Acceptance sampling plans for inflated Pareto processes

Fernanda Figueiredo[1],* Adelaide Figueiredo[2] and M. Ivette Gomes[3]

[1]Faculty of Economics, University of Porto, and CEAUL, Portugal

[2]Faculty of Economics and LIAAD-INESC TEC, University of Porto, Portugal

[3] FCUL and CEAUL, University of Lisbon, Portugal

Abstract

In various industries of consumer goods it is important to control the presence of some chemical substances in the raw material that will affect the quality of the final products. Chromatography analyses are usually performed on samples of items taken from large batches, and on the basis of the obtained measurements, we have to conclude about the absence or presence of such substances, and then decide for the acceptance or rejection of the corresponding lots. However most of the chromatographs in use do not have sufficient precision to detect very low or high concentrations of such substances, and as a result, the data set of the measurements suggest an underlying inflated continuous distribution. In this work we highlight the adequacy of the inflated Pareto distribution to model such type of data, and we define and evaluate acceptance sampling plans under this distributional setup.

Keywords: Acceptance sampling, attributes sampling plans, estimation, inflated Pareto distribution, statistical process control, variables sampling plans.

*Corresponding author. E-mail: otilia@fep.up.pt

1 Introduction

In the food industry as well as in other manufacturing companies, it is very important to control the presence of some chemical substances in the raw material, and in some cases along the different phases of the process production, that will affect the quality of the finished products. See, for instance, Figueiredo *et al.* [5], Owen and DeRouen [10] and Whitaker *et al.* [13].

To measure the concentration of such substances, chromatography analyses are usually performed on samples of items, taken in general from large batches. On the basis of the obtained measurements and with the use of adequate acceptance sampling plans, the most important off-line technique in Statistical Quality Control, we have to conclude about the absence or presence of such substances, and then decide for the acceptance or rejection of the corresponding batches.

However, most of the chromatographs that are used have small precision to detect very low or very high concentrations of such chemical substances, or at least, to provide right quantifications in such cases, being therefore preferable to truncate the results below or above a certain threshold. In some practical situations, the measurement results of the quality characteristic, apart from being truncated, suggest an underlying inflated continuous distribution, possibly with a heavy right-tail.

Inflated discrete and continuous distributions, in particular containing many zero values, have been studied in the literature, and used in many different areas of application including agriculture, biology, ecology, environment, fishery and medicine, among others. For details on applications of this type of models see, for instance, the pioneer works of Aitchison [1] and Owen and DeRouen [10], and others more recent, such as Baksh *et al.* [2], Figueiredo *et al.* [5], Lachenbruch [7], Loganathan and Shalini [8], Rakitzis *et al.* [11], Sileshi [12] and Zidan *et al.* [14], among others.

When dealing with real data, it is also very important and demanding the comparison of different attributes and variables sampling plans, in order to choose the best one, according to the objectives to aim. Although the attributes sampling plans are the most common, both types of acceptance sampling plans have advantages and disadvantages to each other. Some details about acceptance sampling plans can be found in Carolino and Barão [3], Gomes [6], Montgomery [9] and Whitaker *et al.* [13]. For instance, a variables sampling plan usually requires a sample of smaller size than an attributes sampling plan for the same level of protection, although the sampling/observation unit costs can be higher in the variables sampling plan. But the main disadvantage of variables acceptance sampling plans is that the distribution of the quality characteristic under study has to be known (or estimated). In the standard case, a Gaussian distribution is assumed to model the underlying quality characteristic, but if the data is non-Gaussian and we implement sampling plans under this assumption, we can have large deviations from the expected performance of the sampling plans.

In a real framework, we usually only have access to the measurement results of the quality characteristic on a historical data set of items previously inspected. Moreover, in most of the cases, it is very difficult to model the data, or at least, to evaluate the performance of sampling plans on the basis of the fitted distribution. In this case how to compare and evaluate the performance of specific sampling plans? In a previous paper (Figueiredo *et al.* [5]) we considered a real data set from a consulting study, with measurements obtained from chromatography analyses, and using the bootstrap methodology (Efron and Tibshirani [4]) to construct replicates of the lots, combined with Monte Carlo simulations, we compared the performance of complex variables sampling plans. In this study, using a similar data set from the same production process, we highlight the adequacy of the inflated Pareto distribution to model such data. Then, we define and evaluate several acceptance-sampling plans under this distributional setup. We also note that the development of control charts to monitor processes that generate data of this type will also be of great interest, but is out of the scope of this paper.

The paper is organized as follows. After presenting a small introduction and the motivation for the study, Section 2 provides some information about the inflated Pareto distribution, and presents an example, based on a real situation from our consulting activity, for which the inflated Pareto model provides a good fitting. In Section 3, some variables and attributes acceptance sampling plans are developed for lots of items, assuming that the quality characteristic is a *random variable* (rv) with inflated Pareto distribution, and their performance when applied to a real data set is analyzed. Some distributional results about the Pareto, the inflated Pareto and the Poisson distributions used in the construction and evaluation of the sampling plans are also presented, as well as an algorithm for an easy implementation of such acceptance sampling plans. The paper ends with some conclusions in Section 4.

2 Inflated Pareto distribution

Let X be a mixed rv from an inflated Pareto distribution, with *cumulative distribution function* (cdf) given by

$$F(x;p,\xi,\delta) = p + (1-p)(1-(x/\delta)^{-1/\xi}), \quad x \geq \delta, \tag{1}$$

and *probability density function* (pdf) given by

$$f(x;p,\xi,\delta) = \begin{cases} (1-p)(\xi\delta)^{-1}(x/\delta)^{-1/\xi-1}, & x > \delta, \\ p, & x = \delta, \end{cases} \tag{2}$$

where δ and ξ are respectively the scale and the shape parameter, both positive, and the parameter p, is the probability associated to the singular distribution at $x = \delta$. Note that if p and δ are fixed, as larger ξ is, larger is the weight of the right-tail, and higher the frequency of getting high values. Look, for instance, for the shape of the inflated Pareto pdf represented in Figure 1.

