Received: Added at production | Revised: Added at production | Accepted: Added at production
DOI: xxx/xxxx

ARTICLE TYPE

Objective Bayesian Inference for the Capability Index of the Gamma Distribution

Marcello Henrique de Almeida*[1] | Pedro Luiz Ramos[2] | Gadde Srinivasa Rao[3] | Fernando Antonio Moala[1]

[1]Department of Statistics, State University of Sao Paulo, Presidente Prudente, São Paulo, Brazil
[2]Institute of Mathematical and Computer Sciences, University of São Paulo, São Carlos, São Paulo, Brazil
[3]Department of Mathematics and Statistics, University of Dodoma, Dodoma, Tanzania

Correspondence
*Corresponding author name: Pedro L. Ramos Email: pedrolramos@usp.br

Summary

The Gamma distribution has been applied in research in several areas of knowledge, due to its good flexibility and adaptability nature. Process capacity indices like C_{pk} are widely used when the measurements related to the data follow a normal distribution. This article aims to estimate the C_{pk} index for non-normal data using Gamma distribution. We discuss maximum likelihood estimation and a Bayesian analysis through the Gamma distribution using an objective prior, known as matching prior that can return Bayesian estimates with good properties for the C_{pk}. A comparative study is made between classical and Bayesian estimation. The proposed Bayesian approach is considered with the Markov Chain Monte Carlo method to generate samples of the posterior distribution. A simulation study is carried out to verify whether the posterior distribution presents good results when compared with the classical approach in terms of the mean relative errors and the mean square errors, which are the two commonly used metrics to evaluate the parameter estimators. Based on the real data set, Bayesian estimates and credibility intervals for unknown parameters and the prior distribution are achieved to verify if the process is under control.

KEYWORDS:
Objective Bayesian inference, matching prior, Process capacity index, C_{pk}.

1 | INTRODUCTION

Process capability indices are employed to appraise management aspiration to raise its earnings potential and at the same time to compete with existing company's by focus on plummeting the variability in the manufacturing process. Also, at the same time produced items should meet the company's specifications and the customer's anticipation. To estimate the amount of process capability of the manufacturing process, it is required to characterize a quantitative measure of the product that justifies the performance of the company. Sometimes the product quality can be assessed by sampling inspection whereas it may not interpret quality specifications laid down by the company or customer's anticipation. To analyze quality of the product in manufacturing technology by using process capability analysis can be measured through various stages of manufacturing process includes process, product design, manufacturing and manufacturing planning, for more details see Statisti and Tehnike[1]. A number of process capability indices have been proposed to determine the manufactured product is capable or not and also product meets the specifications given by the company or satisfied by the customer. More literature review on process capability indices can be checked in Kotz and Johnson[2], Spiring et al.[3], Sappakitkamjorn and Niwitpong[4] and Yum and Kim[5] and Piña-Monarrez et al[6]. Zhang[7] rightly pointed out that the two most frequently used indices are among the available numerous process capability indices.

To estimate these process capability indices the basic assumption on quality characteristics under study should follows normal distribution. Whereas for industrial production is concern most of quality characteristics may not follows normal distribution or quality characteristics distribution is not known, in these situations the process capability indices are not accurate, for more details see Sennaroglu and Senvar[8], Piña-Monarrez et al.[6], Senvar and Kahraman[9,10]. Hosseinifard et al.[11] studied the effect of non-normal data to estimate the process capability indices and assessed that data lead to inaccurate results. For non-normally distributed data, a quantile approach to measure the parametric estimation of process capability indices were developed by Clements[12]. Hence, more researchers focused on the study of the process capability analysis for the non-normal data, for more details refer Kashif et al[13], Panichkitkosolkul[14], Leiva et al.[15], Senvar and Sennaroglue[16], Saha et al.[17,18], Meng et al.[19], Mahmoud et al[20], and Wu[21].

An important non-normal distribution that has been widely considered in reliability is the Gamma distribution. The inference for the parameter of the model are standard in most statistical books. Under the classical approach the maximum likelihood estimator (MLE) is usually considered due to its good properties such as consistency, efficiency and invariant under one-to-one transformations. On the other hand, this method has two drawbacks, the first is that the method is usually biased for small samples, and secondly due to the mathematical form of the capability index it is very difficulty to apply the delta method to obtain the confidence intervals of the estimates. In order to overcome these obstacles we can apply the objective Bayesian methods. Previously, several researchers discussed procedures to conduct Bayesian inference for the parameters of the Gamma distribution (see Louzada and Ramos[22], Miller[23], Sun and Ye.[24] and Berger et al.[25]). Ramos et al.[26] compared the effect of the different objective priors have in the posterior distribution and concluded that in terms of bias, mean square error and coverage probabilities the matching prior was the most appropriate.

In this article, we consider a fully objective Bayesian analysis to estimate the C_{pk} measure. Due to the complex form of the index, we obtain the posterior distribution using the invariant property of the Markov Chain Monte Carlo (MCMC) methods. The Metropolis-Hasting algorithm is considered to obtain the chains of the parameters as well as the C_{pk}. Further, we show that the results obtained from objective inference are superior to the ones obtained through maximum likelihood. From the obtained results, we can achieve Bayes estimates with smaller bias and construct precise credibility intervals from the posterior density. The proposed approach is applied to check if the control process of powdered juice company is under control or if requires improvements to decrease losses.

The remainder of this paper is organized as follows. Section 2 presents mathematical properties for the Gamma distribution and the C_{pk} measures. Section 3 provides the maximum likelihood estimates. Section 4 and 5 presents the steps to achieve the Bayesian inference assuming an objective prior and the MCMC algorithm to sample from the posterior distribution. In Section 6, we present a simulation study to illustrate the usefullness of our proposed approach. Section 7 presents an analysis of an important data set. Finally, in Section 8 summarizes the study.

