{4}$$

$$H_1 : C_{pk} > c_{pk,0} \text{ the process is capable} \tag{5}$$

where $c_{pk,0}$ is the minimal requirement for C_{pk}.

For testing H_0 versus H_1 Pearn and Chen (1999), assuming a normally distributed quality characteristic, $X \sim N\left(\mu, \sigma^2\right)$, proposed a statistical test (PC-test) based on the distribution of the estimator

$$\tilde{C}_{pk} = b_f \hat{C}''_{pk} \tag{6}$$

where

$$\hat{C}''_{pk} = \frac{d - \left(\bar{X} - m\right) I_A\left(\mu\right)}{3S} \tag{7}$$

$b_f = \frac{\Gamma\left[(n-1)/2\right]}{\Gamma\left[(n-2)/2\right]}\sqrt{\frac{2}{n-1}}$, n is the sample size, $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, $S = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}}$, $I_A\left(\mu\right) = 1$ if $\mu \in A$ and $I_A\left(\mu\right) = -1$ if $\mu \notin A$ with $A = \left\{\mu \middle| \mu \geq m\right\}$.

Given the type I error probability α, the critical value of the test is

$$C_0 = \frac{b_f t_{n-1,\alpha}\left(\delta_c\right)}{3\sqrt{n}} \tag{8}$$

where $t_{n-1,\alpha}\left(\delta_c\right)$ is the upper α quantile of a non-central *t* with $n-1$ degrees of freedom and non-centrality parameter $\delta_c = 3\sqrt{n} c_{pk,0}$.

The power of the PC-test can be computed as

$$\pi_{PC}\left(c_{pk,1}\right) = \Pr\left\{\tilde{C}_{pk} > C_0 \middle| C_{pk} = c_{pk,1}\right\} = \Pr\left\{t_{n-1}\left(\delta\right) > \frac{3\sqrt{n}C_0}{b_f}\right\} \tag{9}$$

where $\delta = 3\sqrt{n} c_{pk,1}$.

Lepore and Palumbo (2015), for testing $H_0 : C_{pk} \leq c_{pk,0}$ versus $H_1 : C_{pk} > c_{pk,0}$, discussed a test (LP-test) based on the estimator (Vännman, 2006)

$$\hat{C}_{pk} = \frac{1 - \left|\hat{\delta}\right|}{3\hat{\gamma}} \tag{10}$$

where $\hat{\gamma} = \frac{\hat{\sigma}}{d}$, $\hat{\delta} = \frac{\bar{X} - m}{d}$ and $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n}}$.

Under the assumption of a normally distributed quality characteristic, the authors obtained the cumulative distribution function of $\hat{C}_{pk}$ as

$$F_{\hat{C}_{pk}}\left(x\right) = \begin{cases} 1 - \Pr\left(t_{n-1}\left(\delta_1\right) \leq -t\right) + \Pr\left(t_{n-1}\left(\delta_2\right) \leq -t\right) & x \leq 0 \\ 1 - Q_{n-1}\left(-t, \delta_1; 0, R\right) + Q_{n-1}\left(t, \delta_2; 0, R\right) & x > 0 \end{cases} \tag{11}$$

where $t_{n-1}(\delta_1)$ and $t_{n-1}(\delta_2)$ are non-central t variables with $n-1$ degrees of freedom and non-centrality parameters $\delta_1 = -3\sqrt{n}\frac{(1-\delta)}{(1-|\delta|)}C_{pk}$ and $\delta_2 = 3\sqrt{n}\frac{(1+\delta)}{(1+|\delta|)}C_{pk}$ respectively, $\delta = \frac{\bar{X}-\mu}{d}$, $t = 3\sqrt{n-1}x$, $R = \sqrt{n-1}(\delta_2 - \delta_1)/2t$ and $Q_f(t,\delta;0,R) = \frac{\sqrt{2\pi}}{\Gamma(f/2)2^{(f-2)/2}}\int_0^R \Phi\left(\frac{tx}{\sqrt{f}} - \delta\right)x^{f-1}\phi(x)dx$ is the Q_f-function proposed by Owen (1965), where Γ is the gamma function, and Φ and ϕ are respectively the normal cumulative distribution function and probability density function.

Lepore and Palumbo (2015) obtained the critical value for the test as

$$c_{pk;\alpha} = \frac{t_{n-1;\alpha}(\delta_0)}{3\sqrt{n-1}} \tag{12}$$

where $t_{n-1;\alpha}(\delta_0)$ is the upper α quantile of a non-central t distribution with $n-1$ degrees of freedom and non-centrality parameter $\delta_0 = 3\sqrt{n}c_{pk,0}$.

The power of the LP-test is then

$$\pi_{LP}(c_{pk,1}) = \Pr\left\{\hat{C}_{pk} > c_{pk;\alpha} \middle| C_{pk} = c_{pk,1}\right\} = 1 - F_{\hat{C}_{pk}}(c_{pk;\alpha}) \tag{13}$$

given that $C_{pk} = c_{pk,1} > c_{pk,0}$.

3. A general sequential method.

To describe the general sequential testing procedure proposed by Hussein et al. (2012) and Hussein (2005), we used $x_1, x_2, ..., x_k, ...,$ to denote a sequence of multivariate independent observations collected over time. We assumed that these data came from a common multivariate distribution with density function $f(x;\boldsymbol{\theta})$ where the vector of parameters $\boldsymbol{\theta}$ is unknown.

We were interested in testing

$$H_0 : h(\boldsymbol{\theta}) = \mathbf{0} \text{ versus } H_1 : h(\boldsymbol{\theta}) \neq \mathbf{0} \tag{14}$$

where $h(\boldsymbol{\theta}) : \mathbb{R}^d \to \mathbb{R}^q$, with $q \leq d$, is a function with first order derivative matrix denoted by $H(\boldsymbol{\theta})$.

Let us assume that for $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ with $q \geq d$ the following regularity conditions hold:

C1. The distribution function $F(x;\boldsymbol{\theta})$ of the vector random variable *X* is identifiable over Θ.

C2. There exists an open subset, $\Theta_0 \subset \Theta$, containing the true value of the parameter under H_0, such that the partial derivatives

$$\frac{\partial}{\partial \theta_i} \ln f(x;\boldsymbol{\theta}),\ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x;\boldsymbol{\theta}),\ \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln f(x;\boldsymbol{\theta})$$

exist and are continuous for all $x \in \mathbb{R}^l$, $\boldsymbol{\theta} \in \Theta_0$.