Please Insert Figure 1 Here

2.1 Maximum likelihood estimation

Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from an inflated Pareto distribution, with $f(x; p, \xi, \delta)$ given in (2). The *maximum likelihood* (ML) estimates of the parameters p, ξ and δ maximize the logarithm of the likelihood function, defined by

$$\ln L(p,\xi,\delta) = n_1 \ln p + n_2 \ln(1-p) - n_2 \ln(\xi\delta) - (1/\xi + 1) \sum_{i=1}^{n_2} \ln(x_i/\delta), \qquad (3)$$

where n_1 and n_2 denote the number of observations respectively equal and greater to δ in the overall sample of size n. Thus, these estimates are the solution of the system of likelihood equations

$$\begin{cases} \dfrac{\partial \ln L(p,\xi,\delta)}{\partial p} = \dfrac{n_1}{p} - \dfrac{n_2}{1-p} = \dfrac{n_1 - np}{p(1-p)} = 0 \\ \dfrac{\partial \ln L(p,\xi,\delta)}{\partial \xi} = \xi n_2 - \sum_{i=1}^{n_2} \ln(x_i/\delta) = 0 \\ \dfrac{\partial \ln L(p,\xi,\delta)}{\partial \delta} = \dfrac{n_2}{\xi\delta} > 0 \end{cases} \iff \begin{cases} \widehat{p} = \dfrac{n_1}{n} \\ \widehat{\xi} = \dfrac{1}{n_2} \sum_{i=1}^{n_2} \ln(x_i/\delta) \\ \widehat{\delta} = \min x_i, \end{cases}$$

given that these values maximize the logarithm of the likelihood function, and consequently, the likelihood function.

2.2 Inflated Pareto model fitted to a real data set

Motivated by a real problem solved in our consulting activity, we will consider in this paper the following example. For confidential reasons we do not refer exactly the industry where some of the sampling plans that will be considered in this paper were applied.

Example: Consider a company that wants to control the quality of lots of raw material. Let X be a continuous rv with unknown distribution, associated with the measurements of the quality characteristic, for instance, the concentration of a chemical substance in an item of raw material. Suppose that the company establishes that all the items must have values of $X \leq 4$, and that, due to the sensitivity and precision of the measurement instrument, a chromatograph, all the observed values will be greater than or equal to 0.5. Assume that the lots for inspection are of very large size, N, say thousands of items, but the sampling rate in each lot is small due to operational problems such as the capacity of the laboratories, the excessive time spending on testing and the expensive costs.

To implement and compare different attributes and variables sampling plans, in order to choose the one that best fits the objectives of the company, suppose we only have access to a historical data set of measurements of the concentration of the chemical substance. Here, for illustration of the adequacy of the inflated Pareto distribution to model such type of data, we consider three data sets, A, B and C, presented in Table 1, corresponding to the measurements of the concentration of the chemical substance in three different types of raw material. Then, when analyzing the performance of the proposed sampling plans, developed in Section 3, we only consider set A, the larger sample.

Please Insert Table 1 Here

From Table 1 we observe that 95%, 92.3% and 97.6% of the items type A, B and C respectively, have a concentration level of this chemical substance smaller or equal to 4.0, and therefore, approximately 5%, 7.7% and 2.4% of these items do not satisfy the requirements of the company. It is also important to refer that 56.7%, 53.6% and 59.6% of the measurements in set A, B and C respectively, are equal to 0.5, which means either the absence of the chemical substance or wrong quantification due to the low sensitivity of the equipment. Finally the underlying distribution in all the three cases presents a heavy right-tail. This preliminary analysis of the data samples suggest the use of an inflated Pareto distribution to model measurements of this type.

Model fitting: Accordingly to the historical data, we have fitted an inflated Pareto model to the data, with cdf given in (1). We have considered $\delta = 0.5$, because the equipment has not precision to measure with rigor values below this threshold, and for this reason all the observations smaller or equal to 0.5 were registered as 0.5. We have further estimated the other parameters, ξ and p, by ML. Thus, we proceeded as follows: split the sample of size n into observations equal to $\delta = 0.5$ (subsample of size n_1), and greater than δ (subsample of size n_2);

- To estimate p, consider the proportion of observations equal to $\delta = 0.5$ in the overall sample, i.e., $\hat{p} = n_1/n$; we have got $\hat{p} = 908/1600 = 0.5675$ for set A, $\hat{p} = 403/752 = 0.5359$ for set B and $\hat{p} = 505/848 = 0.5955$ for set C.

- To estimate ξ, consider $\hat{\xi} = \sum_{i=1}^{n_2} \ln(x_i/\delta)/n_2 = \overline{y}$, with $y_i = \ln(x_i/\delta)$; we have got $\hat{\xi} = 0.9288$ for set A, $\hat{\xi} = 1.1035$ for set B and $\hat{\xi} = 0.7507$ for set C.

In Figure 2, we present the histograms associated with the measurements of data sets A, B and C, and the estimated pdf curves corresponding to the inflated Pareto distributions fitted to the data. As we can observe, the inflated Pareto model seems to be adequate to fit such type of measurements. Accordingly to the fitted model, the estimated probability of occurrence of a value higher than 4.0 is 4.61%, 7.05% and 2.53% for items of raw material of type A, B and C, respectively.

Please Insert Figure 2 Here

3 Acceptance sampling plans for inflated Pareto data

Suppose large lots of items, of size N, coming from an inflated-Pareto process X, with cdf given in (1), and assume that we have only one *upper specification limit* (USL), for the quality characteristic X. After a prior estimation of p and δ, on the basis of a historical data set, assume these parameters fixed and known. In case of having only one *lower specification limit* (LSL), or instead, two specification limits, USL and LSL, the development of acceptance sampling plans is similar but with more computations.

3.1 Some preliminaries

There are several types of acceptance sampling plans, but the most common are outlined for controlling the fraction of defective items, in our case θ, given by

$$\theta = \mathbb{P}(X > \mathrm{USL}) = (1-p)(\delta/\mathrm{USL})^{1/\xi}, \tag{4}$$

or a process parameter associated with the production of defectives, such as the parameter ξ, that can be written as function of θ and USL, through the expression

$$\xi = \ln(\delta/\mathrm{USL})/\ln(\theta/(1-p)). \tag{5}$$

Note that for δ and p fixed, θ will be small if ξ is small.

