Rescaled bootstrap confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest

Paper published in

Communications in Statistics-Simulation and Computation

Full citation to this publication:

https://doi.org/10.1080/03610918.2020.1859539

- **APA**

Moya, P. J., Munoz, J. F., Verdejo, E. Á., & Blanco-Encomienda, F. J. (2023). Rescaled bootstrap confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest. *Communications in Statistics-Simulation and Computation*, *52*(2), 466-481.

- **ISO 690**

MOYA, P. J., et al. Rescaled bootstrap confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest. *Communications in Statistics-Simulation and Computation*, 2023, vol. 52, no 2, p. 466-481.

- **MLA**

Moya, P. J., et al. "Rescaled bootstrap confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest." *Communications in Statistics-Simulation and Computation* 52.2 (2023): 466-481.

Rescaled bootstrap confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest

Abstract

Confidence intervals for the population variance in the presence of outliers or spikes in the distribution of a variable of interest are topics that have not been investigated in depth previously. Results derived from a first Monte Carlo simulation study reveal the limitations of the customary confidence interval for the population variance when the underlying assumptions are violated, and the use of alternative confidence intervals is thus justified. We suggest confidence intervals based on the rescaled bootstrap method for many reasons. First, this is a simple technique that can be easily applied in practice. Second, it is free of probabilistic distributions. Finally, it can be easily applied to the cases of finite populations and samples selected from complex sampling designs. Results derived from a second Monte Carlo simulation study indicate that the suggested confidence intervals have desirable coverage rates with smaller average widths. Accordingly, an advantage of the suggested confidence intervals is that they offer a good compromise between simplicity and desirable properties. The various simulation studies are based on different scenarios that may arise in practice, such as the presence of outliers or spikes, and the fact that the underlying assumptions of the customary confidence interval are violated.

Keywords: Finite population, coverage rate, missing data, Normal distribution

1. Introduction

Many quantitative techniques are based on the variance of a random variable, and they are applied to a wide range of disciplines and topics, such as social sciences, econometrics, engineering, etc. Some known examples are the statistical inference, the hypothesis testing or the goodness of fit (see [46]). In particular, the variance has a relevant application in Statistical Quality Control (SQC), since this parameter is commonly used to analyze the variability of the production process, the construction of different limits (natural tolerance limits of the process, control limits, etc.), and the implementation of various SQC techniques (control charts, process capability index, etc.). See also [27], [29], [41] and [42].

The variance is usually an unknown parameter in practice. In this situation, it is quite common to estimate the unknown parameter by using the information collected from a random sample. For the problem of estimating the population variance, the popular sample variance based on the Bessel's correction is the most common estimator. In addition, most studies report the corresponding confidence intervals along with the point estimators. Let $X_1, ..., X_n$ be a sample, with size n, taken from a given variable of interest x. The traditional confidence interval for the population variance of x (σ^2) is based on the following assumptions:

(A1) x follows a Normal distribution with mean μ and variance σ^2, i.e., $x \rightarrow N(\mu, \sigma^2)$.

(A2) $X_1, ..., X_n$ are independent and identically distributed (iid).

For instance, the Assumption ($A2$) holds when the sample is extracted from an infinite population. However, this is not the case of many situations in practice. For example, many surveys conducted by official governments or institutions, such as the European Union Statistics on Income and Living Conditions (see [15]), are based on finite populations. Similarly, many SQC techniques are based on finite populations, since they are applied in finite lots of products. The traditional confidence interval for the population variance may have a poor performance in the case of finite population with large sampling fraction $f = n/N$, where N is the size of the finite population. It is well known that the approximation to an infinite population may fit well when f is small (see, for example,

[29]). Estimators of the population variance under different sampling designs can be seen in [9], [12], [17], [36], etc. Some references that propose estimators of the population variance in the presence of auxiliary variables are [16], [24], [38], [40] and [39].