C3. For each $\boldsymbol{\theta} \in \Theta_0$ and $k = 1,2,3,...,$ the score equation

$$\sum_{i=1}^{k} \frac{\partial \ln f(x;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$$

has a unique solution.

C4. There are functions, $M_1(x)$ and $M_2(x)$, that have finite expectations under any of the parameter values, $\boldsymbol{\theta} \in \Theta_0$, such that

$$\left|\frac{\partial}{\partial \theta_i} \ln f(x;\boldsymbol{\theta})\right| \leq M_1(x),\ \left|\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x;\boldsymbol{\theta})\right| \leq M_2(x),\ \left|\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln f(x;\boldsymbol{\theta})\right| \leq M_2(x)$$

for all $\boldsymbol{\theta} \in \Theta_0$, $1 \leq i \quad j,k \leq d$.

C5. $E_\theta(\partial/\partial\theta_i)\ln f(x;\boldsymbol{\theta}) = 0$, $1 \leq i \leq d$, $\boldsymbol{\theta} \in \Theta_0$.

C6. $I^{-1}(\theta)$ and its elements $I_{ij}(\theta) = E_\theta\left[\left(\partial^2/\partial\theta_i\partial\theta_j\right)\ln f(x;\boldsymbol{\theta})\right]$ exist and are continuous for all $\boldsymbol{\theta} \in \Theta_0$, $1 \leq i,\ j \leq d$.

C7. $\text{Var}_\theta\left[\left(\partial^2/\partial\theta_i\partial\theta_j\right)\ln f(x;\boldsymbol{\theta})\right] < \infty$ for $1 \leq i,\ j \leq d$.

Let us further assume that:

C8. $E_\theta\left|(\partial/\partial\theta_i)\ln f(x;\boldsymbol{\theta})\right|^{2+\delta} < \infty$, $i = 1,2,...,d$, and for some $\delta > 0$

C9. The function $h(\boldsymbol{\theta})$ is continuously differentiable over Θ_0 and its first-order derivative matrix $H(\boldsymbol{\theta})$, is bounded and of rank *q*.

Let us now consider a fixed sample design with sample size equal to *k* and let us consider the Wald's statistic

$$W_k = kh\left(\hat{\boldsymbol{\theta}}_k\right)\left[H^t(\boldsymbol{\theta}) I^{-1}(\boldsymbol{\theta}) H(\boldsymbol{\theta})\right]^{-1} h\left(\hat{\boldsymbol{\theta}}_k\right)^t \tag{15}$$

where $\hat{\boldsymbol{\theta}}_k$ is a consistent estimator of $\boldsymbol{\theta}$.

Hussein et al. (2012) in Theorem 1 showed that under H_0, and if conditions C1-C9 hold, there exists an independent Wiener process, $B_j(t)$, $j = 1, 2, ..., q$, such that for $\alpha \leq \frac{1}{2} - 1/(2+\delta)$ with δ>0,

$$\sup_{1 \leq t < \infty} \left| W_{[kt]} - U_{[kt]} \right| = O\left(k^{-\alpha} (\ln \ln k)^{1/2}\right) \tag{16}$$

Where

$$U(x) = \frac{1}{x} \sum_{j=1}^{q} B_j^2(x) \tag{17}$$

$$W_{[kt]} = [kt] h\left(\hat{\boldsymbol{\theta}}_{[kt]}\right)\left[H^t(\boldsymbol{\theta}) I^{-1}(\boldsymbol{\theta}) H(\boldsymbol{\theta})\right]^{-1} h\left(\hat{\boldsymbol{\theta}}_{[kt]}\right)^t \tag{18}$$

and $[.]$ denotes the integer part of its argument.

The statistic W_k can therefore be approximated by a functional of Brownian motions. Furthermore, the authors derived the limiting distribution of W_k. In detail they showed that (Corollary 1):

- under the conditions of Theorem 1

$$\max_{1 < k \leq n} \left[\frac{k}{n} W_k\right]^{1/2} \xrightarrow{D} \sup \left(\sum_{j=1}^{q} B_j^2(t)\right)^{1/2} \tag{19}$$

  where $\xrightarrow{D}$ denotes convergence in distribution;
- replacing the unknown $\boldsymbol{\theta}$ in the term $H^t(\boldsymbol{\theta}) I^{-1}(\boldsymbol{\theta}) H(\boldsymbol{\theta})$ by any almost surely convergent estimator, Corollary 1 remains valid.

Therefore, Hussein e*t al.* (2012) defined as test statistic

$$W_k^* = kh\left(\hat{\boldsymbol{\theta}}_k\right)\left[H^t\left(\hat{\boldsymbol{\theta}}_k\right) I^{-1}\left(\hat{\boldsymbol{\theta}}_k\right) H\left(\hat{\boldsymbol{\theta}}_k\right)\right]^{-1} h\left(\hat{\boldsymbol{\theta}}_k\right)^t \tag{20}$$

where $\hat{\boldsymbol{\theta}}_k$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, and proposed the following α-level sequential test truncated at the maximal allowable sample size n_0.

The sequential test procedure is performed as follows:

- for $k = 2, 3, ..., n_0$ compute of the statistic

$$W_k^{*(1)} = \sqrt{k/n_0} \sqrt{W_k^*}\,; \tag{21}$$

- the hypothesis H_0 is rejected the first time that $W_k^{*(1)}$ exceeds the critical value w_α;
- if $W_k^{*(1)}$ does not exceed w_α by n_0 then do not reject H_0.

The maximal sample size n_0 can be decided on the basis of financial, ethical or statistical reasons as, for example, to achieve a desired power level.

Given the Type I error probability α, the critical value w_α can be obtained from Borodin and Salminen (1996).

4. A sequential test for C_{pk}

Let us now consider the hypothesis

$$H_0 : C_{pk} = c_{pk,0} \tag{22}$$

versus

$$H_1 : C_{pk} \neq c_{pk,0} \tag{23}$$

and assume that the quality characteristic of interest *X* is normally distributed: $X \sim N\left(\mu, \sigma^2\right)$.

For $C_{pk} \geq 0$, H_0 is equivalent to

$$H_0 : \ln\left(\left(C_{pk}\right)^2\right) - \ln\left(\left(c_{pk,0}\right)^2\right) = 0 \tag{24}$$

and the alternative hypothesis is equivalent to

$$H_1 : \ln\left(\left(C_{pk}\right)^2\right) - \ln\left(\left(c_{pk,0}\right)^2\right) \neq 0 \tag{25}$$

Let us define the function $h(\boldsymbol{\theta})$ as

$$h(\boldsymbol{\theta}) = \ln\left(\left(C_{pk}\right)^2\right) - \ln\left(\left(c_{pk,0}\right)^2\right) = \ln\left[\frac{\left(d - |\mu - m|\right)^2}{9\sigma^2\left(c_{pk,0}\right)^2}\right] \tag{26}$$

where $\boldsymbol{\theta} = \left(\mu, \sigma^2\right)$. Note that in this framework $h(\boldsymbol{\theta}) : \mathbb{R}^d \to \mathbb{R}^q$ with *d*=2 and *q*=1.