To develop acceptance sampling plans, we must consider consistent estimators for θ or ξ, with known distribution. Noting that the ML estimate of ξ is obtained with the observations of the sample greater than δ, the Pareto distribution has a crucial role in the development and implementation of such sampling plans.

First we must sample from the process until obtaining n_2 observations greater than δ, say $(X_1, ..., X_{n_2})$. The rv's X_i, $1 \le i \le n_2$ of this sample follow a Pareto distribution with cdf given by $F(x; \xi, \delta) = 1 - (x/\delta)^{-1/\xi}$, $x > \delta$. Note that the overall sample size

n, taken from the lot, is a rv with a negative binomial distribution, $NB(n_2, 1-p)$, being the mean sample size, $\mathbb{E}(NB)$, given by $n_2/(1-p)$.

Then, based on the sample $(Y_i = \ln(X_i/\delta)\,,\, 1 \le i \le n_2)$, where the rv's Y_i, $1 \le i \le n_2$ are distributed as an exponential standard distribution, we will consider the following consistent estimators of ξ to develop acceptance variables sampling plans: the sample mean statistic,

$$\overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i, \tag{6}$$

with $2n_2\overline{Y}/\xi$ following a $\chi^2_{2n_2}$ distribution, and the statistic T, defined by

$$T = \frac{Y_{n_2:n_2}}{\log n_2 + \gamma}, \tag{7}$$

where $\gamma = 0.5772$ is the Euler constant and $Y_{n_2:n_2}$ denotes the maximum of the sample. This statistic has a cdf given by

$$\mathbb{P}(T \le t) = P\left(Y_{n_2:n_2} \le t(\log n_2 + \gamma)\right) = \left(1 - \exp\left(-t(\log n_2 + \gamma)/\xi\right)\right)^{n_2}, \; t > 0. \tag{8}$$

To develop attributes sampling plans for lots of size N with a fraction of defectives $\theta = (1-p)(\delta/\mathrm{USL})^{1/\xi}$, we consider the number of defectives in the sample $(Y_1, ..., Y_{n_2})$, i.e., the statistic W defined by

$$W = \sum_{i=1}^{n_2} \mathbb{I}\left(Y_i - \mathrm{USL}\right), \tag{9}$$

based on the indicator function $\mathbb{I}\left(Y_i - \mathrm{USL}\right) = \begin{cases} 1, & \text{if } Y_i - \mathrm{USL} > 0 \\ 0, & \text{otherwise.} \end{cases}$

The rv W has a hypergeometric distribution. However, if $N >> n$ and θ small, we can approximate it either by a Binomial or a Poisson distribution, i.e., $W \sim H(N, n, \theta) \simeq Bi(n, \theta) \simeq P_o(\lambda = n\theta)$. For the computation of the attributes sampling plan parameters, the following result that approximates the cdf of W is used: if W is a rv with a $P_o(\lambda)$ distribution, for c integer, we can use the approximation

$$\mathbb{P}(W \le c) = \sum_{w=0}^{c} \frac{\exp(-\lambda)\,\lambda^w}{w!} \simeq 1 - F_{\chi^2_{2(c+1)}}(2\lambda), \tag{10}$$

with $F_{\chi^2_{2(c+1)}}$ denoting the cdf of a $\chi^2_{2(c+1)}$ distribution.

3.2 Design of variables and attributes sampling plans

Designing a single sampling plan usually consists of determining the values of the sample size and the acceptance number which allow us to obtain a sampling plan with a specified performance, in general, predetermined producer's and consumer's risks, or a specific *operating characteristic* (OC) curve. Sometimes, imposed by operational and cost constraints, it is necessary to design sampling plans for a fixed sample size and in order to obtain a fixed producer's risk.

Let $\mathbb{P}(A|\theta)$ denote the probability of acceptance of a lot with a fraction defective θ. The *acceptable quality level* (AQL), is the poorest level of quality for the supplier's process that the consumer would consider to be acceptable as a process average. It is only a standard, and it is hoped that the supplier's process will operate with a level quality better than AQL. The *lot tolerance percent defective* (LTPD), also called *limiting quality level*, is the poorest level of quality that the consumer is willing to accept in an individual lot. The producer's and the consumer's risks, α and β respectively, are defined by

$$\alpha = \mathbb{P}(\overline{A}|\theta = \text{AQL}) \quad \text{and} \quad \beta = \mathbb{P}(A|\theta = \text{LTPD}). \tag{11}$$

Thus, acceptance sampling plans designed to obtain fixed α and β risks for given quality levels AQL and LTPD, are designed such that

$$\begin{cases} \mathbb{P}(A|\theta = \text{AQL}) = 1 - \alpha \\ \mathbb{P}(A|\theta = \text{LTPD}) = \beta \end{cases} \iff \begin{cases} \mathbb{P}\left(A|\xi = \dfrac{\ln(\delta/\text{USL})}{\ln(\text{AQL}/(1-p))}\right) = 1 - \alpha \\ \mathbb{P}\left(A|\xi = \dfrac{\ln(\delta/\text{USL})}{\ln(\text{LTPD}/(1-p))}\right) = \beta, \end{cases}$$

Here we will consider the following ones:

PLAN I: Accept the lot if $\overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \leq k$. The parameters n_2 and k of this plan must satisfy the conditions

$$\begin{cases} \mathbb{P}\left(\overline{Y} \leq k | \xi = \dfrac{\ln(\delta/\text{USL})}{\ln(\text{AQL}/(1-p))}\right) = 1-\alpha \\ \mathbb{P}\left(\overline{Y} \leq k | \xi = \dfrac{\ln(\delta/\text{USL})}{\ln(\text{LTPD}/(1-p))}\right) = \beta. \end{cases}$$

From these equations, we determine the value n_2 such that

$$n_2 : \ F^{-1}_{\chi^2_{2n_2}}(1-\alpha) = \frac{\ln(\text{AQL}/(1-p))}{\ln(\text{LTPD}/(1-p))} F^{-1}_{\chi^2_{2n_2}}(\beta) \tag{12}$$

and then, k is given by

$$k = \frac{1}{2n_2} \frac{\ln(\delta/\text{USL})}{\ln(\text{AQL}/(1-p))} F^{-1}_{\chi^2_{2n_2}}(1-\alpha), \tag{13}$$

where $F^{-1}_{\chi^2_{2n_2}}$ denotes the inverse of the cdf of a $\chi^2_{2n_2}$ distribution.

