Nonparametric Test on Process Capability

Stefano Bonnini

Abstract The study of process capability is very important in designing a new product or service and in the definition of purchase agreements. In general we can define capability as the ability of the process to produce conforming products or deliver conforming services. In the classical approach to the analysis of process capability, the assumption of normality is essential for the use of the indices and the interpretation of their values make sense but also to make inference on them. The present paper focuses on the two-sample testing problem where the capabilities of two processes are compared. The proposed solution is based on a nonparametric test. Hence the solution may be applied even if normality or other distributional assumptions are not true or not plausible and in the presence of ordered categorical variables. The good power behaviour and the main properties of the power function of the test are studied through Monte Carlo simulations.

Keywords Process capability · Permutation test · Two-sample test

1 Introduction

To ensure a high quality of product or service, the production process or service delivery process should be stable and a continuous quality improvement should be pursued. Control charts are the basic instruments for a statistical process control (SPC). One of the main goals of these and other statistical techniques consists in studying and controlling the capability of the process. A crucial aspect which should be studied and controlled is the process variability.

Every process, even if well-designed, presents a natural variability due to unavoidable random factors. In the presence of specific factors that cause systematic variability, the process is out of control and its performances are unacceptable. In these situations the process variability is greater than the natural variability and high per-

S. Bonnini (✉)
Department of Economics and Management, University of Ferrara,
Via Voltapaletto 11, Ferrara, Italy
e-mail: stefano.bonnini@unife.it

R. Cao et al. (eds.), *Nonparametric Statistics*, Springer Proceedings
in Mathematics & Statistics 175, DOI 10.1007/978-3-319-41582-6_2

centage of outputs (products, services, etc) could be nonconforming, that is the process would produce high percentages of waste. In other words, when the process is in control, most of the values of the response variable under monitoring falls between the specification limits. When the process is out of control, the probability that the response variable takes values outside the specification limits is high. Hence the main purpose of a SPC is to minimize the process variability.

The study of process capability is very important in designing a new product or service and in the definition of purchase agreements. In general we can define capability as the ability of the process to produce conforming products/services. In other words the greater the probability of observing values of the response in the interval [*LSL*, *USL*], the greater the process capability, where *LSL* and *USL* are lower specification limit and upper specification limit respectively.

In the statistical literature several works have been dedicated to process capability indices. For a deep discussion see, among the others, [5, 6, 9–11, 14, 15].

By assuming normality for the response, a simple way of measuring the process capability is based on the index

$$C_p = (USL - LSL)/(6\sigma), \tag{1}$$

where σ is the standard deviation of the response. For a non centred process, that is when the central tendency of the distribution of the response is not centred in the specification interval, a more appropriate measure of process capability is provided by

$$C_{pk} = \min[(USL - \mu), (\mu - LSL)]/(3\sigma), \tag{2}$$

where μ is the process mean. C_p can be considered as potential capacity of the process, while C_{pk} can be considered as actual capacity. When the process is centred $C_p = C_{pk}$. If $LSL \leq \mu \leq LSL$ then $C_{pk} \geq 0$ and when $\mu = LSL$ or $\mu = USL$ we have $C_{pk} = 0$.

The assumption of normality is essential for the use of the indices and the interpretation of their values make sense. Some approaches, proposed in the presence of non normal data, are based on a suitable transformation of data. Alternative solutions consist in defining general families of distributions like those of Pearson and Johnson (see [14]).

When the capabilities of two or more processes are compared, we should consider that a given value of C_{pk} could correspond to one process with centred mean and high variability or to another process with less variability and non centred mean. As a consequence, high values of C_{pk} may correspond to a non centred process with low variability. To take into account the centering of the process we should jointly consider C_p and C_{pk}. An alternative is represented by the following index of capability

$$C_{pkm} = (USL - LSL)/(6\sqrt{\sigma^2 + (\mu - T)^2}), \tag{3}$$

where T is the target value for the response. It is worth noting that $C_{pkm} = C_p/\sqrt{1+\theta^2}$, where $\theta = (\mu - T)/\sigma$.