In the case at hand, where *q*=1, the statistic W_k^* can be written as

$$W_k^* = kh^2\left(\hat{\boldsymbol{\theta}}_k\right)\left[H^t\left(\hat{\boldsymbol{\theta}}_k\right)I^{-1}\left(\hat{\boldsymbol{\theta}}_k\right)H\left(\hat{\boldsymbol{\theta}}_k\right)\right]^{-1} \tag{27}$$

where $\hat{\boldsymbol{\theta}}_k = \left(\bar{X}_k, S_k^2\right)$ with $\bar{X}_k = \frac{\sum_{i=1}^{k} X_i}{k}$ and $S_k^2 = \frac{\sum_{i=1}^{k}\left(X_i - \mu\right)^2}{k}$.

The function $h\left(\hat{\boldsymbol{\theta}}_k\right)$ is therefore given by

$$h\left(\hat{\boldsymbol{\theta}}_k\right) = \ln\left(\frac{\left(d - \left|\bar{X}_k - m\right|\right)^2}{9S_k^2\left(c_{pk,0}\right)^2}\right) \tag{28}$$

and the partial derivative matrix of $h(\boldsymbol{\theta})$ computed at $\hat{\boldsymbol{\theta}}_k$ can be written as

$$H\left(\hat{\boldsymbol{\theta}}_k\right) = \begin{bmatrix} -\frac{2signum\left[\bar{X}_k - m\right]}{d - \left|\bar{X}_k - m\right|} \\ -\frac{1}{S_k^2} \end{bmatrix} \tag{29}$$

where $signum[a] = \frac{a}{|a|}$ if $a \neq 0$; $signum[a] = 0$ if $a = 0$.

Since the Fisher information matrix for normally distributed data is $I(\boldsymbol{\theta}) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/2\sigma^4 \end{bmatrix}$, it follows that the statistic W_k^* can be written as

$$W_k^* = k\left[\ln\left(\frac{\left(d - \left|\bar{X}_k - m\right|\right)^2}{9S_k^2\left(c_{pk,0}\right)^2}\right)\right]^2 \times \left[\frac{4\bullet\left(signum\left[\bar{X}_k - m\right]\right)^2 S_k^2}{\left[d - \left|\bar{X}_k - m\right|\right]^2} + 2\right]^{-1} \tag{30}$$

Therefore, given the value of α and the maximal allowable sample size n_0, the test is performed by computing, for *k*=2,3,…, n_0, the statistic

$$W_k^{*(1)} = \sqrt{\left(k/n_0\right)}\sqrt{W_k^*} = \sqrt{\left(k/n_0\right)}\sqrt{k\left[\ln\left(\frac{\left(d - \left|\bar{X}_k - m\right|\right)^2}{9S_k^2\left(c_{pk,0}\right)^2}\right)\right]^2 \times \left[\frac{4\bullet\left(signum\left[\bar{X}_k - m\right]\right)^2 S_k^2}{\left[d - \left|\bar{X}_k - m\right|\right]^2} + 2\right]^{-1}} \tag{31}$$

Let n_{stop} be the first integer k=2,3,…, n_0 for which $W_k^{*(1)} > w_\alpha$:

- we reject H_0 if $W_{n_{stop}}^{*(1)} > w_\alpha$:
- we do not reject H_0 if $W_k^{*(1)}$ does not exceed w_α by n_0.

In this framework n_{stop} is the stopping sample size of the test.

In the case at hand, where q=1, the critical value w_α is obtained from the distribution of $\sup_{0<t<1}|B(t)|$. In particular w_α is such that (Feller 1970)

$$1-\alpha = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(\frac{-(2k+1)^2\pi^2}{8w_\alpha^2}\right) \tag{32}$$

As an example, for α= 0.02, 0.1 and 0.2, the values of w_α are 2.576, 1.96 and 1.645 respectively.

5. Simulation studies

In this section we study the properties of the sequential procedure by comparing its performances, under H_0 and H_1, with those of the LP and PC-tests. More precisely, we compare the tests in terms of the sample size required for achieving a given value of power.

Note that the sequential test is two sided with composite alternative hypothesis $H_1: C_{pk} \neq c_{pk,0}$, while the LP and PC-tests are unilateral. In order to correctly compare the performances of the tests, we considered cases under H_1 where $C_{pk} = c_{pk,1}$ with $c_{pk,1} > c_{pk,0}$. In this manner the sequential bilateral test with Type I error probability α can be compared with the non-sequential unilateral tests with Type I error probability equal to $\alpha_u = \alpha/2$.

To study the properties of the sequential procedure under H_1, we examined several scenarios (details on how the scenarios were built are reported in Appendix A), which are also discussed in Lepore and Palumbo (2015), where several values of C_{pk} under H_1 ($c_{pk,1}$) were considered for the unilateral test with α_u =0.01, 0.05, 0.1 and $c_{pk,0}$ =1.00, 1.33, 1.67.

For the LP and PC-tests we determined the minimum sample size, $n_{LP;0.80}$ and $n_{PC;0.80}$, using formulas (9) and (13) respectively to achieve a power at least equal to or greater than 0.80: i.e. $\pi_{LP}(c_{pk,1}) \geq 0.80$ and $\pi_{PC}(c_{pk,1}) \geq 0.80$.

As far as the sequential test is concerned we used a set of simulation studies. For each value of α, $c_{pk,0}$ and $c_{pk,1}$ we generated, using R (R core team 2013), 10^4 replicates from a normally distributed quality characteristic. The aim of these simulations was to determine the smallest maximal allowable sample size, $n_{0;\hat{\pi}_s>0.80}$, which gives an empirical power $\hat{\pi}_s$ greater than 0.80: i.e. $\hat{\pi}_s > 0.80$. The empirical power $\hat{\pi}_s$ of the sequential test is estimated as the proportion of correctly rejected H_0.