2 | BACKGROUND

In this section, we provide the background necessary. Firstly we revisit the properties of the Gamma distribution and further, we provide a brief introduction of process capability index. Hence, let X be a random variable with Gamma distribution where its probability distribution is defined by

$$f(x\,|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \tag{1}$$

whose parameters are $\alpha > 0$ and $\beta > 0$ and $\Gamma(\phi) = \int_0^{\infty} e^{-x} x^{\phi-1} dx$ is the gamma function. The mean and variance of the gamma distribution is, respectively given by

$$\mathbb{E}(\mathbb{X}) = \int_0^{\infty} \frac{\beta^{\alpha} x^{\alpha} e^{-\beta x}}{\Gamma(\alpha)} dx = \frac{\alpha}{\beta} \tag{2}$$

$$\mathbb{V}ar(\mathbb{X}) = \int_0^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha+1} e^{-\beta x} dx - \mathbb{E}(\mathbb{X})^2 = \frac{\alpha}{\beta^2}. \tag{3}$$

This distribution belongs to the exponential family of continuous probability distributions with two parameters. In the literature, there are reparametrized version of this model, hereafter, we will use the form given in (1). The Gamma distribution is an important generalization of the exponential distribution ($\alpha = 1$), assuming that $\alpha = \frac{n}{2}$, $n \in \mathbb{Z}$ and $\beta = \frac{1}{2}$ one can show that the model reduces to the chi-square distribution with n degrees of freedom.

Since the Gamma distribution is used to describe non-negative data with positive skewness, the model is widely used in survival and reliability analysis (see, for example, Lawless 1982), as it is a model that is adaptable and returns, in most cases, a good adjustment for the event of interest, for example, in the areas of engineering, meteorology, climatology, among other situations.

Another important topic that is usually considered in reliability analysis is related to quantitative measures considered in the manufacturing processes that are often used to measure its reliability, which is known as process capability index. The PCI is usually a measure that is achieved through the percentiles of the target distribution and the appropriate specifications for the process. The specifications are given by the lower specification limit (LSL) and upper specification limit (USL) and can be established according to the interest of the researcher or through regulatory quality control agencies, as we will see in the application section. There are several indexes available in the literature, among them, we can mention some of the commonly used indices in statistical quality control which are referred to as C_p and C_{pk}.

Process capabilities indices are tools with the capacity to process information where it is possible to assess whether a particular process is capable of generating products that are within the required specifications. The indexes C_p and C_{pk} appear in the literature after studies presented by the engineer, physicist, with important contributions of the statistician Walter Shewhart during his work with Statistical Process Control.

The first index proposed in the literature was C_p. This index considers that the process is always centered around the average. Here, the width of the process amplitude is evaluated in relation to the width of the specification given in the experiment. As C_p does not take the process average into account, therefore, it will not provide relevant information on whether the process is centralized[27]. In practice, this result is important, later Kane[28] released the C_{pk} index which returned improve results compared to the previous index and it is applied to processes that are not centered around the average of the specification range. According to Silveira[29], this improved index measures the distance between the specification limit closest to the average from the quality characteristic of interest.

For the application of these two methods given above, it is necessary to fulfill two requirements: the process must be under statistical control and a characteristic of the process of interest must be distributed normally. However, the data not always follow a normal distribution. For different processes with non-normal distributions, only the mean and standard deviation may not be sufficient to generate the characteristics of a given product. In order to overcome the insufficiency presented by the C_p index, the C_{pk} was firstly defined for normally distributed data and can be computed from

$$C_{pk} = \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right), \tag{4}$$

where μ and σ represent the mean and standard deviation of the distribution normal, respectively.

Further, Clements then proposed another method for estimating the process capacity index. Also referred to as C_{pk} this index is computed for data sets that follow non-normal distribution, and become one of the most popular approaches for this purpose (Kotz and Johnson,[2]). The method employs both a new percentile-based process capability index and a distribution fit approach. The index can be calculated given by:

$$C_{pk} = \min\left(\frac{USL - M}{U_p - M}, \frac{M - LSL}{M - L_p}\right), \tag{5}$$

where U_p, L_p and M are the percentiles 99.865^{th}, 0.135^{th} and 50^{th} from the Gamma distribution, respectively, USL and LSL denote the upper and lower specification limits. It is important to point out that the percentiles U_p, L_p and M mentioned above, can be easily computed from a function already implemented in the statistical software R.

3 | MAXIMUM LIKELIHOOD ESTIMATION

The maximum likelihood function is one of the most commonly used estimation procedures in frequentist inference, the method is usually considered due to its flexibility in construct the likelihood function, that contains all the information of θ obtained from the data. The MLEs have important properties such as invariance, consistency, and efficiency. Moreover, due to its asymptotic properties, we may be able to construct confidence intervals for the parameters of interest. Let $X_1, ..., X_n$ a random sample of size n of the random variable give in (1) then the α and β likelihood function is give by

$$L(\alpha,\beta|\boldsymbol{x}) = \frac{\beta^{n\alpha}}{[\Gamma(\alpha)]^n}\left\{\prod_{i=1}^{n} x_i^{\alpha-1}\right\}\exp\left\{-\beta\sum_{i=1}^{n} x_i\right\}. \tag{6}$$

Given that the maximum likelihood of θ, $\theta = \{(\alpha,\beta);\alpha > 0,\beta > 0\}$, which is represented by $\hat{\theta}$ that maximize $L(\alpha,\beta|\boldsymbol{x})$, it may be difficult to direct maximize $L(\cdot)$ and as the logarithm of the likelihood function $l(\alpha,\beta|\boldsymbol{x})$ is a growing function, we have that maximize $L(\alpha,\beta|\boldsymbol{x})$ is equivalent to maximizing its logarithm. Hence, in this case the logarithm of the likelihood function is given by

$$l(\alpha,\beta|\boldsymbol{x}) = n\alpha\log\beta - n\log\Gamma(\alpha) + (\alpha-1)\sum_{i=1}^{n}\log x_i - \beta\sum_{i=1}^{n} x_i. \tag{7}$$