PLAN II: Accept the lot if $T = Y_{n_2:n_2}/(\log n_2 + \gamma) \leq k$, $\gamma \simeq 0.5772$, with $Y_{n_2:n_2}$ denoting the maximum of the sample. The parameters n_2 and k of this plan must satisfy the conditions

$$\begin{cases} \mathbb{P}(T \leq k|\text{AQL}) = \left(1 - \exp\left(-\dfrac{k(\ln n_2 + \gamma)\ln(\text{AQL}/(1-p))}{\ln(\delta/\text{USL})}\right)\right)^{n_2} = 1-\alpha \\ \mathbb{P}(T \leq k|\text{LTPD}) = \left(1 - \exp\left(-\dfrac{k(\ln n_2 + \gamma)\ln(\text{LTPD}/(1-p))}{\ln(\delta/\text{USL})}\right)\right)^{n_2} = \beta. \end{cases}$$

Thus, we determine the value n_2 such that

$$n_2 : \ \left(1 - (1-\alpha)^{1/n_2}\right)^{1/\ln(\text{AQL}/(1-p))} = \left(1 - \beta^{1/n_2}\right)^{1/\ln(\text{LTPD}/(1-p))} \tag{14}$$

and then, k is given by

$$k = -\frac{\ln(\delta/\text{USL}) \ln\left(\left(1-\beta^{1/n_2}\right)^{1/\ln(\text{LTPD}/(1-p))}\right)}{\ln n_2 + \gamma}. \tag{15}$$

PLAN III: Accept the lot if $W \leq c$, with W denoting the number of defectives in the sample $(Y_1, ..., Y_{n_2})$. The parameters n_2 and c of this plan are such that

$$\begin{cases} \mathbb{P}(W \leq c|\text{AQL}) \simeq \sum_{w=0}^{c} \dfrac{\exp(-n_2\text{AQL})(n_2\text{AQL})^w}{w!} = 1-\alpha \\ \mathbb{P}(W \leq c|\text{LTPD}) \simeq \sum_{w=0}^{c} \dfrac{\exp(-n_2\text{LTPD})(n_2\text{LTPD})^w}{w!} = \beta, \end{cases}$$

or that verify

$$\begin{cases} F_{\chi^2_{2(c+1)}}(2n_2\text{AQL}) = \alpha \\ F_{\chi^2_{2(c+1)}}(2n_2\text{LTPD}) = 1-\beta. \end{cases}$$

Thus, we determine the value c such that

$$c: \; F^{-1}_{\chi^2_{2(c+1)}}(\alpha) \simeq (\text{AQL}/\text{LTPD}) \times F^{-1}_{\chi^2_{2(c+1)}}(1-\beta) \tag{16}$$

and then,

$$n_2 = F^{-1}_{\chi^2_{2(c+1)}}(\alpha)/(2\text{AQL}), \tag{17}$$

where $F^{-1}_{\chi^2_{2(c+1)}}$ denotes the inverse of the cdf of a $\chi^2_{2(c+1)}$ distribution.

3.3 Performance of the previous sampling plans

In the context of the example in Subsection 2.2, we assume lots of large size N, and an upper specification limit USL $= 4$ for the items. Thus, items associated with measurements of concentration of the chemical substance above 4 are considered defective (or nonconforming). To illustrate the performance of the previous sampling plans I, II and III, we only consider the larger historical data set A of the measurements of the chemical substance. To determine acceptance sampling plans for the fraction of defective items in the batches, we assume δ known, equal to 0.5, and p fixed, equal to the obtained ML estimate, $\widehat{p} = 0.5675$. We consider that the deterioration of the quality of the lots of items is essentially due to changes in the parameter ξ of the distribution. As we referred before, for fixed p, the probability of occurring very high values of X increases with ξ.

The parameters n_2, k and c of the sampling plans I, II and III are determined in order to obtain plans with specific OC curves, i.e., predetermined α and β risks for given AQL and LTPD levels. We compare the performance of these sampling plans by analyzing the OC curve, as well as the *average outgoing quality* (AOQ) curve, and in particular, its maximum value, called AOQ limit. The OC curve, i.e., the curve fitted to the points $(\theta, \mathbb{P}(A|\theta))$, for $\theta = 0, 1/N, \ldots, 1$, shows the discriminatory power of the sampling plan. Comparing the OC curves, the most severe plan is the one associated to the OC curve that decreases faster. The AOQ curve, i.e., the curve fitted to the points $(\theta, \text{AOQ}(\theta) = \theta \times \mathbb{P}(A|\theta))$, for $\theta = 0, 1/N, \ldots, 1$, describes the average quality of the lot that results from a 100% rectifying inspection program over a long sequence of lots, from a process with fraction defective θ. A 100% rectifying inspection program means that when the lot is rejected, all the items of the lot are inspected, and the discovered defective items are replaced by good ones. The AOQ limit represents the worst possible average quality that would result from the rectifying inspection program. In our case it is approximately measured by the proportion of items in the lot with a concentration of the chemical substance above 4, that pass the control, i.e. that is forwarded to the production line or to the buyers.