In order to obtain $n_{0;\hat{\pi}_s>0.80}$ we implemented an iterative search algorithm with an initial value for $n_{0;\hat{\pi}_s>0.80}$ given by $n_{start} = n_{PC;0.80}$. The algorithm works as follows:

1. With fixed n_{start} as the maximal allowable sample size of the sequential test, the empirical power of the test $\hat{\pi}_s$ is estimated as the proportion of correctly rejected H_0 on $m = 10^4$ simulations.
2. If

$$0 < \frac{\hat{\pi}_s - 0.80}{0.80} \leq 0.025$$

then $n_{0;\hat{\pi}_s>0.80}$ is set equal to n_{start} and the search algorithm stops. At the same time the average stopping sample size n_{avg} was empirically assessed as the average of the stopping sample sizes n_{stop} required by the sequential test to correctly reject H_0 when the maximal allowable sample size is equal to $n_{0;\hat{\pi}_s>0.80}$.

3. Otherwise: if $\hat{\pi}_s \leq 0.80$, then $n_{start} = n_{start} + 1$; if $\hat{\pi}_s > 0.80$, then $n_{start} = n_{start} - 1$ and the algorithm starts other *m* simulations.

The simulation results are summarized in Tables 1-9 where, for each combination of α, $c_{pk,0}$ and $c_{pk,1}$, the following quantities are reported: $n_{LP;0.80}$ the minimum sample size required by the LP-test for achieving a power level ≥ 0.80; $n_{PC;0.80}$ the minimum sample size required by the PC-test for achieving a power level ≥ 0.80; $n_{0;\hat{\pi}_s>0.80}$ the smallest maximal allowable sample size for the sequential test for achieving an empirical power $\hat{\pi}_s > 0.80$; n_{avg} the average of the stopping sample sizes n_{stop} required for the sequential test with maximal allowable sample size $n_{0;\hat{\pi}_s>0.80}$ for concluding in favor of H_1; $\text{S.D.}(n_{stop})$ the standard deviation of the final sample sizes n_{stop}; $\hat{\pi}_s$ the estimated power of the sequential test.

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.20	>200	195	182	123.6	31.5	0.807
1.30	107	94	88	59.4	15.7	0.817
1.40	66	59	51	34.3	9.7	0.806
1.50	46	41	35	23.2	7.0	0.810
1.60	35	32	25	16.4	5.3	0.805
1.70	28	26	19	12.3	4.3	0.802
1.80	23	21	16	10.1	3.8	0.820
1.90	20	18	13	8.2	3.1	0.817
2.00	17	16	11	6.9	2.7	0.816

Table 1: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.00$ and α=0.02 ($w_\alpha = 2.576$).

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.20	140	116	116	71.2	23.3	0.804
1.30	68	57	56	33.6	12.2	0.820
1.40	42	36	32	18.5	7.6	0.813
1.50	29	25	21	11.8	5.3	0.816
1.60	22	20	15	8.2	3.9	0.806
1.70	18	16	11	6.0	2.9	0.815
1.80	15	13	9	4.9	2.3	0.812
1.90	13	12	7	4.0	1.7	0.801
2.00	11	10	6	3.5	1.4	0.808

Table 2: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.00$ and α=0.1 ($w_\alpha = 1.96$).

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.20	104	84	83	46.7	18.6	0.803
1.30	51	42	40	21.3	9.9	0.811
1.40	31	26	22	11.2	5.8	0.806
1.50	22	19	14	7.0	3.7	0.804
1.60	17	14	10	5.0	2.6	0.810
1.70	13	12	8	4.1	2.0	0.819
1.80	11	10	6	3.3	1.4	0.807
1.90	10	9	5	2.9	1.1	0.800
2.00	8	8	4	2.6	0.8	0.801

Table 3: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.00$ and α=0.2 ($w_\alpha = 1.645$).

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.60	199	178	171	116.1	29.8	0.811
1.70	115	104	96	64.7	17.2	0.815
1.80	75	68	62	41.7	11.2	0.809
1.90	55	50	44	29.4	8.4	0.820
2.00	42	39	33	21.8	6.6	0.814
2.10	34	32	26	17.1	5.5	0.815
2.20	29	27	21	13.7	4.7	0.808
2.30	25	23	17	11.0	3.9	0.807

Table 4: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.33$ and α=0.02 ($w_\alpha = 2.576$).

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.60	124	108	107	65.8	21.5	0.820
1.70	70	62	60	36.0	12.7	0.817
1.80	47	41	39	23.0	8.8	0.819
1.90	34	30	27	15.5	6.5	0.818
2.00	27	24	19	10.6	4.9	0.804
2.10	22	20	15	8.2	3.9	0.807
2.20	18	17	12	6.5	3.1	0.809
2.30	16	14	10	5.5	2.6	0.816

Table 5: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.33$ and α=0.1 ($w_\alpha = 1.96$).

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.60	92	77	77	43.2	17.3	0.808
1.70	52	44	44	23.7	10.6	0.816
1.80	35	30	27	14.1	7.0	0.811
1.90	25	22	19	9.7	5.1	0.816
2.00	20	18	13	6.5	3.5	0.802
2.10	16	14	10	5.0	2.6	0.809
2.20	14	12	8	4.1	2.0	0.804
2.30	12	11	7	3.7	1.7	0.817

Table 6: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.33$ and α=0.2 ($w_\alpha = 1.645$).

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.90	>200	>200	>200	>200	-	-
2.00	198	181	173	117.2	29.9	0.816
2.10	125	115	106	71.6	18.3	0.812
2.20	88	82	74	49.7	13.4	0.819
2.30	67	62	53	35.6	9.8	0.804
2.40	51	47	42	28.0	8.0	0.817
2.50	42	39	33	22.0	6.6	0.808
2.60	35	33	27	17.9	5.5	0.803

Table 7: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.67$ and α=0.02 ($w_\alpha = 2.576$).

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.90	>200	>200	208	129.3	40.7	0.802
2.00	123	109	106	65.2	21.4	0.809
2.10	78	69	65	39.4	13.6	0.800
2.20	53	48	45	27.0	10.0	0.813
2.30	40	36	32	18.7	7.6	0.806
2.40	32	29	25	14.4	6.1	0.807
2.50	26	24	19	10.7	4.9	0.806
2.60	22	20	16	8.8	4.1	0.812

Table 8: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.67$ and α=0.1 ($w_\alpha = 1.96$).

$c_{pk,1}$	$n_{LP;0.80}$	$n_{PC;0.80}$	$n_{0;\hat{\pi}_s>0.80}$	n_{avg}	S.D.(n_{stop})	$\hat{\pi}_s$
1.90	175	151	154	89.3	32.8	0.806
2.00	91	79	79	44.5	17.6	0.812
2.10	56	49	49	26.8	11.6	0.814
2.20	39	35	32	16.8	8.2	0.811
2.30	30	26	23	11.9	6.1	0.813
2.40	24	21	17	8.6	4.6	0.812
2.50	19	17	14	7.0	3.7	0.818
2.60	16	15	11	5.5	2.9	0.817

Table 9: Simulation results under H_1 with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.67$ and α=0.2 ($w_\alpha = 1.645$).