From

$$\frac{\partial\log L(\alpha,\beta|\boldsymbol{x})}{\partial\alpha} \quad \text{and} \quad \frac{\partial\log L(\alpha,\beta|\boldsymbol{x})}{\partial\beta},$$

the score functions of α and β are given by for

$$\frac{\partial Log L(\alpha,\beta|\boldsymbol{x})}{\partial\alpha} = n\log(\beta) - n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} + \sum_{i=1}^{n}\log(x_i) = 0 \tag{8}$$

$$\frac{\partial Log L(\alpha,\beta|\boldsymbol{x})}{\partial\beta} = \frac{n\alpha}{\hat{\beta}} - \sum_{i=1}^{n} x_i = 0. \tag{9}$$

Performing some changes algebraically in (8) and (9), we find the maximum likelihood estimators of α and β given, respectively, by

$$\log(\hat{\alpha}) - \psi(\hat{\alpha}) = \log(\overline{X}) - \frac{1}{n}\sum_{i=1}^{n}\log(x_i) \tag{10}$$

$$\hat{\beta} = \frac{\hat{\alpha}}{\overline{X}} \tag{11}$$

where $\psi(\hat{\alpha}) = \dfrac{\partial\log\Gamma(\alpha)}{\partial\alpha} = \dfrac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})}$ and $\overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$.

There are situations, mainly, in the case of more complex models, the maximum likelihood estimators does not have closed-form expressions. In this case, the maximum likelihood estimators must be solved using numerical methods. There is also the possibility of making some modifications to the maximum likelihood estimator that does not have closed-form expression that lead to estimators with closed-form[30]. Through an important result that guarantees that if the derivatives up to the second order of the likelihood function are defined and the fisher information matrix is not null, then the distribution of $\hat{\alpha}$ and $\hat{\beta}$ tends to a normal distribution with mean α, β and equal variance to the elements I_{11} and I_{22} of the matrix $I(\alpha,\beta)^{-1}$, respectively. The Fisher information matrix is given by

$$I(\alpha,\beta) = \begin{bmatrix} \psi'(\alpha) & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{bmatrix},$$

where $\psi'(\hat{\alpha}) = \dfrac{\partial^2\log\Gamma(\alpha)}{\partial\alpha^2}$ is known as the trigamma function.

4 | BAYESIAN INFERENCE

Here, we will consider that the vector of parameters θ=(α, β) are treated as random variables. The distribution that represents the information about θ before the data is observed is known as prior distribution and represented by $\pi(\theta)$. Given a sample, all the information contained in the sample can be represented by the likelihood function, and after the experiment is conducted the distribution of the parameters given the observed data is achieved through Bayes' theorem and refereed as posterior distribution, $\pi(\theta|x)$. As there are many ways to obtain the prior distribution we considered the matching prior suggested by Ramos et al.[26], its construction is described below.

An important objective prior was introduced by Tibshirani[31] assuming that for a prior distribution $\pi(\theta_1, \theta_2)$ where the parameter of interest is θ_1 and θ_2 is a nuisance parameter, the credibility interval for that parameter in has a convergence error $O(n^{-1})$ in the frequentist sense, that is,

$$P[\theta_1 < \theta_1^{1-\alpha}(\pi, X)|(\theta_1, \theta_2)] = 1 - \alpha - O(n^{-1}). \tag{12}$$

The class of priors that satisfies the condition above is usually referred to as matching priors with a convergence order of $O(n^{-1})$. Mukerjee and Dey[32] further discussed necessary and sufficient conditions for this class matching priors also be a matching prior but with a convergence of order $o(n^{-1})$. For the Gamma distribution, Sun and Ye[24] showed that the matching prior of order $o(n^{-1})$ when β is the parameter of interest is given by

$$\pi(\alpha, \beta) \propto \frac{\alpha\psi'(\alpha) - 1}{\beta\sqrt{\alpha}}.$$

As discussed earlier this prior is one of the many possible objective priors for the parameters of the Gamma distribution. On the other hand, Ramos et al.[26] showed that the cited prior returned a posterior distribution that provided excellent posterior estimates in terms of smaller bias and MSE. Moreover, the obtained posterior has precise coverage probabilities for both parameters and important properties such as invariance under one-to-one transformation, hence, we will consider such prior to performing the inference for the C_{pk} index.

Through the matching prior we obtain the posterior joint distribution for α and β.

$$\pi(\alpha, \beta|\boldsymbol{x}) \propto \frac{(\alpha\psi'(\alpha) - 1)}{\sqrt{\alpha}} \frac{\beta^{n\alpha-1}}{[\Gamma(\alpha)]^n} \left\{\prod_{i=1}^{n} x_i^{\alpha-1}\right\} \exp\left\{-\beta\sum_{i=1}^{n} x_i\right\}. \tag{13}$$

The matching prior is improper and may lead to improper posterior. However, Ramos et al.[26] proved that the posterior distribution (13) is proper if and only if $n \geq 2$ with $n \in \mathbb{N}$ as well as its higher moments (posterior mean, variance among others). Therefore we can use such posterior to conduct the Bayesian inference.

By integrating (13) with respect to the parameter β we find the later marginal distribution for α which is given by

$$\pi(\alpha|\boldsymbol{x}) \propto \frac{(\alpha\psi'(\alpha) - 1)}{\sqrt{\alpha}} \frac{\Gamma(n\alpha)}{\Gamma(\alpha)^n} \left(\frac{\sqrt[n]{\prod_{i=1}^{n} x_i}}{\sum_{i=1}^{n} x_i}\right)^{n\alpha}. \tag{14}$$

The conditional posterior distribution for β is

$$\pi(\beta|\alpha, \boldsymbol{x}) \sim \text{Gamma}\left(n\alpha, \sum_{i=1}^{n} x_i\right). \tag{15}$$

5 | MARKOV CHAIN MONTE CARLO METHOD

Extracting information from the posterior distributions is usually not an easy task. The joint posterior distribution follows an unknown model and the normalized constant is difficult to be computed. In order to extract the information of interested we can considered the Metropolis-Hastings (MH) algorithm to sample from the posterior distribution. The algorithm is used as follows:

1. Select an initial value for $\alpha^{(1)}$ and take $j = 1$ as the loop counter iteration.
2. Generate a candidate $\alpha^{(*)}$ from the proposed distribution Gamma$(\alpha^{(j)}, b)$;
3. Calculate the probability of acceptance given by
$$\lambda\left(\alpha^{(j)}, \alpha^{(*)}\right) = \min\left(1, \frac{p\left(\alpha^{(*)}|\boldsymbol{x}\right)}{p\left(\alpha^{(j)}|\boldsymbol{x}\right)} \frac{\mathrm{q}\left(\alpha^{(j)}, \alpha^{(*)}, b\right)}{\mathrm{q}\left(\alpha^{(*)}, \alpha^{(j)}, b\right)}\right),$$
where $p(\cdot)$ is the marginal posterior distribution given in (14). Select a random value u from the uniform distribution in the interval of $(0, 1)$;
4. If $\lambda\left(\alpha^{(j)}, \alpha^{(*)}\right) \geq u(0, 1)$ then accept the new value $\alpha^{(j+1)} = \alpha^{(*)}$, otherwise reject it and consider $\alpha^{(j+1)} = \alpha^{(j)}$;

5. Generate a random value $\beta^{(j+1)}$ from the conditional posterior distribution for β given by
$$\beta^{(j+1)} \sim \text{Gamma}\left(n\alpha^{(j+1)}, \sum_{i=1}^{n} x_i\right).$$
6. Let
$$C_{pu} = \frac{USL - M(\alpha^{(j+1)}, \beta^{(j+1)})}{U_p(\alpha^{(j+1)}, \beta^{(j+1)}) - M(\alpha^{(j+1)}, \beta^{(j+1)})}$$
$$C_{pl} = \frac{M(\alpha^{(j+1)}, \beta^{(j+1)}) - LSL}{M(\alpha^{(j+1)}, \beta^{(j+1)}) - L_p(\alpha^{(j+1)}, \beta^{(j+1)})},$$
where $\mathrm{U}_p(\alpha^{(j+1)}, \beta^{(j+1)})$, $\mathrm{L}_p(\alpha^{(j+1)}, \beta^{(j+1)})$ e $\mathrm{M}(\alpha^{(j+1)}, \beta^{(j+1)})$ are the percentiles given respectively by 99.865, 0.135 e 50 obtained from the gamma distribution.
7. Compute
$$C_{pk}^{(j+1)} = \min\left(C_{pu}, C_{pl}\right).$$
8. Increase the counter from j to $j+1$. The process must be repeated until a specified number of iterations is achieved.

It it is important to mention that when using marginal posterior distribution of α, there is significant difficulty in calculating the ratio
$$\frac{p\left(\alpha^{(*)}|\alpha^{(t)}, \boldsymbol{x}\right)}{p\left(\alpha^{(t)}|\alpha^{(*)}, \boldsymbol{x}\right)},$$
this impasse can be explained by the fact that $p\left(\alpha^{(*)}|\alpha^{(t)}, \boldsymbol{x}\right)$ has the terms $\Gamma(n\alpha)$ and $(\Gamma(\alpha)^n)^{-1}$ and depending on the values of ϕ and n the software may not compute such terms. For instance, consider $\alpha = 3$ and assuming that $n \geq 100$, the R will not be able to calculate such terms, returning the symbol ∞ in the first case and 0 for the second case. This difficulty can be resolved by rewriting the way in which we calculate the probability of acceptance, given by
$$\lambda\left(\alpha^{(t)}, \alpha^{(*)}\right) = \min\left(1, e^{\Omega}\right),$$
where $\Omega = \log\left(p\left(\alpha^{(*)}|\alpha^{(t)}, \boldsymbol{x}\right)\right) + \log\left(q\left(\alpha^{(t)}|\alpha^{(*)}\right)\right) - \log\left(p\left(\alpha^{(t)}|\alpha^{(t)}, \boldsymbol{x}\right)\right) - \log\left(q\left(\alpha^{(*)}|\alpha^{(t)}\right)\right)$.

This method overcome the problem, as it is easy to calculate the terms $n\log(\Gamma(\alpha))$ and $\log(\Gamma(n\alpha))$ by using the approximation
$$\log(\Gamma(x)) = x\log(x) - x - \frac{1}{2}\log\left(\frac{x}{12\pi}\right) + \frac{1}{12x} - \frac{1}{360x^3} + \frac{1}{1260x^5} + \dots$$

6 | SIMULATION STUDY

A simulation study was proposed to verify whether the posterior distribution presents good results when compared with the classical approach in terms of the mean relative errors (MRE) and the mean square errors (MSE), which are the two commonly used metrics to evaluate the parameter estimators. In practice, these metrics together allows the researcher to observe if the Bayes estimates obtained from the posterior distribution are, in general, close to the true parameters. The metrics are computed by
$$\text{MRE} = \frac{1}{N}\sum_{j=1}^{N}\frac{\hat{\theta}_j}{\theta} \quad \text{and} \quad \text{MSE} = \sum_{j=1}^{N}\frac{(\hat{\theta}_j - \theta)^2}{N},$$
where $\theta = C_{pk}$ and $N = 10,000$ is the number of estimates obtained through the posterior means of the parameters. Additionally, the 95% coverage probability ($CP_{95\%}$) of the credibility intervals are computed from the Bayesian credibility intervals (CI). Under this approach, the estimator is expected to return MREs and MSEs values closer to one and zero, respectively. In addition, considering a credibility level of 95% the ranges achieved are expected to cover the true values of the $\boldsymbol{\theta}$ with a proportion of 0.95. The simulation was done with using the R software (R Core Development Team). An implemented function to compute the C_{pk} under the Bayesian approach is available in the supplemental material. Considering $n = (10, 20, 30, \dots, 150)$ the results were presented assuming different values of the C_{pk}.

Here, we considered four scenarios assuming different values for the parameters α, β, and the specification limits USL and LSL. For scenario 1 we assume $\alpha = 2$, $\beta = 0.5$, $USL = 10$, and $LSL = 0.5$ with the true $C_{pk} = 0.4599$. In the second scenario we use $\alpha = 2$, $\beta = 1$, $USL = 14.5$ and $LSL = 0.1$ and find the real $C_{pk} = 0.9710$. Similarly, we analyzed the third scenario with the respective values $\alpha = 1.1$, $\beta = 0.2$, $USL = 10$ and $LSL = 0.1$ given the true $C_{pk} = 0.1992$. Finally, in the fourth scenario we assume $\alpha = 7$, $\beta = 1.2$, $USL = 25$ and $LSL = 0.01$ and the real $C_{pk} = 1.3140$.