For illustration, we represented in Figures 3 and 4, the OC and the AOQ curves of the plans I, II and III designed to have risks $\alpha = 15\%$ and $\beta = 30\%$ when AQL = 5% and LTPD = 10%, and risks $\alpha = 5\%$ and $\beta = 10\%$ when AQL = 2.5% and LTPD = 10%, respectively. When analyzing the OC curves in Figure 3, we observe that plan I requires smaller sample size than the others for the desired levels of protection, and even though, it is the one with the best performance in terms of $\mathbb{P}(A, \theta)$, AOQ and AOQ limit. Plan II is the one with the larger sample size, and in terms of performance we have no gain with it. However, with all the plans, the probability of accepting a lot with a large percentage of defectives is high, and the AOQ limit is not significantly small when compared with θ. This can be explained by the large values considered for the α and β risks, and consequently, by the small sample sizes here considered. Thus, there is a chance to design improved sampling plans. For instance, analyzing the performance of the plans represented in Figure 4, we conclude that plan I is significantly better than the others in terms of $\mathbb{P}(A, \theta)$, AOQ and AOQ limit, with the expense of a larger sample. When we reduce the α and β risks, and in this case we also set the acceptable quality level AQL in a smaller value, the sample size increases significantly. Accordingly to the objectives to aim, plan I or plan II are the best ones.

Please Insert Figure 3 Here

Please Insert Figure 4 Here

To better compare the performance of the sampling plans I, II and III, based on the statistics $\overline{Y}$, T and W defined in (6), (7) and (9), respectively, we have designed them for a fixed value $n_2 = 25, 50$, and with acceptance numbers k and c determined through the equations (13), (15) and (16), in order to obtain a predetermined risk, $\alpha = 5\%$ in cases of AQL $= 2.5\%$ and AQL $= 1\%$. From Figures 5 and 6, we observe that plan I is significantly better than the others, and as expected, its performance increases with the sample size n_2. For smaller AQL values we get similar conclusions, with improvements in the performance of the sampling plans, as expected.

Please Insert Figure 5 Here

Please Insert Figure 6 Here

3.4 Algorithm for the implementation of acceptance sampling plans for inflated Pareto distribution

To promote the use of the previous acceptance sampling plans by practitioners, we provide the following algorithm for an easy implementation of sampling plans designed to obtain fixed α and β risks for given quality levels AQL and LTPD.

Algorithm:

1. Fix the standards AQL and LTPD, and the risks α and β;

2. Determine the parameters of the sampling plans, i.e. the sample size and the acceptance number. In case of:

i. Plan I: Determine n_2 using equation (12), and k using equation (13);

ii. Plan II: Determine n_2 using equation (14), and k using equation (15);

iii. Plan III: Determine n_2 using equation (17), and c using equation (16);

3. Sample from the process until obtaining n_2 observations greater than δ, say $(X_1, ..., X_{n_2})$.

4. Compute the values $Y_i = \ln(X_i/\delta)$, $1 \leq i \leq n_2$;

5. Compute the control statistic and take decisions about the lot. In case of:

i. Plan I: Compute $\overline{Y}$ using equation (6). If $\overline{Y} \leq k$, accept the lot; otherwise, reject the lot;

ii. Plan II: Compute T using equation (7). If $T \leq k$, accept the lot; otherwise, reject the lot;

iii. Plan III: Compute W using equation (9). If $W \leq c$, accept the lot; otherwise, reject the lot.

4 Conclusions

In this paper we refer the importance of inflated models in applications, and in particular, we present some motivation for the use of the inflated Pareto distribution, that is a very manageable mixture-type model, with simple distributional properties. We derive some variables and attributes acceptance sampling plans for inflated Pareto data, assuming that the deterioration of the quality of the lots is essentially due to changes in the shape parameter of the distribution, and considering the other parameters of the model fixed and known. Simple analytical expressions are provided to determine the parameters of the different sampling plans here considered and that allow to achieve the desired performance in terms of the producer's and consumer's risks for given AQL and LTPD levels. We illustrate the performance of such sampling plans in terms of the obtained OC and AOQ curves, and as expected, we concluded that it is possible to design variables sampling plans with the same or higher level of protection as attributes sampling plans, on the basis of a smaller sample size. To promote and facilitate the use of such sampling plans by practitioners, an algorithm for its implementation is provided. Finally, the implementation of control charts to monitor inflated Pareto data is also of great interest, as well as the design of acceptance sampling plans for inflated Pareto models in the case of all the parameters unknown. These topics will be addressed in a future work.

Acknowledgements

This research was partially supported by National Funds through **FCT** – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology), through the projects PEst-OE/MAT/UI0006/2014 (CEA/UL) and UID/EEA/50014/2013.

References

[1] Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistics Association*, 50, 901–908.

[2] Baksh, M.F., Bohning, D. and Lerdsuwansri, R. (2011). An extension of an over-dispersion test for count data. *Computational Statistics and Data Analysis*, 55, 466–474.

[3] Carolino, E. and Barão, I. (2013). Robust methods in acceptance sampling. *Revstat* 11(1), 67–82.

[4] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

[5] Figueiredo, F., Figueiredo, A. and Gomes, M.I. (2014). Comparison of sampling plans by variables using the bootstrap and Monte Carlo simulations. *AIP Conference Proceedings* 1618, 535–538.

[6] Gomes, M.I. (2011). Acceptance sampling. In *International Encyclopedia of Statistical Science, Part 1,* Lovric, M. (ed.), Springer, 5–7.

[7] Lachenbruch, P.A. (2001). Comparison of two-part models with competitors. *Statistics in Medicine*, 20, 1215–1234.

[8] Loganathan, A. and Shalini, B.K. (2014). Determination of single sampling plans by attributes under the conditions of zero-inflated Poisson distribution. *Communications in Statistics — Simulation and Computation*, 43, 538–548.

[9] Montgomery, D.C. (2009). *Introduction to Statistical Quality Control: a Modern Introduction*, 6th edition, John Wiley & Sons.

[10] Owen, W.J. and DeRouen, T.A. (1980). Estimation of the mean for lognormal data containing zeros and left-censored values, with applications to the measurement of worker exposure to air contaminants. *Biometrics*, 36 (4), 707–719.

[11] Rakitzis, A., Maravelakis, P. and Castagliola, P. (2015+). CUSUM Control Charts for the Monitoring of Zero-Inflated Binomial Processes. *Quality and Reliability Engineering International* (in press).