To study the properties of the sequential procedure under $H_0 : C_{pk} = c_{pk,0}$, for each combination of $c_{pk,0}$ (1.00, 1.33, 1.67) and α (0.02, 0.1, 0.2), we generated 10^4 replicates from a normally distributed process. The aim of these simulations was to determine the smallest maximal allowable sample size, $n_{0;1-\hat{\alpha}>1-\alpha}$, which gives an empirical type I error probability $\hat{\alpha}$ smaller than the nominal α value: $\hat{\alpha} < \alpha$ or equivalently $1-\hat{\alpha} > 1-\alpha$.

In order to obtain $n_{0;1-\hat{\alpha}>1-\alpha}$ we implemented an iterative search algorithm with initial value for $n_{0;1-\hat{\alpha}>1-\alpha}$ given by $n_{start} = 75$. The algorithm works as follows:

1. With fixed n_{start} as the maximal allowable sample size of the sequential test, the value of $1-\hat{\alpha}$ is estimated as the fraction of correctly accepted H_0 on $m = 10^4$ simulations.
2. If

$$0 < \frac{(1-\hat{\alpha})-(1-\alpha)}{(1-\alpha)} \leq 0.025$$

then $n_{0;1-\hat{\alpha}>1-\alpha}$ is set equal to n_{start} and the search algorithm stops. Contextually the average stopping sample size under H_0, $n_{H_0,avg}$, is empirically assessed as the average of the final sample sizes n_{stop} required by the sequential test to correctly accept H_0 when the maximal allowable sample size is equal to $n_{0;1-\hat{\alpha}>1-\alpha}$.

3. Otherwise: if $(1-\hat{\alpha}) \le (1-\alpha)$, then $n_{start} = n_{start} + 1$; if $(1-\hat{\alpha}) > (1-\alpha)$, then $n_{start} = n_{start} - 1$ and the algorithm starts other *m* simulations.

The simulation results are summarized in Tables 10-12, where for each combination of α and $c_{pk,0}$ the following quantities are reported: $n_{0;1-\hat{\alpha}>1-\alpha}$ the smallest maximal allowable sample size for the sequential test for achieving an empirical type I error probability $\hat{\alpha}$ smaller than the nominal α $(\hat{\alpha} < \alpha)$; $n_{H_0,avg}$ the average of the final sample sizes n_{stop} required for the sequential test with maximal sample size $n_{0;1-\hat{\alpha}>1-\alpha}$ for correctly concluding in favor of H_0; S.D.(n_{stop}) the standard deviation of the final sample sizes n_{stop}; $\hat{\alpha}$ the estimated type I error probability; $1-\hat{\alpha}$.

α	$n_{0;1-\hat{\alpha}>1-\alpha}$	$n_{H_0,avg}$	S.D.(n_{stop})	$\hat{\alpha}$	$1-\hat{\alpha}$
0.02	127	100.9	21.9	0.0191	0.9809
0.1	82	56.8	19.5	0.099	0.9010
0.2	62	37.7	16.5	0.1949	0.8051

Table 10: Simulation results under $H_0 : C_{pk} = c_{pk,0}$ with $c_{pk,0} = 1.00$.

α	$n_{0;1-\hat{\alpha}>1-\alpha}$	$n_{H_0,avg}$	S.D.(n_{stop})	$\hat{\alpha}$	$1-\hat{\alpha}$
0.02	110	86.8	19.3	0.0195	0.9805
0.1	67	44.6	16.6	0.0989	0.9011
0.2	69	43.0	18.1	0.1958	0.8042

Table 11: Simulation results under $H_0 : C_{pk} = c_{pk,0}$ with $c_{pk,0} = 1.33$.

α	$n_{0;1-\hat{\alpha}>1-\alpha}$	$n_{H_0,avg}$	S.D.(n_{stop})	$\hat{\alpha}$	$1-\hat{\alpha}$
0.02	90	69.2	17.6	0.0198	0.9802
0.1	61	41.2	14.6	0.0982	0.9018
0.2	74	47.0	19.0	0.1978	0.8022

Table 12: Simulation results under $H_0 : C_{pk} = c_{pk,0}$ with $c_{pk,0} = 1.67$.

6. Discussion

As far as the behavior of the test under H_1 is concerned, by examining the averages of the final sample sizes n_{avg} the results show that the sequential test, with the same power of the LP and PC-tests, saves a lot of sample size. Furthermore, the maximum allowable sample size $n_{0;\hat{\pi}_s>0.80}$ required to achieve the desired power is almost always smaller than $n_{LP;0.80}$ and $n_{PC;0.80}$. This indicates that even in the worst cases the sequential test needs a maximum allowable sample size not greater than the sample size of the non-sequential tests.

As an example, under $H_1 : C_{pk} = c_{pk,1}$ with $c_{pk,1} = 1.30$, when $c_{pk,0} = 1.00$ and α=0.02, we have $n_{LP;0.80} = 107$, $n_{PC;0.80} = 94$ (Table 1 and Figure 1), while with a maximum allowable sample size equal to $n_{0;\hat{\pi}_s>0.80} = 88$ the power of the sequential test is $\hat{\pi}_s > 0.80$ with an $n_{avg} = 59.4(\simeq 60)$.