Under the assumption of normality, it is possible to compute confidence intervals for the capability indices by means of point estimates of μ and σ. Common and very useful testing problems consider the null hypothesis $H_0 : C = C_0$ against the alternative $H_1 : C > C_0$, where C is a given index of capability and C_0 is a specific reference value for C (see for example [9]). We wish to focus on the two-sample testing problem where the capabilities of two processes, C_1 and C_2 are compared. The goal consists in testing the null hypothesis $H_0 : C_1 = C_2$ against the alternative $H_1 : C_1 > C_2$. Typical situations are related to the comparison between sample data drawn from a given process under study and sample data from an in-control process or to the comparison between the capabilities of the processes associated to different industrial plants, operators, factories, offices, corporate headquarters, etc. Some interesting contributions about capability testing are provided by [7, 8, 12, 13].

The proposal of the present paper is based on a nonparametric solution. Hence the test may be applied even if normality or other distributional assumptions are not true or not plausible. The method is based on a permutation test and neither requires distributional assumptions nor needs asymptotic properties for the null distribution of the test statistic. Hence, it is a very robust procedure and can also be applied for small sample sizes and for ordered categorical data.

The basic idea is to transform the continuous response variable into a categorical variable through a suitable transformation of the support of the original response into a set of disjoint regions and to perform a test for comparing the heterogeneities of two categorical distributions. In Sect. 2 the procedure is described. Section 3 presents the results of a simulation study for proving the good power behaviour of the proposed test. Final conclusions are given in Sect. 4.

2 Permutation Test on Capability

Let X be a continuous random variable representing the response under study in the SPC. The probability that X takes values in the region $R \in \Re$ is

$$\pi_R = P[X \in R] = \int_R f(x)dx, \tag{4}$$

where $f(x)$ is the (unknown) density function of X. Let us define $R_T = [LSL, USL]$ the target region for X, $R_L = (-\infty, LSL)$ and $R_U = (USL, +\infty)$. A reasonable assumption, unless the process is severely out of control, is that most of the probability mass is concentrated in R_T, i.e., the probability that X falls in the target region is greater than the probability than X takes values in the lower tail or in the upper tail. Formally

$$\pi_{R_T} = \max(\pi_{R_L}, \pi_{R_T}, \pi_{R_U}), \tag{5}$$

with $\pi_{R_L} + \pi_{R_T} + \pi_{R_U} = 1$. The ideal situation, when the process is in control, is that the probability of producing waste is null, that is $\pi_{R_L} = \pi_{R_U} = 0$ and $\pi_{R_T} = 1$. The worst situation, when π_{R_T} takes its absolute minimum under the constrain defined in Eq. 5, consists in the uniform distribution, where $\pi_{R_L} = \pi_{R_T} = \pi_{R_U} = 1/3$. Hence a suitable index of capability could be the one's complement of a normalized measure of heterogeneity for categorical variables. A solution could be based on the use of the index of *Gini*

$$C^{(G)} = 1 - (3/2)[1 - (\pi_{R_L}^2 + \pi_{R_T}^2 + \pi_{R_U}^2)]. \tag{6}$$

The famous entropy of *Shannon* may be also considered for computing a normalized index of capability

$$C^{(S)} = 1 + (\pi_{R_L} \ln \pi_{R_L} + \pi_{R_T} \ln \pi_{R_T} + \pi_{R_U} \ln \pi_{R_U})/\ln 3. \tag{7}$$

Other alternatives can be provided by the family of indices proposed by *Rényi*

$$C^{(\omega)} = 1 - (1-\omega)^{-1} \ln(\pi_{R_L}^{\omega} + \pi_{R_T}^{\omega} + \pi_{R_U}^{\omega})/\ln 3. \tag{8}$$

Each normalized index of heterogeneity takes value 1 in case of maximum heterogeneity (uniform distribution), value 0 in case of minimum heterogeneity (degenerate distribution) and greater values when moving from the degenerate towards the uniform distribution (see [4]). Hence the greater the value of the index of heterogeneity the lower the capability of the process because the capability is non decreasing function of the probability concentration. For this reason, if the probabilities were known, the comparison of two process capabilities could be done by comparing the cumulative ordered probabilities $\Pi_{1(s)} = \sum_{t=1}^{s} \pi_{1(t)}$ and $\Pi_{2(s)} = \sum_{t=1}^{s} \pi_{2(t)}$ with $\pi_{jR_T} = \pi_{j(1)} \geq \pi_{j(2)} \geq \pi_{j(3)}$, $j = 1, 2$, $s = 1, 2, 3$. Thus the hypotheses of the problem are

$$H_0 : [C_1 = C_2] \equiv [\Pi_{1(s)} = \Pi_{2(s)} \forall s], \tag{9}$$

and

$$H_1 : [C_1 > C_2] \equiv [\Pi_{1(s)} \geq \Pi_{2(s)} \forall s \text{ and } \exists s \text{ s.t. } \Pi_{1(s)} > \Pi_{2(s)}]. \tag{10}$$