The MCMC approach was used to find the estimates a posterior. To achieve that, we considered the Metropolis-Hastings algorithm that was described in detail in Section 5. For each simulated data we obtain two chains of size 5500, the first 500 samples were discarded as a burn-in sample. As elements generated in the chain has usually a dependence between, we considered the thin 5 (jump between the elements) to decrease the autocorrelation. Further, we considered a convergence diagnostic criteria proposed by Geweke[33] to confirm the convergences of chains generated via MCMC assuming a 95% confidence level. The samples produced were used to estimate the posterior mean from the marginal distribution of C_{pk}. Here, we considered the posterior mean since the posterior moments are finite for $n \geq 2$ and have optimality under the Kullback-Leibler divergence. Tables 1 and 2 present the MREs, MSEs and $CP_{95\%}$ from the different estimators of C_{pk}.

As can be seen from the results above, the results obtained from the posterior distribution using the matching prior are more precise than the ones achieved through maximum likelihood in terms of MREs and MSEs. Under the classical approach, we were not able to construct the confidence intervals due to the difficulty to apply the delta method. On the other hand, we constructed the credibility intervals from the Bayesian estimates directly from its percentiles which returned accurate results, this is expected

TABLE 1 MRE, MSE and CP from the estimates of Cpk considering different values of n with $N = 10.000$ simulated samples using Classical and Bayesian inference.

θ	n	MLE		Bayes		
		MRE	MSE	MRE	MSE	CP
$\alpha = 2, \beta = 0.5$	10	1.1636	0.0439	0.9742	0.0343	0.964
	20	1.1237	0.0272	1.0384	0.0185	0.953
	30	1.0821	0.0168	1.0305	0.0131	0.950
	40	1.0583	0.0118	1.0219	0.0099	0.951
	50	1.0470	0.0088	1.0191	0.0077	0.954
	60	1.0398	0.0072	1.0170	0.0064	0.952
	70	1.0321	0.0060	1.0128	0.0055	0.952
	80	1.0286	0.0053	1.0118	0.0049	0.947
	90	1.0284	0.0047	1.0135	0.0044	0.945
	100	1.0238	0.0041	1.0105	0.0039	0.948
	110	1.0198	0.0037	1.0078	0.0035	0.948
	120	1.0167	0.0033	1.0057	0.0031	0.951
	130	1.0160	0.0031	1.0058	0.0029	0.946
	140	1.0181	0.0029	1.0086	0.0028	0.948
	150	1.0160	0.0027	1.0073	0.0026	0.946
$\alpha = 2, \beta = 1$	10	1.0231	0.0039	0.9471	0.0097	0.950
	20	1.0175	0.0023	0.9943	0.0025	0.950
	30	1.0109	0.0011	1.0004	0.0011	0.953
	40	1.0081	0.0007	1.0021	0.0007	0.950
	50	1.0066	0.0005	1.0026	0.0005	0.947
	60	1.0052	0.0004	1.0023	0.0004	0.948
	70	1.0046	0.0004	1.0023	0.0003	0.946
	80	1.0042	0.0003	1.0022	0.0003	0.950
	90	1.0034	0.0003	1.0017	0.0002	0.948
	100	1.0030	0.0002	1.0015	0.0002	0.945
	110	1.0026	0.0002	1.0013	0.0002	0.949
	120	1.0026	0.002	1.0014	0.0002	0.946
	130	1.0024	0.0002	1.0012	0.0002	0.950
	140	1.0021	0.0001	1.0010	0.0001	0.947
	150	1.0020	0.0001	1.0011	0.0001	0.944

TABLE 2 MRE, MSE and CP from the estimates of Cpk considering different values of n with $N = 10.000$ simulated samples using Classical and Bayesian inference.

θ	n	MLE		Bayes		
		MRE	MSE	MRE	MSE	CP
$\alpha = 1.1, \beta = 0.2$	10	1.3253	0.0299	1.0753	1.3957	0.956
	20	1.1497	0.0097	1.0895	0.0079	0.952
	30	1.0957	0.0055	1.0585	0.0048	0.948
	40	1.0692	0.0037	1.0425	0.0033	0.948
	50	1.0549	0.0028	1.0338	0.0026	0.951
	60	1.0455	0.0023	1.0283	0.0021	0.946
	70	1.0351	0.0019	1.0206	0.0018	0.948
	80	1.0333	0.0016	1.0205	0.0015	0.946
	90	1.0307	0.0014	1.0195	0.0014	0.948
	100	1.0234	0.0012	1.0134	0.0012	0.948
	110	1.0264	0.0012	1.0173	0.0011	0.945
	120	1.0219	0.0011	1.0136	0.0010	0.945
	130	1.0237	0.0010	1.0160	0.0009	0.947
	140	1.0192	0.0009	1.0122	0.0008	0.950
	150	1.0161	0.0008	1.0095	0.0008	0.947
$\alpha = 7, \beta = 1.2$	10	0.9450	0.0100	0.8520	0.0450	0.898
	20	1.0125	0.0094	0.9754	0.0096	0.960
	30	1.0179	0.0092	0.9949	0.0079	0.962
	40	1.0187	0.0083	1.0018	0.0070	0.953
	50	1.0167	0.0067	1.0033	0.0058	0.949
	60	1.0124	0.0054	1.0014	0.0049	0.946
	70	1.0102	0.0043	1.0009	0.0039	0.950
	80	1.0087	0.0038	1.0006	0.0035	0.950
	90	1.0089	0.0034	1.0016	0.0031	0.949
	100	1.0082	0.0031	1.0016	0.0028	0.948
	110	1.0078	0.0028	1.0019	0.0026	0.947
	120	1.0066	0.0024	1.0011	0.0023	0.950
	130	1.0054	0.0023	1.0004	0.0021	0.948
	140	1.0054	0.0021	1.0008	0.0020	0.946
	150	1.0046	0.0019	1.0002	0.0018	0.951

as the matching prior has frequentist coverage close to the nominal level. Overall, we conclude that the posterior distribution obtained with matching prior should be used to estimate the Cpk measure related to the Gamma distribution.