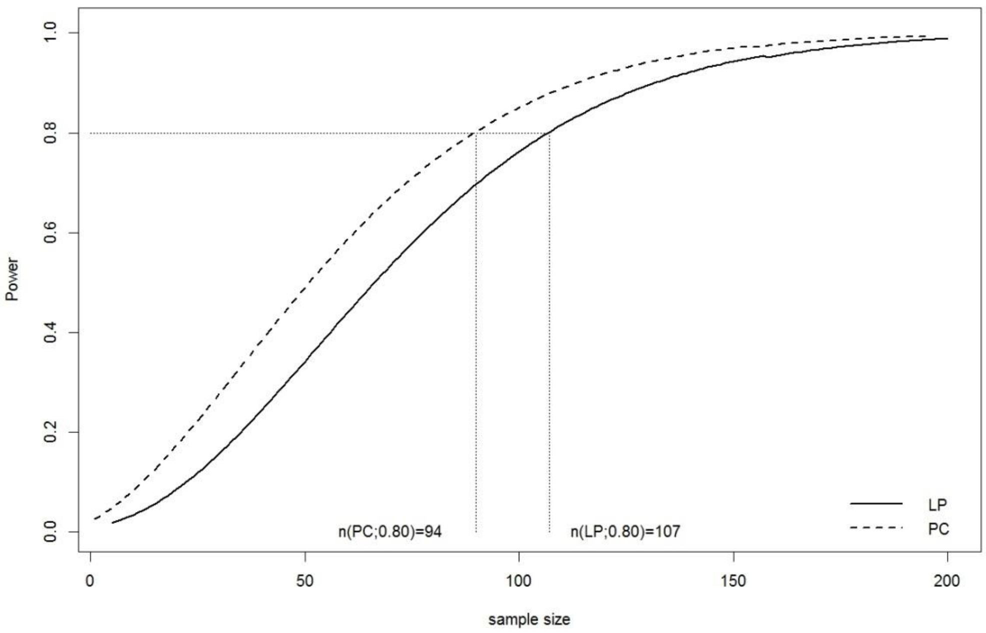

Figure 1: Power functions of the LP and PC tests as a function of the sample size for the case $c_{pk,0} = 1.00$, $\alpha = 0.02(\alpha_u = 0.01)$ and $c_{pk,1} = 1.30$

In this case the sequential procedure saves, on average, 43.9% of the sample size as to the LP-test and 36.2% as to the PC-test.

Under $H_1 : C_{pk} = c_{pk,1}$ with $c_{pk,1} = 1.60$, when $c_{pk,0} = 1.33$ and α=0.1, we have $n_{LP;0.80} = 124$ and $n_{PC;0.80} = 108$ (Figure 2 and Table 5) while with a maximum allowable sample size equal to $n_{0;\hat{\pi}_s > 0.80} = 107$ the power of the sequential test is $\hat{\pi}_s > 0.80$ with an $n_{avg} = 65.8 (\simeq 66)$.

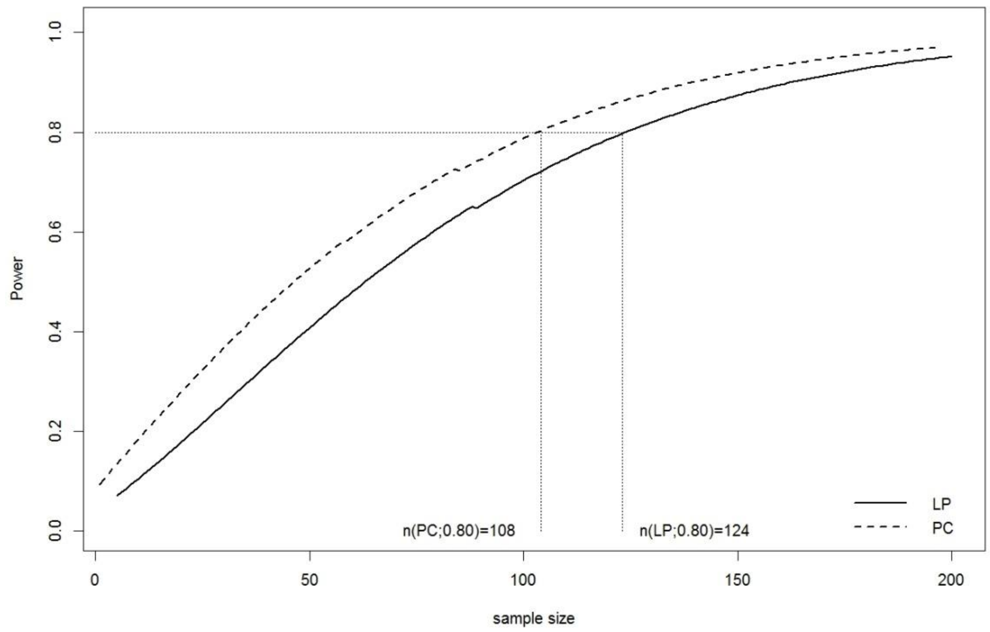

Figure 2: Power functions of the LP and PC tests as a function of the sample size for the case $c_{pk,0} = 1.33$, $\alpha = 0.1 (\alpha_u = 0.05)$ and $c_{pk,1} = 1.60$

In this case the sequential procedure therefore saves, on average, 46.8% of the sample size as to the LP test and 38.9% as to the PC test.

Finally, under $H_1 : C_{pk} = c_{pk,1}$ with $c_{pk,1} = 1.90$, when $c_{pk,0} = 1.67$ and α=0.2, we have $n_{LP;0.80} = 175$ and $n_{PC;0.80} = 151$, (Figure 3 and Table 9) while with a maximum allowable sample size equal to $n_{0;\hat{\pi}_s > 0.80} = 154$ the power of the sequential test is $\hat{\pi}_s > 0.80$ with an $n_{avg} = 89.3 (\simeq 90)$.

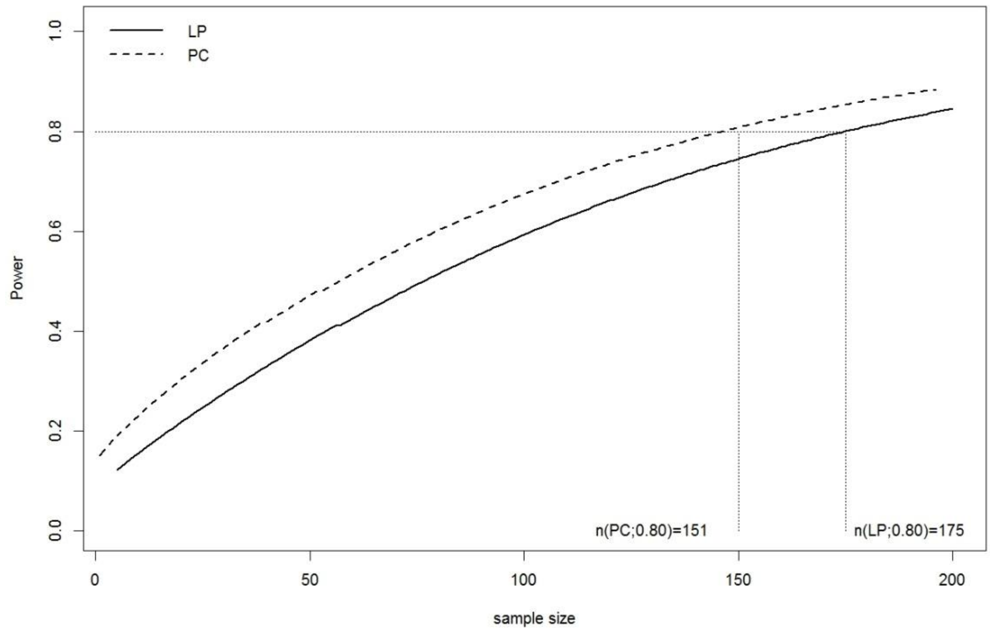

Figure 3: Power functions of the LP and PC tests as a function of the sample size for the case $c_{pk,0} = 1.67$, $\alpha = 0.2(\alpha_u = 0.1)$ and $c_{pk,1} = 1.90$.