The main contribution of this paper is to introduce an alternative method to construct confidence intervals for the population variance, which may offer a good performance in the presence of different scenarios that may arise in practice. In particular, we first analyze the presence of outliers, since they significantly affect the estimation of the population variance. Second, the presence of spikes in the distribution of the variable of interest may introduce biases in the estimation of the population variance, hence this topic is also investigated. For instance, this problem may arise when certain imputation techniques are used in the presence of missing data, which is quite common in practice. Additional scenarios based on simulated and real populations are also studied, and they are used to analyze the empirical performance of the various confidence intervals when the Assumptions ($A1$) and ($A2$) are violated.

The suggested confidence intervals are based on the bootstrap methodology, which is a common technique in social sciences when the theoretical distribution of an interest statistic is unknown ([10, 14, 47]). In particular, we consider the rescaled bootstrap resampling method to construct confidence intervals for the population variance. This method was originally proposed by [33], and numerous extensions and applications have been subsequently proposed in the literature. The rescaled method suggested by [33] is only applicable to smooth statistics due to the fact that the rescaling factors are applied to the survey data values. However, [34] proposed a new version of this method where the corresponding rescaling factors are applied to the survey weights (see [35]), and both smooth and non-smooth statistics can be thus used. [2] proposed a rescaled bootstrap method that can be applied to the case of without replacement sampling designs. Similarly, [8] modified the rescaled bootstrap method to the situation where the bootstrap samples are selected without replacement. [32] proposed an extension of the rescaled bootstrap to stratified multistage designs, which has relevant applications in large scaled sample surveys and a wide range of reweighting methods, such as the calibration estimators ([11, 25]). [3] proposed a novel method for the problem of estimating quantiles, and the corresponding confidence intervals are based on the rescaled bootstrap method. Assuming also the context of finite populations, [4] applied the rescaled bootstrap technique to the case of samples selected by ranked set sampling. Additional details related to the rescaled bootstrap and other bootstrap methods can be found in [37]. On the other hand, the limits of the suggested confidence intervals are obtained using the percentile and the studentized bootstrap approximations.

The suggested confidence intervals have some desirable properties. First, an advantage of this method is its simplicity, i.e., it has a simple implementation in practice. Second, many methods are based on a certain probabilistic distribution, and the performance of such methods may be poor if this assumption is violated. In this sense, the suggested confidence intervals are free of probabilistic distributions. Finally, they can be easily applied to finite populations and samples selected from complex sampling designs. A Monte Carlo simulation study is carried out to analyze the empirical performance of the proposed confidence intervals, and this study is based on multiple populations and scenarios that may arise in practice. In particular, we investigate different cases where the underlying assumptions of the customary confidence interval are violated, such as the presence of outliers or spikes. Confidence intervals with desirable properties are obtained, i.e., they have empirical coverage rates close to the required nominal level, and their average widths are smaller than the average widths of alternative confidence intervals.

This article is organized as follows. In Section 2, some existing confidence intervals of the population variance are introduced. In Section 3, we identify the limitations of the customary confidence interval, and they are empirically analyzed via a first Monte Carlo simulation study. Results derived from this section justify that the customary confidence interval may have a poor performance under the situations investigated in this paper, and the use of alternative confidence intervals is thus required in such cases. In Section 4, we propose using the rescaled bootstrap method to construct confidence intervals for the population variance. In Section 5, the empirical performance of the various confidence intervals is analyzed via a second Monte Carlo simulation study, and desirable results are obtained. Note that multiple scenarios are investigated in both simulation studies, such as the fact that the underlying assumptions of the customary confidence interval are violated. Such scenarios include the presence of outliers or spikes in the distribution. The main conclusions are summarized in Section 6.

2. Some confidence intervals for the population variance

Assuming that both Assumptions $(A1)$ and $(A2)$ hold, the customary $100(1-\alpha)\%$ confidence interval for the population variance σ^2 is given by [44]:

$$CI_{\chi^2} = [L_{\chi^2}, U_{\chi^2}], \tag{1}$$

where the lower and upper limits are defined, respectively, as

$$L_{\chi^2} = \frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}$$

and

$$U_{\chi^2} = \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}},$$

and where

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

is the sample variance, $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ is the sample mean, x_i denotes the ith observed value of the variable of interest x, and $\chi^2_{df,a}$ is the ath quantile of the Chi-square distribution with df degrees of freedom. Note that the confidence interval (1) is highly sensitive if the Assumptions $(A1)$ and $(A2)$ are violated ([22, 26]), hence such confidence interval may have a poor performance under different scenarios. Some examples are the situations investigated in this paper, i.e., presence of outliers or spikes in the distribution.