7 | APPLICATION

Over time technologies have played an extremely important role in all areas of knowledge. In general, one of its main functions today is to make work easier, more productive, and especially, to generate quality products that meet the wishes and needs of customers.

In this perspective, thinking about industries related manufacture in general, they are always prone to have failures in their manufacturing stages and at the end of the process, the generated products may not meet the specifications. In this sense, there are two possibilities: the customer ends up receiving a product that does not match their expectations or there is the possibility that the company is having losses due to lack of control manufacturing process. Another concern is related to the losses of customers due to the discontent of low quality of the products.

Under this motivation, the proposed methodology was applied to verify whether the product for a specific company meets the desired specifications. The datasets presented in this analysis were obtained from powder juice packages with two different flavors: grape and strawberry. The database was presented by Molina[34] who analyzed the C_{pk} assuming that the data follow a normal distribution assuming the specification limits as $LSL = 18$ and $USL = 22$, according to the standards established by the National Institute of Metrology, Quality and Technology (INMETRO), which is the body responsible for quality control in Brazil.

The data set available in Table 3 represents the weight (in grams) with an accuracy of 3 decimal places of 30 juice packs of the strawberry (Juice I) and grape (Juice II) flavors.

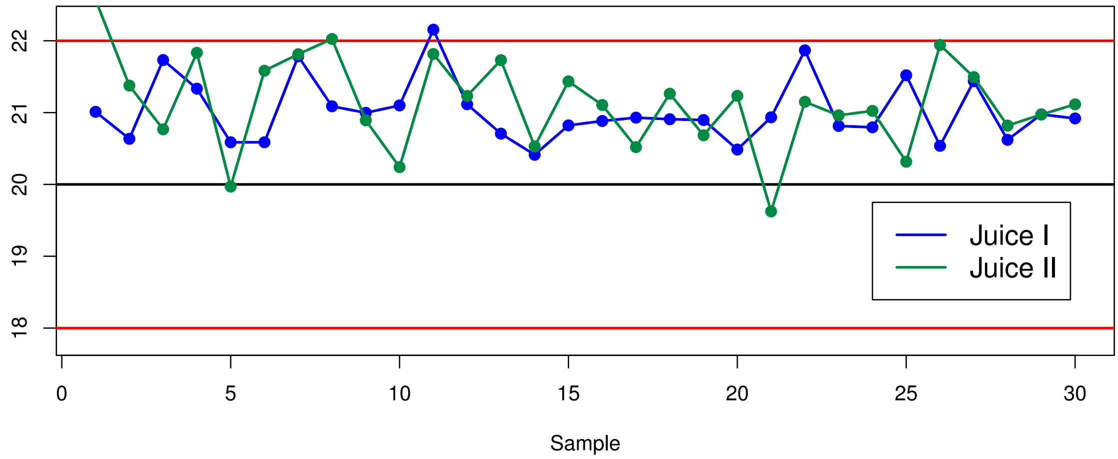

FIGURE 1 Control Chart for the Juice dataset.

Figure 1 presents the control chart for the average of juices I and II. The lower and upper specification limits are 18 and 22 respectively, therefore, the average process is expected to be around the mean (20 grams). From the two samples analyzed we observe that process does not meet the requirements, especially because they are not centered around the mean. In fact, the estimated mean for both cases are closed to 21 grams, i.e, during production, each package is receiving an additional 1 gram, which leads to a loss of one package for each 20 produced.

On the other hand, the standard statistical approach may not provide us an objective index to verify if the process is out of control, only a few samples were out of the USL, therefore, we will estimate the C_{pk}. Based on the data obtained, we considered the Kolmogorov-Smirnov (KS) test to check if the distribution can be assumed as obtained from the Gamma distribution. The

TABLE 3 Data set related to 30 packs of strawberry and grape juice.

Juice I	21.011, 20.635, 21.732, 21.333, 20.587 20.587, 21.784, 21.088, 20.997, 21.100 22.155, 21.116, 20.707, 20.413, 20.822 20.883, 20.930, 20.908, 20.897, 20.486 20.935, 21.867, 20.814, 20.795, 21.520 20.537, 21.438, 20.621, 20.975, 20.919
Juice II	22.572, 21.376, 20.768, 21.833, 19.970 21.583, 21.813, 22.025, 20.892, 20.241 21.816, 21.232, 21.730, 20.529, 21.435 21.106, 20.519, 21.263, 20.684, 21.233 19.624, 21.150, 20.962, 21.024, 20.316 21.942, 21.495, 20.819, 20.973, 21.115