In this case the sequential procedure saves, on average, 48.6% of the sample size as to the LP-test and 40.4% as to the PC-test.

Also under H_0 the proposed sequential test displays features that are very interesting. For each value of α and $c_{pk,0}$ the smallest maximal allowable sample sizes $n_{0;1-\hat{\alpha}>1-\alpha}$, required by the sequential test for ensuring an empirical type I error probability $\hat{\alpha}$ smaller than the nominal value α are always smaller than the values of $n_{0;\hat{\pi}_s>0.80}$ corresponding to the worst scenario examined under H_1: i.e. the value of $c_{pk,1}$ closer to $c_{pk,0}$.

Let us consider the following cases:

1. when $H_0: C_{pk} = c_{pk,0}$ holds, with $c_{pk,0} = 1.00$ and $\alpha = 0.02$, we have that $n_{0;1-\hat{\alpha}>1-\alpha} = 127$ (Table 10) while for $c_{pk,1} = 1.20$ it results that $n_{0;\hat{\pi}_s>0.80} = 182$ (Table 1).
2. when $H_0: C_{pk} = c_{pk,0}$ holds, with $c_{pk,0} = 1.33$ and $\alpha = 0.2$, we have that $n_{0;1-\hat{\alpha}>1-\alpha} = 69$ (Table 11) while for $c_{pk,1} = 1.60$ it results that $n_{0;\hat{\pi}_s>0.80} = 77$ (Table 6)
3. when $H_0: C_{pk} = c_{pk,0}$ holds, with $c_{pk,0} = 1.67$ and $\alpha = 0.1$, we have that $n_{0;1-\hat{\alpha}>1-\alpha} = 61$ (Table 12) while for $c_{pk,1} = 1.90$ it results that $n_{0;\hat{\pi}_s>0.80} = 208$ (Table 8).

These results ensure that for all the cases examined the empirical type I error probability does not exceed the nominal α -level of the test.

Furthermore, by examining $n_{H_0,avg}$, the averages of the final sample sizes n required for the sequential test for correctly concluding in favor of H_0, we can assert that the sequential procedure allow early stopping sample sizes also in those cases where the process is not capable.

7. Conclusions

In this article we proposed a sequential procedure for hypothesis testing on the C_{pk} index. We studied the statistical properties of the sequential test with an extensive simulation study with regard to the type I error, the average of the sample sizes for correctly deciding, for H_0 and H_1, the maximum allowable sample size required to achieve a pre-set power level and for ensuring that the empirical type I error probability does not exceed the nominal α -level of the test. We compared the performances of the sequential procedure with two non-sequential tests.

The results showed that the sequential test allows on average smaller stopping sample sizes as compared with the fixed sample size tests while maintaining the desired α -level and power.

Furthermore, the maximum allowable sample sizes required by the sequential test to achieve the desired power level are smaller than, or at most equal to, the sample sizes required by the non-sequential tests: this means that even in the worst cases the sequential procedure uses a sample size that does not exceed the sample size of the non-sequential tests with the same power level (under H_1) or without exceeding the type I error probability (under H_0).

Summarizing, the proposed sequential procedure has several interesting features: it offers a substantial decrease in sample size compared with the non-sequential tests, while type I and II error probabilities are correctly maintained at their desired values.

We consider these results as valuable, because in a highly competitive context where both cost and quality are relevant, the availability of statistical methods which make it possible to save sampling size can be directly translated into saved resources and reduced costs.

Furthermore, process capability analysis is increasingly used in healthcare-related studies (Chen *et al.* 2014, Liu *et al.* 2010) where, in addition to the economic matter, important ethical issues must be taken into account. It is worth noting that within this particular framework, methods capable of shortening the time span or reducing the sample size required for testing process capability can be of great value.

Appendix A

To explain how the scenarios for our study have been built let us consider, without loss of any generality, the case where $c_{pk,0} = 1.33$.

We considered, under H_0, a process where: $USL = 25$, $LSL = 15$, $\mu = 23$ and $\sigma = 0.501253$ in such a way that $C_{pk} = c_{pk,0} = 1.33$.

Under H_1 we examined cases where $c_{pk,1}$ ranged from 1.60 to 2.30 (see e.g. Table 4). Therefore the alternative hypothesis consists of scenarios where the capability was improving: these scenarios can be obtained by improving process centering or by reducing the process variability.

As an example, the case $c_{pk,1} = 1.60$ can be obtained with a mean level closer to the center, $\mu = 22.59398$, with fixed $\sigma = 0.501253$ or by reducing the variability, $\sigma = 0.416667$, with fixed $\mu = 23$.

For the sake of completeness in the simulations, we initially examined the scenarios for both the possible situations: an improvement in process centering and a reduction in process variability. However, the results are very similar therefore in the paper we reported only the results concerning the improvement in the process centering. As an example, in Table A1 for the case $c_{pk,0} = 1.33$ and $\alpha = 0.02$ (the same as Table 4), the following are reported: $n_{0;\hat{\pi}_s > 0.80}$ the smallest maximal allowable sample size for the sequential test for achieving an empirical power $\hat{\pi}_s > 0.80$; $n_{avg;\mu}$ the average of the stopping sample sizes $n_{stop;\mu}$ required for the sequential test with maximal allowable sample size $n_{0;\hat{\pi}_s > 0.80}$ for concluding in favor of H_1 when process centering is improved; $\text{S.D.}(n_{stop;\mu})$ the standard deviation of the final sample sizes $n_{stop;\mu}$; $\hat{\pi}_{s;\mu}$ the estimated power of the sequential test when process centering is improved; $n_{avg;\sigma}$ the average of the stopping sample sizes $n_{stop;\sigma}$ required for the sequential test with maximal allowable sample size $n_{0;\hat{\pi}_s > 0.80}$ for concluding in favor of H_1 when the process variability is reduced; $\text{S.D.}(n_{stop;\sigma})$ the standard deviation of the final sample sizes $n_{stop;\sigma}$; $\hat{\pi}_{s;\sigma}$ the estimated power of the sequential test when the process variability is reduced $\hat{\pi}_{s;\sigma}$.