[12] Sileshi, G. (2008). The excess-zero problem in soil animal count data and choice of appropriate models for statistical inference. *Pedobiologia*, 52, 1–17.

[13] Whitaker, T.B., Doko, M.B., Maestroni, B.M. and Ogunbanwo, B.F. (2007). Evaluating the performance of sampling plans to detect Fumonisin B_1 in maize lots marketed in Nigeria. *Journal of AOAC International*, 90(4), 1050–1059.

[14] Zidan, M., Wang, J.-C. and Niewiadomska-Bugaj, M. (2011). Comparison of k independent, zero-heavy lognormal distributions. *The Canadian Journal of Statistics*, 39(4), 690–702.

Table 1: Measurements of the concentration of the chemical substance in three different types of raw material, given in data sets A, B and C, of size 1600, 752 and 848 items respectively.

Classes of measurements	Set A Number of items (%)	Set B Number of items (%)	Set C Number of items (%)
0.5	908 (56.7%)	403 (53.6%)	505 (59.6%)
]0.5,1.0]	357 (22.3%)	157 (20.9%)	200 (23.6%)
]1.0,2.0]	187 (11.7%)	91 (12.1%)	96 (11.3%)
]2.0,3.0]	53 (3.3%)	31 (4.1%)	22 (2.6%)
]3.0,4.0]	16 (1.0%)	12 (1.6%)	4 (0.5%)
]4.0,5.0]	20 (1.3%)	13 (1.7%)	7 (0.8%)
]5.0,7.5]	17 (1.1%)	12 (1.6%)	5 (0.6%)
]7.5,10.0]	16 (1.0%)	10 (1.3%)	6 (0.7%)
]10.0,20.0]	10 (0.6%)	8 (1.1%)	2 (0.2%)
>20.0	16 (1.0%)	15 (2.0%)	1 (0.1%)

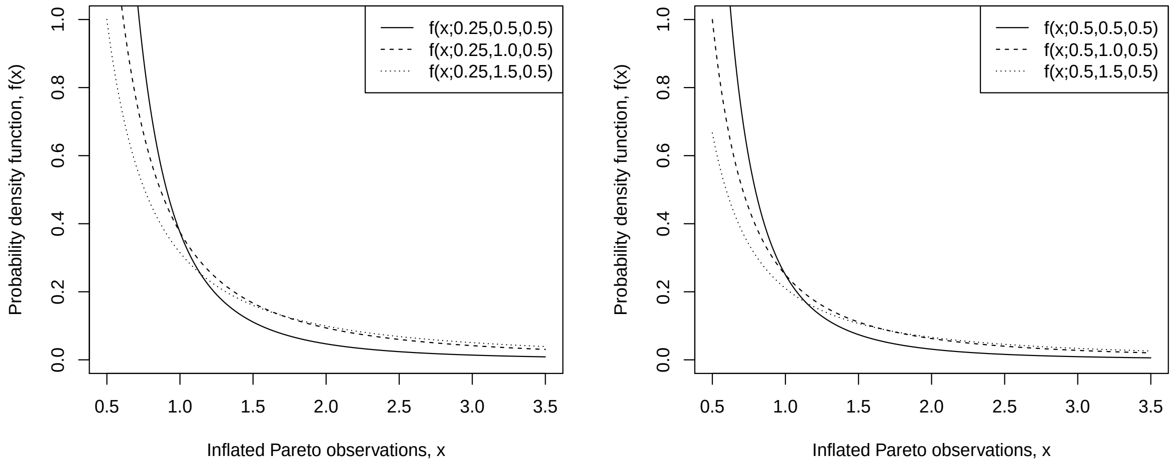

Figure 1: Probability density function of inflated Pareto distributions with parameters $p = 0.25$ (left), 0.5 (right), $\delta = 0.5$ and $\xi = 0.5, 1, 1.5$.

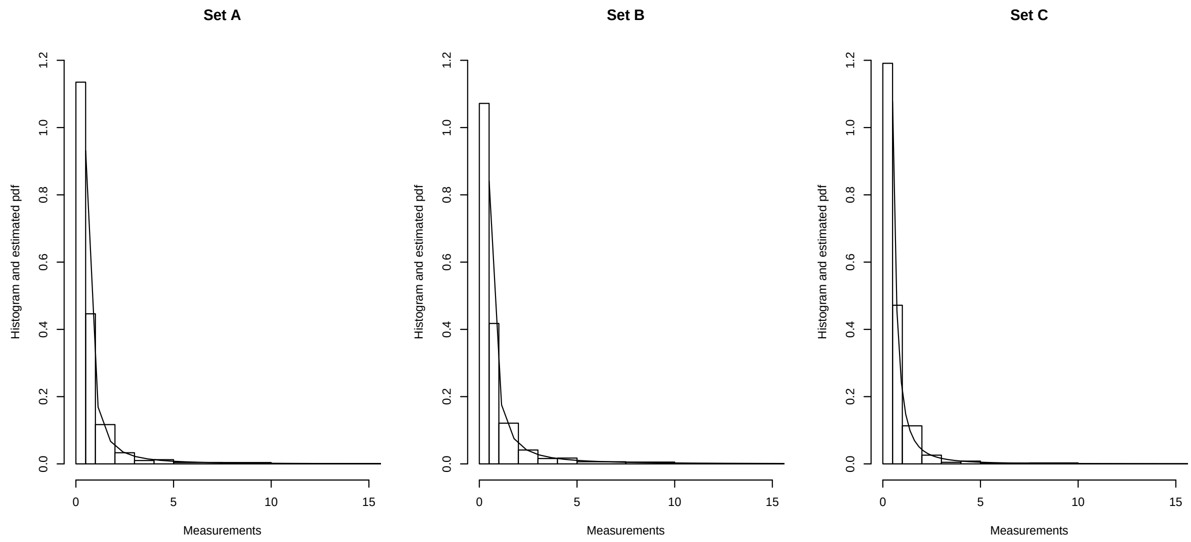

Figure 2: Histograms and pdf of the inflated Pareto distributions fitted to data sets A, B and C.