Under the null hypothesis, when the cumulative ordered probabilities are equal, exchangeability holds. But $\pi_{j(t)}$, $j = 1, 2$, $t = 1, 2, 3$ are unknown parameters of the distribution and need to be estimated by using the observed ordered frequencies $\hat{\pi}_{j(t)} = n_{j(t)}/n_j$, where $n_{j(t)}$ is the tth ordered absolute frequency for the j-th sample and n_j is the size of the j-th sample. Hence the real ordering of the probabilities is estimated and the exchangeability under H_0 is approximated and not exact.

[1, 3] suggest that a test statistic for the similar problem of two-sample test on heterogeneity may be based on the difference of the sampling estimates of the indices of heterogeneity. By adapting this approach to our specific problem, we suggest to

use, as test statistic, the difference of the sampling estimates of the process capabilities under comparison: $T = \hat{C}_1 - \hat{C}_2$, where $\hat{C}_j$ is computed like C_j but by replacing $\pi_{j(t)}$ with $\hat{\pi}_{j(t)}$, with $j = 1, 2$ and $t = 1, 2, 3$. Hence we have $T_G = \hat{C}_1^{(G)} - \hat{C}_2^{(G)}$, $T_S = \hat{C}_1^{(S)} - \hat{C}_2^{(S)}$ and $T_{R_\omega} = \hat{C}_1^{(\omega)} - \hat{C}_2^{(\omega)}$.

An alternative solution could be based on the combination of more than one statistic, by considering the information provided by different indices. For example, according to the additive combining rule, we have

$$T_C = T_G + T_S + T_{R_3} + T_{R_\infty}, \tag{11}$$

where T_{R_3} and T_{R_∞} are the test statistics based on the indices of *Rényi* of order 3 and ∞ respectively. Whatever the statistics used for the problem, the null hypotheses must be rejected for large values of this statistic.

The first step of the testing procedure consists of the computation of the observed ordered table, that is $\{n_{j(t)}; j = 1, 2; t = 1, 2, 3\}$ and the observed value of the test statistic $T^{(0)}$. By performing B independent permutations of the dataset, then obtaining B permuted ordered tables $\{n^*_{j(t)}; j = 1, 2; t = 1, 2, 3\}$ and B corresponding permutation values of the test statistic $T^{*(1)}, \ldots, T^{*(B)}$, the p-value, according to the permutation distribution, can be computed as

$$p = \sum_{b=1}^{B} I(T^{*(b)} \geq T^{(0)})/B, \tag{12}$$

where $I(E) = 1$ iff the event E is true, and $I(E) = 0$ otherwise. An alternative resampling strategy may be based on a bootstrap approach but [2] proves that this solution is usually not as powerful as the permutation one.

3 Monte Carlo Simulation Study

To analyze the power behaviour of the proposed tests, a Monte Carlo simulation study was performed. Data for the j-th sample were randomly generated by the following variable:

$$X_j = 1 + int[3U^{\gamma_j}], \tag{13}$$

where U is a uniform random variable, and $\gamma_j \in (0, 1]$ is the heterogeneity parameter: the greater γ_j the higher the heterogeneity of X_j (thus the lower C_j), hence $C_1 > C_2$ iff $\gamma_1 < \gamma_2$. For each specific setting, defined in terms of γ_1, γ_2, n_1 and n_2 values, $CMC = 1000$ datasets where generated and, for each dataset, $B = 1000$ permutations were performed to estimate the p-values and compute the rejection rates of the tests. The estimated power (rejection rates) of the tests on capability based on the indices of *Gini*, *Shannon*, *Rényi* (order 3 and order ∞) and on the direct (additive) combination of the four mentioned tests were computed.