$c_{pk,1}$	$n_{0;\hat{\pi}_s>0.80}$	$n_{avg;\mu}$	S.D.$(n_{stop;\mu})$	$\hat{\pi}_{s;\mu}$	$n_{avg;\sigma}$	S.D.$(n_{stop;\sigma})$	$\hat{\pi}_{s;\sigma}$
1.60	171	116.1	29.8	0.811	115.7	29.3	0.815
1.70	96	64.7	17.2	0.815	64.8	16.9	0.816
1.80	62	41.7	11.2	0.809	42.0	11.3	0.813
1.90	44	29.4	8.4	0.820	29.6	8.4	0.818
2.00	33	21.8	6.6	0.814	22.1	6.5	0.805
2.10	26	17.1	5.5	0.815	17.2	5.5	0.807
2.20	21	13.7	4.7	0.808	13.6	4.7	0.812
2.30	17	11.0	3.9	0.807	10.9	3.9	0.802

Table A1: Results of the simulations for the case $c_{pk,0} = 1.33$, α=0.02 ($w_\alpha = 2.576$) obtained by improving process centering and with a reduction in process variability

It can be noted that the results of the two cases (improving the process centering vs reduction of the process variability) are very similar. For this reason in the paper we focused only on the scenarios concerning the improvement in process centering.

References

Borodin, A.B. and Salminen, P. (1996). "*Handbook of Brownian Motion-Facts and Formulae*". Birkhäuser Verlag, Basel.

Bothe, D.R. (1997). *Measuring Process Capability*. McGraw-Hill, New York.

Chen, M.S., Wu, M.S. and Lin, C. M. (2014). "Application of Indices C_p and C_{pk} to Improve Quality Control Capability in Clinical Biochemistry Laboratories". *Chinese Journal of Physiology* 57, 2, pp. 63-68.

Chen, S.M. and Hsu, Y.S. (2004). "Uniformly most powerful test for process capability index C_{pk}". *Quality Technology and Quantitative Management*, 1, 2. pp. 257-269.

Fan, S. and Kao, C. (2006). “Development of confidence interval and hypothesis testing for Taguchi capability index using a Bayesian approach”. *International Journal of Operations Research*, 3, 1, pp. 56-75.

Kane, V.E. (1986). “Process capability indices”. *Journal of Quality Technology*, 18, 1, pp. 41-52.

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

Kotz, S. and Johnson, N.L. (2002). “Process capability indices-a review, 1992–2000. *Journal of Quality Technology* 34, 1, pp.2-53.

Kotz, S, and Lovelace, C.R. (1998). *Introduction to Process Capability Indices: Theory and Practice*. Arnold, London.

Feller, W. (1970). *An Introduction to probability theory and its Applications Vol II*. John Wiley & Son, New York.

Hussein, A. (2005). “Sequential Comparison of Two Treatments Using Weighted Wald-Type Statistics”. *Communications in Statistics-Theory and Methods*, 34, 7, pp. 1631-1641.

Hussein, A., Ahmed, S.E. and Bhatti, S. (2012). “Sequential testing of process capability indices”. *Journal of Statistical Computation and Simulation*, 82, 2, pp.279-282.

Lepore, A. and Palumbo, B. (2015). “New Insights into the Decisional Use of Process Capability Indices via Hypothesis Testing”. *Quality and Reliability Engineering International*, 31, 8, pp.1725-1741.

Lin, G.H. (2006). “A hypothesis testing procedure on assessing process performance”. *Journal of Statistics and Management Systems*, 9, 2, pp. 319-339.

Liu, Y., Hourd, P., Chandra, A. and Williams, D.J. (2010). “Human cell culture process capability: a comparison of manual and automated production”. *Journal of Tissue Engineering and Regenerative Medicine*, 4, 1, pp. 45-54.

Montgomery, D. (2009). *Introduction to Statistical Quality Control (6th ed).* John Wiley & Sons, Hoboken, NJ.

Owen, D.B. (1965). "A Special Case of a Bivariate Non-Central t-Distribution". *Biometrika*, 52, 3/4, pp. 437-446.

Pearn, W.L. and Chen, K.S. (1999). "Making decision in assessing process capability index C_{pk}" *Quality and Reliability Engineering International*, 15, 4, pp. 321-326.

Pearn, W.L. and Kotz, S. (2006). *Encyclopedia and Handbook of Process Capability Indices*: *A Comprehensive Exposition of Quality Control Measures*. World Scientific, Singapore.

Pearn, W.L. and Lin, P.C. (2002). "Computer program for calculating the p-value in testing process capability index C_{pmk}". *Quality and Reliability Engineering International*, 18, 4, pp. 333-342.

Pearn, W.L. and Lin P.C. (2004). "Testing process performance based on capability index C_{pk} with critical values". *Computers and Industrial Engineering*, 47, 4, pp. 351-369.

Pearn, W.L., Shu, M.H. and Hsu, B.M. (2005). "Monitoring manufacturing quality for multiple Li-BPIC processes based on capability index C_{pmk}". *International Journal of Production Research*, 43, 12, pp. 2493-2512.

Perakis, M., and Xekalaki, E. (2003). "On a comparison of the efficacy of various approximations of the critical values for tests on the process capability indices C_{PL}, C_{PU}, C_{pk}". *Communications in Statistics-Simulation and Computation*, 32, 4, pp. 1249-1264.

Polansky A.M. and Kirmani, S.N.U.A. (2003). *Quantifying the capability of industrial processes*. In *Handbook of Statistics 22*, *ch. 17*. Khattree B., and Rao, C.R. (Eds.). Elsevier Science, Amsterdam.

R Core Team, (2013). *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Spiring, F. Leung, B. Cheng, S. and Yeung, A. (2003). "A bibliography of process capability papers". *Quality and Reliability Engineering International*, 19, 5, pp. 445-460.

Tartakovsky, A., Nikiforov, I. and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Chapman & Hall/CRC Press, Boca Raton.

Vännman, K. (2006). “Safety regions in process capability plots”. *Quality technology and quantitative management*, 3, 2, pp. 227-820.

Vännman, K. and Kulahci, M. (2008). “A model-free approach to eliminate autocorrelation when testing for process capability”. *Quality and Reliability Engineering International*, 24, 2, pp. 213-228.

Wheeler, D.J. (1999). *Beyond Capability Confusion*. SPC Press, Knoxville.

Yum, B.J. and Kim, K.W. (2011). “A bibliography of the literature on process capability indices: 2000-2009”. *Quality and Reliability Engineering International*, 27, 3, pp. 251-268.