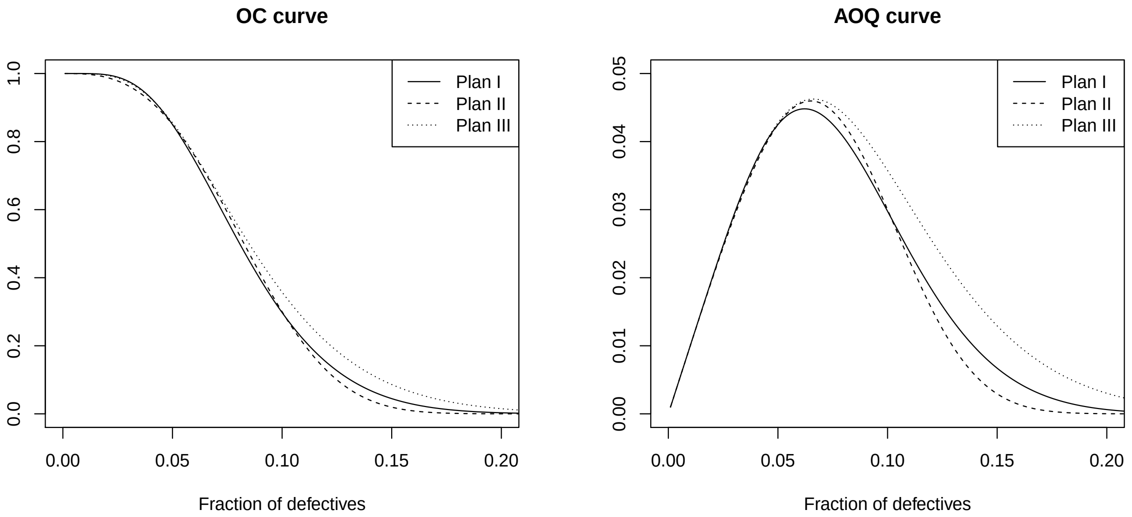

Figure 3: OC and AOQ curves of the Plan I ($n_2 = 16$, $k = 1.213$), Plan II ($n_2 = 81$, $k = 1.204$) and Plan III ($n = 55$, $c = 4$), designed to have risks $\alpha = 15\%$ and $\beta = 30\%$, when AQL $= 5\%$ and LTPD $= 10\%$.

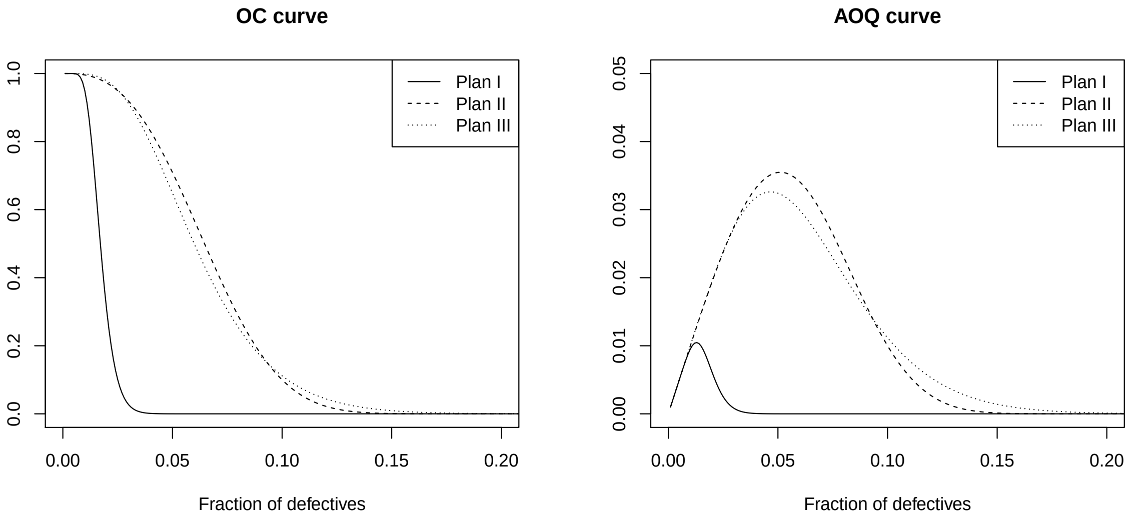

Figure 4: OC and AOQ curves of the Plan I ($n_2 = 109$, $k = 0.642$), Plan II ($n_2 = 125$, $k = 1.052$) and Plan III ($n = 78$, $c = 4$), designed to have risks $\alpha = 5\%$ and $\beta = 10\%$, when AQL $= 2.5\%$ and LTPD $= 10\%$.