Table 1 Simulation results under $H_0 : C_1 = C_2$, $n_1 = n_2 = 50$, $\alpha = 0.05$, $B = 1000$, $CMC = 1000$

			Rejection rates				
γ_1	γ_2	$\gamma_1 - \gamma_2$	T_G	T_S	T_{R_3}	T_{R_∞}	T_C
0.2	0.2	0.0	0.055	0.054	0.055	0.055	0.055
0.4	0.4	0.0	0.053	0.052	0.053	0.054	0.054
0.6	0.6	0.0	0.040	0.048	0.040	0.025	0.040

Table 2 Simulation results under $H_1 : C_1 > C_2$, $n_1 = n_2 = 50$, $\alpha = 0.05$, $B = 1000$, $CMC = 1000$

			Rejection rates				
γ_1	γ_2	$\gamma_1 - \gamma_2$	T_G	T_S	T_{R_3}	T_{R_∞}	T_C
0.8	1.0	0.2	0.044	0.042	0.047	0.060	0.052
0.6	1.0	0.4	0.375	0.399	0.372	0.314	0.360
0.4	1.0	0.6	0.945	0.955	0.933	0.863	0.935

In Table 1, the rejection rates under the null hypothesis of equality in capability are reported with samples sizes equal to 50. Three different capability levels are considered. The powers of all the tests seem to increase with the capability: as a matter of fact capability is negatively related to heterogeneity, hence lower capability implies greater heterogeneity and greater heterogeneity means greater uncertainty. Table 1 shows that all the tests are well approximated, because the rejection rates are very similar to the nominal α level, even if, in the presence of high capabilities, the tests tend to be slightly anticonservative. The test based on the *Rényi* index of order ∞ is less stable than the others because of its very low power in the presence of low capabilities.

Table 2 shows the estimated power of the tests under H_1, when the capability of the second process is at the minimum level and for three different capability levels of the first process. As expected, the greater the difference in capability, the greater the power of the tests. When the difference in capability is low, the most powerful tests are those based on the direct combination and on the *Rényi* index of order ∞. Instead, when the difference in capability is high, the latter test is the less powerful, the power performance of the others is similar and the test based on the *Shannon* index is slightly preferable.

In Table 3 the behaviour of the rejection rates as function of the sample sizes, when the parameter difference is equal to 0.4, can be appreciated. The consistency of the tests is evident because larger sample sizes correspond to higher power. Again the power behaviours of the tests are very similar and, for small sample sizes, the test based on the *Rényi* index of order ∞ is the most powerful but for large sample sizes this test is the less powerful.

Table 4 focuses on the power comparison of the tests for different sample sizes when the difference between the capabilities is small. Even in this case, the test based on *Rényi* index of order ∞ is the best in the presence of small sample sizes

Table 3 Simulation results under $H_1 : C_1 > C_2,\ \gamma_1 = 0.6,\ \gamma_2 = 1.0,\ \alpha = 0.05,\ B = 1000,$ $CMC = 1000$

			Rejection rates				
n_1	n_2	$n_1 - n_2$	T_G	T_S	T_{R_3}	T_{R_∞}	T_C
20	20	0	0.109	0.109	0.108	0.114	0.108
60	60	0	0.440	0.455	0.424	0.365	0.426
100	100	0	0.759	0.769	0.742	0.640	0.740

Table 4 Simulation results under $H_1 : C_1 > C_2,\ \gamma_1 = 0.8,\ \gamma_2 = 1.0,\ \alpha = 0.05,\ B = 1000,$ $CMC = 1000$

			Rejection rates				
n_1	n_2	$n_1 - n_2$	T_G	T_S	T_{R_3}	T_{R_∞}	T_C
20	20	0	0.019	0.019	0.020	0.028	0.019
40	40	0	0.046	0.050	0.045	0.048	0.054
60	60	0	0.052	0.052	0.048	0.058	0.058
100	100	0	0.109	0.113	0.104	0.109	0.109

and this is not true in the presence of large sample sizes. In the intermediate case of sample sizes equal to 40, the most powerful test seems to be the one based on direct combination.

4 Conclusions

The two-sample nonparametric test on process capability is a robust solution and allows inferential comparative analysis of process capabilities even when distributional assumptions (e.g., normality) do not hold or cannot be tested. Under the null hypothesis of equality in heterogeneity, data exchangeability is not exact but the good approximation of the permutation test is proved by the Monte Carlo simulation study.