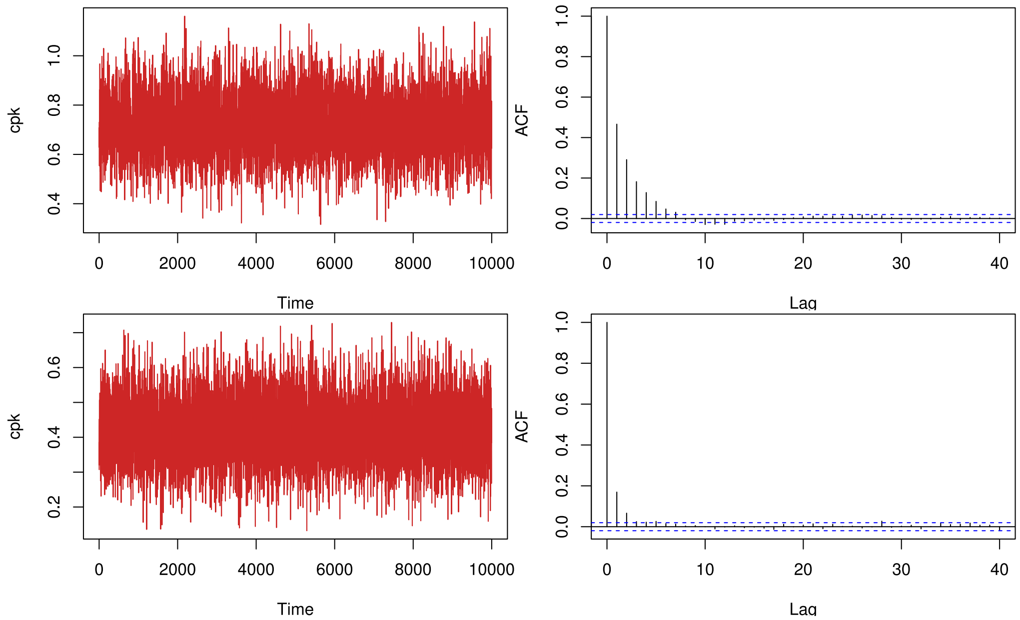

FIGURE 2 Time series plot and autocorrelation plot for the sample of the posterior distribution of the C_{pk} index for Juice I (Upper panel) and Juice II (Lower panel).

results of the test were significant, for Juice I the KS statistics was 0.17674 (p-value 0.3059) while for Juice II the Ks statistic was 0.065418 (p-value 0.9985).

Under the Bayesian approach, we need to check the convergence of the chains. Therefore, Figure 2 presents the time series of the chains (in red) for the c_{pk} of both data sets. This graph allows us to confirm the convergence of the chain. Additionally, the autocorrelation for both chains is small. In this case, we sample $505,000$ values, we discarded the first $5,000$ as a burn-in sample, further we considered a thin of 50, obtaining at the et end two chains of $10,000$. Table 4 presents Bayesian estimates and credibility intervals for the parameters α, β and C_{pk}.

TABLE 4 Bayesian estimates and 95% credibility intervals for α, β and C_{pk}.

	Juice I		Juice II	
θ	Bayes	$CI_{95\%}$	Bayes	$CI_{95\%}$
α	2281.802	(1233.944; 3638.767)	1019.621	(547.594; 1613.765)
β	108.558	(58.740; 173.087)	48.249	(25.948; 76.516)
C_{pk}	0.723	(0.497; 0.960)	0.423	(0.256; 0.591)

According to the literature, a process is said capable or adequate when the variable C_{pk} is close to 1.33. From the results presented in Table 4 we observe that in both cases the process capability index C_{pk} is below the desirable value indicating that both processes are out of control. Therefore, the process can be considered as not capable. These results provide us an objective measure to confirm our visual analysis provided by Figure 1 which shows from the control chart that the company is producing juice bags out of the customer's tolerance range.

8 | CONCLUSIONS

In this article, Bayesian inference for the capability index of the gamma distribution is developed. We proposed a Bayesian analysis through the gamma distribution together with an objective prior, called previous by matching prior that can return us a

Bayesian estimator with good properties in terms of the mean relative errors, the mean square errors, and credibility intervals for the parameters and process capability index. Markov Chain Monte Carlo algorithm is used to generate a sample of the posterior distribution under matching prior to achieving the Bayesian estimates. The simulation study shows that the posterior mean using the matching prior indicates better performance than the obtained with maximum likelihood estimator in terms of MREs and MSEs. The better performance of this approach is also confirmed through the coverage probability obtained from the credibility intervals. Additionally, under this Bayesian approach, we show that we can easily construct confidence intervals without the limitations of the delta method for complex parametrizations. It is very important to note that the matching prior also has frequentist coverage close to the nominal. We advised the practitioners that the posterior distribution obtained with this matching prior should be used to estimate the process capability index when data distributed as Gamma. The proposed estimation methodologies are demonstrated using powder juice packages with two different flavors data sets. The proposed methodology can be extended for future research using other distributions as well as other capability indices.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s)

ACKNOWLEDGEMENTS

We are grateful to the anonymous referees and the editor for the very useful comments and suggestions, which greatly improved this paper. Pedro L. Ramos acknowledges the support of the São Paulo State Research Foundation (FAPESP Proc. 2017/25971-0).

References

1. Statisti U, Tehnike N. Improving the process capability of a turning operation by the application of statistical techniques. *Materiali in tehnologije* 2009; 43(1): 55–59.

2. Kotz S, Johnson NL. Process capability indices—a review, 1992–2000. *Journal of quality technology* 2002; 34(1): 2–19.

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

4. Sappakitkamjorn J, Niwitpong S. Biased estimation of the capability indices, C_p , C_{pk} and C_{pm} using bootstrap and jackknife methods. *Asia Pacific Management Review* 2006; 10: 1–7.

5. Yum BJ, Kim KW. A bibliography of the literature on process capability indices: 2000–2009. *Quality and Reliability Engineering International* 2011; 27(3): 251–268.

6. Piña-Monarrez MR, Ortiz-Yañez JF, Rodríguez-Borbón MI. Non-normal capability indices for the Weibull and lognormal distributions. *Quality and Reliability Engineering International* 2016; 32(4): 1321–1329.

7. Zhang J. Conditional confidence intervals of process capability indices following rejection of preliminary tests. 2010.

8. Sennaroglu B, Senvar O. Performance comparison of box-cox transformation and weighted variance methods with weibull distribution. *J. Aeronaut. Space Technol* 2015; 8(2): 49–55.

9. Senvar O, Kahraman C. Type-2 fuzzy process capability indices for non-normal processes. *Journal of Intelligent & Fuzzy Systems* 2014; 27(2): 769–781.

10. Senvar O, Kahraman C. Fuzzy Process Capability Indices Using Clements' Method for Non-Normal Processes.. *Journal of Multiple-Valued Logic & Soft Computing* 2014; 22.

11. Hosseinifard Z, Abbasi B, Niaki S. Process capability estimation for leukocyte filtering process in blood service: A comparison study. *IIE Transactions on Healthcare Systems Engineering* 2014; 4(4): 167–177.

12. Clements JA. Process capability calculations, for non-normal distributions. *Quality progress* 1989; 22: 95–100.

13. Kashif M, Aslam M, Al-Marshadi AH, Jun CH. Capability indices for non-normal distribution using Gini's mean difference as measure of variability. *IEEE Access* 2016; 4: 7322–7330.