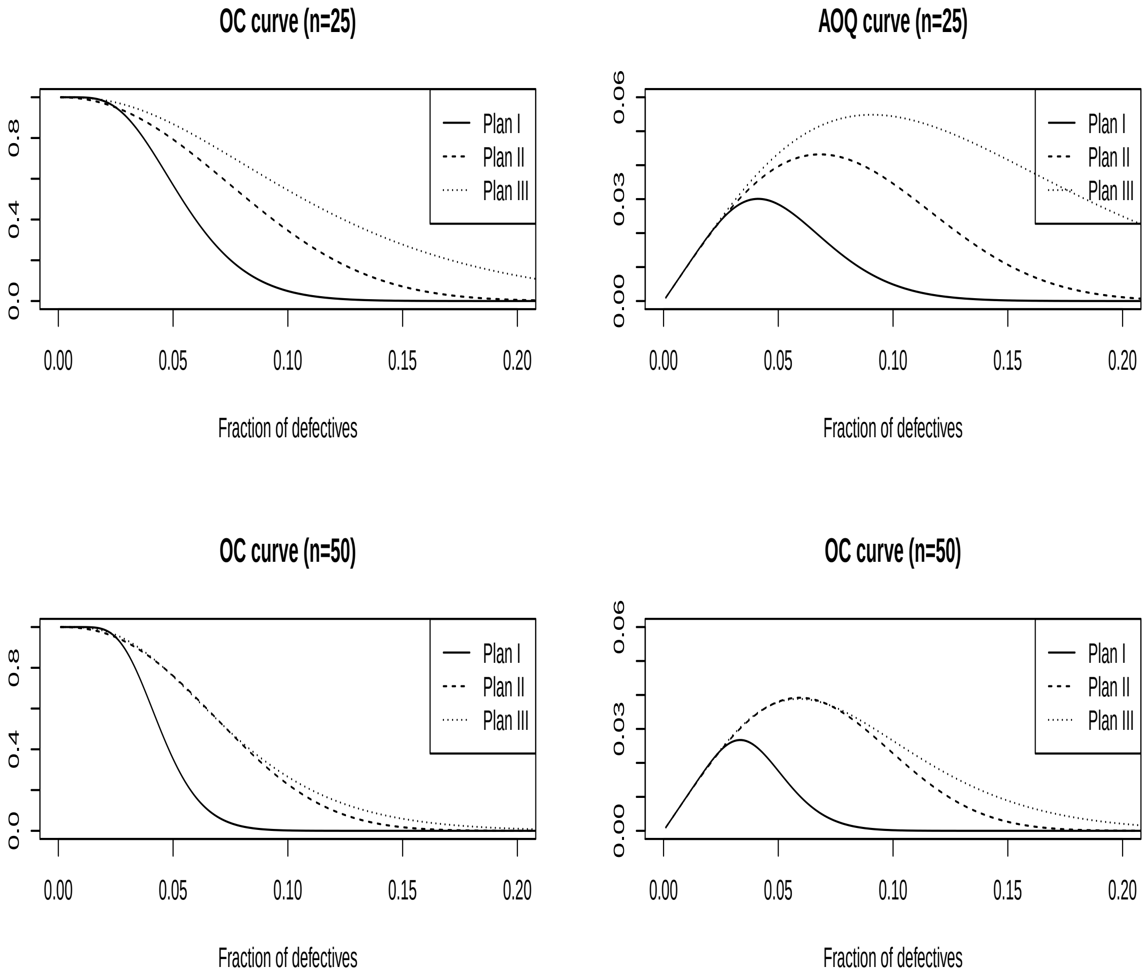

Figure 5: OC and AOQ curves of the sampling plans designed to have an α-risk 5% when AQL = 2.5%

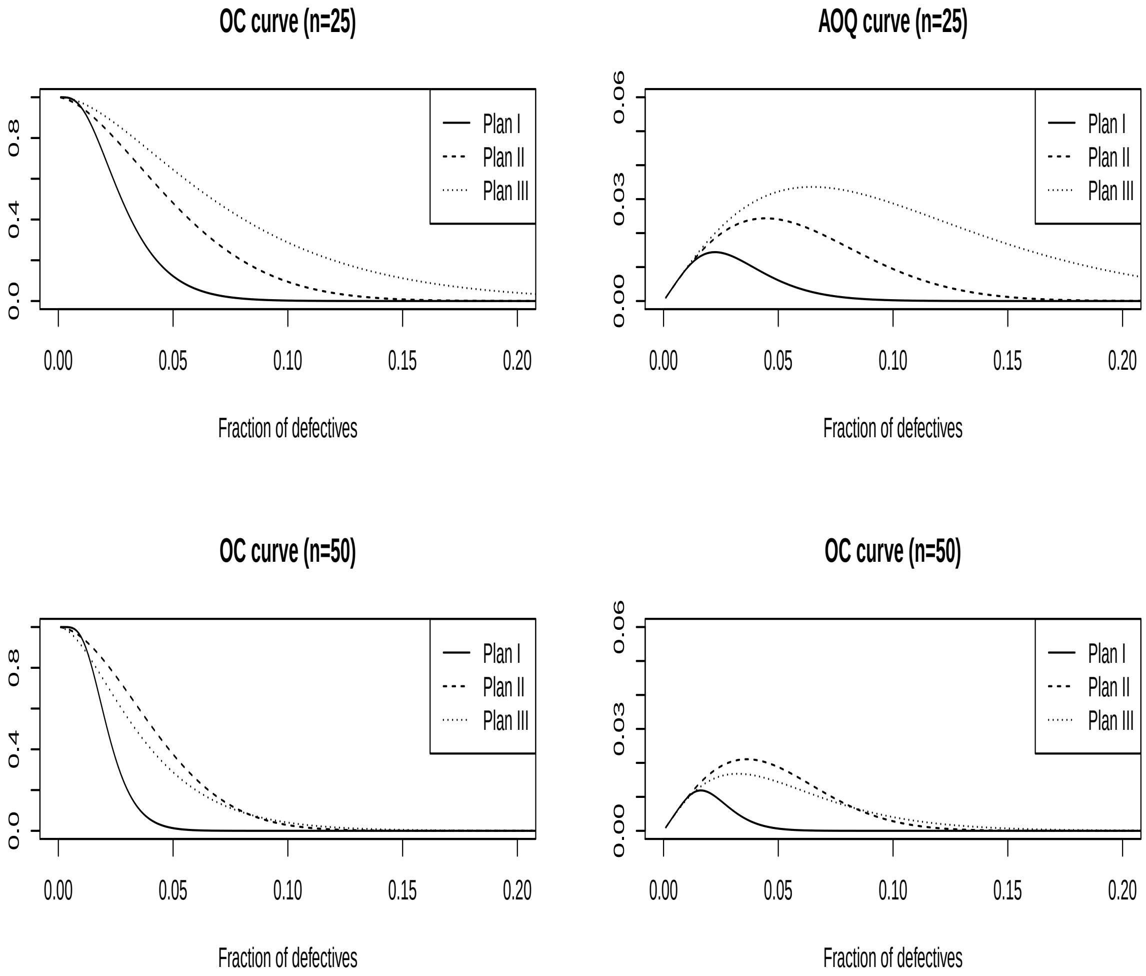

Figure 6: OC and AOQ curves of the sampling plans designed to have an α-risk 5% when AQL $= 1\%$