According to this proposal, the test statistic is based on the comparison of the two-sample heterogeneities, computed by using suitable indices of heterogeneity, like the *Gini* index, the *Shannon* entropy, the *Rényi* family of indices, or a suitable combination of test statistics based on different indices, for example on the sum of these different test statistics.

The Monte Carlo simulation study proves that the power of all the tests seems to increase with the capability: as a matter of fact capability is negatively related to heterogeneity, hence lower capability implies greater heterogeneity and consequently greater uncertainty.

All the considered tests are well approximated, because under the null hypothesis of equality in capability, the rejection rates are very similar to the nominal α level.

Under the alternative hypothesis, when the difference in capability is low, the most powerful tests are those based on the direct combination and on the *Rényi* index of order ∞. Instead, when the difference in capability is high, the latter test is the less powerful and the test based on the *Shannon* index is slightly preferable.

The tests are consistent because if sample sizes increase then power increases. For small sample sizes the test based on the *Rényi* index of order ∞ is the most powerful but for large sample sizes it is the less powerful. In the presence of small difference in the capabilities of the two compared processes, again the test based on the *Rényi* index of order ∞ is the best in the presence of small sample sizes but not in the presence of large sample sizes. In the case of intermediate sample sizes, the test based on the direct combination seems to be the most powerful. Hence, if we consider the instability of the *Rényi* index of order ∞, the test based on the direct combination is the best solution under the alternative hypothesis, when it is difficult to detect the difference in the capabilities of the two processes, i.e., near the null hypothesis.

References

1. Arboretti, G.R., Bonnini, S., Pesarin, F.: A permutation approach for testing heterogeneity in two-sample problems. Stat. Comput. **19**, 209–216 (2009)
2. Bonnini, S.: Testing for heterogeneity for categorical data: permutation solution vs. bootstrap method. Commun. Stat. A-Theor. **43**(4), 906–917 (2014)
3. Bonnini, S.: Combined tests for comparing mutabilities of two populations. In: Topics in Statistical Simulation. Book of Proceedings of the Seventh International Workshop on Simulation 2013, Rimini, 21–25 May 2013, pp. 67–78. Springer, New York (2014)
4. Bonnini, S., Corain, L., Marozzi, M., Salmaso, L.: Nonparametric Hypothesis Testing: Rank and Permutation Methods with Applications in R. Wiley, Chichester (2014)
5. Boyles, R.A.: The Taguchi capability index. J. Qual. Technol. **23**, 17–26 (1991)
6. Chan, L.K., Cheng, S.W., Spiring, F.A.: A new measure of process capability: Cpm. J. Qual. Technol. **20**, 162–175 (1988)
7. Chen, J.P., Tong, L.I.: Bootstrap confidene interval of the difference between two process capability indices. Int. J. Adv. Manuf. Tech. **21**, 249–256 (2003)
8. Choi, Y.M., Polansky, A.M., Mason, R.L.: Transforming non-normal data to normality in statistical process control. J. Qual. Technol. **30**(2), 133–141 (1998)
9. Kane, V.E.: Process capability indices. J. Qual. Technol. **18**, 41–52 (1986)
10. Kotz, S., Johnson, N.L.: Process Capability Indices. Chapman & Hall, London (1993)
11. Pearn, W.L., Kotz, S., Johnson, N.L.: Distributional and inferential properties of process capability indices. J. Qual. Technol. **24**, 216–231 (1992)
12. Pearn, W.L., Lin, P.C.: Testing process performance based on capability index C_{pk} with critical values. Comput. Ind. Eng. **47**, 351–369 (2004)
13. Polansky, A.M.: Supplier selection based on bootstrap confidence regions of process capability indices. Int. J. Rel. Qual. Saf. Eng. **10**, 1 (2003). doi:10.1142/S0218539303000968
14. Rodriguez, R.N.: Recent developments in process capability analysis. J. Qual. Technol. **24**, 176–187 (1992)
15. Vannman, K.: A unified approach to capability indices. Stat. Sin. **5**, 805–820 (1995)

http://www.springer.com/978-3-319-41581-9

Nonparametric Statistics
2nd ISNPS, Cádiz, June 2014
Cao, R.; González Manteiga, W.; Romo, J. (Eds.)
2016, XI, 224 p. 40 illus., 19 illus. in color., Hardcover
ISBN: 978-3-319-41581-9