14. Panichkitkosolkul W. Confidence intervals for the process capability index Cp based on confidence intervals for variance under non-normality. *Malays. J. Math. Sci* 2016; 10(1): 101–115.

15. Leiva V, Marchant C, Saulo H, Aslam M, Rojas F. Capability indices for Birnbaum–Saunders processes applied to electronic and food industries. *Journal of Applied Statistics* 2014; 41(9): 1881–1902.

16. Senvar O, Sennaroglu B. Comparing performances of clements, box-cox, Johnson methods with weibull distributions for assessing process capability. *Journal of Industrial Engineering and Management* 2016; 9(3): 634–656.

17. Saha M, Dey S, Yadav AS, Kumar S. Classical and Bayesian inference of C py for generalized Lindley distributed quality characteristic. *Quality and Reliability Engineering International* 2019; 35(8): 2593–2611.

18. Saha M, Dey S, Yadav AS, Ali S, others . Confidence intervals of the index $C\_\{pk\}$ for normally distributed quality characteristics using classical and Bayesian methods of estimation. *Brazilian Journal of Probability and Statistics*; 35(1): 138–157.

19. Meng F, Yang J, Huang S. Hypothesis testing of process capability index C pk from the perspective of generalized fiducial inference. *Quality and Reliability Engineering International* 2020.

20. Mahmoud MAW, Kilany NM, El-Refai LH. Inference of the lifetime performance index with power Rayleigh distribution based on progressive first-failure–censored data. *Quality and Reliability Engineering International* 2020; 36(5): 1528–1536.

21. Wu CH. Modified processes capability assessment with dynamic mean shift. *Quality and Reliability Engineering International* 2020; 36(4): 1258–1271.

22. Louzada F, Ramos PL. Efficient closed-form maximum a posteriori estimators for the gamma distribution. *Journal of Statistical Computation and Simulation* 2018; 88(6): 1134–1146.

23. Miller RB. Bayesian analysis of the two-parameter gamma distribution. *Technometrics* 1980; 22(1): 65–69.

24. Sun D, Ye K. Frequentist validity of posterior quantiles for a two-parameter exponential family. *Biometrika* 1996; 83(1): 55–65.

25. Berger JO, Bernardo JM, Sun D, others . Overall objective priors. *Bayesian Analysis* 2015; 10(1): 189–221.

26. Ramos PL, Dey DK, Louzada F, Ramos E. On Posterior Properties of the Two Parameter Gamma Family of Distributions. *Anais da Academia Brasileira de Ciências* 2021.

27. Bordignon S, Scagliarini M. Statistical analysis of process capability indices with measurement errors. *Quality and reliability engineering international* 2002; 18(4): 321–332.

28. Kane VE. Process capability indices. *Journal of quality technology* 1986; 18(1): 41–52.

29. Silveira CB. Cp e Cpk – Índices de Capacidade de um processo.. 2006.

30. Louzada F, Ramos PL, Ramos E. A note on bias of closed-form estimators for the gamma distribution derived from likelihood equations. *The American Statistician* 2019; 73(2): 195–199.

31. Tibshirani R. Noninformative priors for one parameter of many. *Biometrika* 1989; 76(3): 604–608.

32. Mukerjee R, Dey DK. Frequentist validity of posterior quantiles in the presence of a nuisance parameter: higher order asymptotics. *Biometrika* 1993; 80(3): 499–505.

33. Geweke J, others . *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN . 1991.

34. Molina R. *Técnicas de Avaliação de Medidas e Verificação dos Índices de Capacidade.* . 2018.

AUTHOR BIOGRAPHIES

Marcello Henrique de Almeida has a degree in Mathematics from Universidade Federal da Grande Dourados (UFGD), 2017, Brazil. In 2019, he started pursuing master's degree in Applied and Computational Mathematics at São Paulo State University, Brazil, as a regular student of the program, currently he is developing research in the field of parametric inference with a focus on classical and Bayesian inference, reliability, and survival analysis.

Pedro Luiz Ramos is a Postdoctoral Fellow at the Institute for Mathematical Science and Computing, University of São Paulo (USP), Brazil. He received his Ph.D. degree in Statistics from the University of São Paulo in 2018, his M.S. degree in Applied and Computational Mathematic from the São Paulo State University, Brazil, and his B.S. degree in Statistics from the same university. His main research interests are in survival analysis, Bayesian inference, classical inference, and probability distribution theory.

Gadde Srinivasa Rao received his MSc in Statistics (1988), MPhil in Statistics (1994) and PhD in Statistics (2002) from the Acharya Nagarjuna University, Guntur, India. He is presently working as Professor of Statistics at the Department of Mathematics and Statistics, The University of Dodoma, Tanzania. He boasts more than 125 publications in different peer-reviewed journals in national and international well reputed journals. His papers have been cited more than 1183 times with h-index 18 and i-10 index 29 (Google Coalitions). His papers have been cited more than 369 times with h-index 10 (Web of Science Coalitions). He is reviewer of more than 60 well-reputed international journals. He has reviewed more than 120 research papers for various well-reputed international journals. He is the author of two book/ book chapters published in Springer Nature Switzerland AG. His research interests include statistical inference, statistical process control, applied Statistics, acceptance sampling plans and reliability estimation. .

Fernando Antonio Moala obtained the BSc (1988) in Mathematics, MSc in Computer Science and Computational Mathematics (emphasis in Statistics) from USP (1993), PhD (2006) in Statistics from the University of Sheffield (UK). In 2014-2015 he completed a Post-doctoral Fellow at McMaster University, Canada. He is currently Assistant Professor at São Paulo State University (UNESP) and his research is in the theory and applications of Bayesian Inference. The main areas of research are in the elicitation of expert, objective prior, nonparametric bayesian modelling and Reliability Analysis..

How to cite this article: M.H Almeida, P.L. Ramos, G.S. Rao, and F.A. Moala (2021), Objective Bayesian Inference for the Capability Index of the Gamma Distribution, *00:1–15*.