Bonett ([5]) proposed the $100(1-\alpha)\%$ confidence interval

$$CI_B = [L_B, U_B], \tag{2}$$

where the lower and upper limits are given, respectively, by

$$L_B = \texttt{exp}\left\{\texttt{ln}(cS^2) - Z_{\frac{1-\alpha}{2}} se\right\},$$

$$U_B = \texttt{exp}\left\{\texttt{ln}(cS^2) + Z_{\frac{1-\alpha}{2}} se\right\},$$

Z_a is the ath quantile of the standard Normal distribution,

$$se = c\left[\frac{\hat{\gamma_4}(n-3)/n}{(n-1)}\right]^{1/2},$$

$c = n/(n - Z_{\alpha/2})$ is an empirically determined, small-sample adjustment that helps equalize the tail probabilities,

$$\hat{\gamma_4} = n\frac{\sum_{i=1}^{n}(x_i - m)^4}{[\sum_{i=1}^{n}(x_i - \bar{x})^2]^2}$$

is the estimator of the coefficient of kurtosis

$$\gamma_4 = \frac{\mu^4}{\sigma^4},$$

and m is a trimmed mean with trim-proportion equal to $1/[2(n-4)^{1/2}]$. A Monte Carlo simulation study based on different probabilistic distributions was carried out in [5]. Normal and non-normal populations were used to compare the empirical performance of the customary (CI_{χ^2}) and Bonett (CI_B) confidence intervals, and both methods were not empirically investigated in the presence of outliers or spikes in the distribution.

3. Limitations of the customary confidence interval under the investigated scenarios

3.1. Description of populations and scenarios

Table 1 summarizes the different populations and scenarios considered in the various Monte Carlo simulation studies discussed in this paper. First, we observe that real and simulated populations are used to evaluate the empirical performance of the various confidence intervals. Note that the real population, named as "*Rods*", is used to analyze the accuracy of the confidence intervals under this real situation. This population contains 183 measurements of the inside diameter of connecting rods manufactured for large diesel engines, where the average diameter is 112.015 millimeters, and the variance is given by 6.401×10^{-4}. The quality of the commented connecting rods is controlled and improved via different SQC techniques that depend on the variance, and for this reason, the problem of estimating the population variance is an important issue in this population. Although there has been an improvement in the accuracy, traditional methods for data collection and processing, that is, manuals, are still used. Accordingly, the presence of spikes is likely due to such manual methods, and the fact that most measurements are around the target value for the quality characteristic. A bar chart for this real population can be seen in Figure 1. This data set is available from the authors upon request. Note that the Shapiro-Wilk test of normality is applied and the null hypothesis is rejected, and this implies that the Assumption ($A1$) is violated in the *Rods* population.

Table 1 about here

Figure 1 about here

In addition, a deeper analysis is also obtained by using the simulated populations, since they have different characteristics and probabilistic properties based on the Normal and Uniform distributions. For each probabilistic distribution, the parameters are chosen so that the correspondent population mean is 10, and the population variance takes the values $\sigma^2 = \{1, 5\}$. Note that the $Normal$ population, with variance $\sigma^2 = 1$, is used to create the *Outliers* and *Spikes* populations, and the Assumption ($A1$) is violated in the *Outliers*, *Spikes* and $Uniform$ populations.

The customary confidence interval defined in (1) depends on a Chi-square distribution. However, this distribution may not fit well when the Assumptions ($A1$) and ($A2$) are violated, and the interval (1) may have a poor performance in this situation. From Table 1 we observe that the sizes of simulated populations are $N = \{1000, 10000\}$, and the sample size is fixed at $n = 200$. This implies that the sampling fraction $f = n/N$ takes the values 0.02 and 0.2 in the case of simulated populations. Thus, we expect that the confidence interval (1) performs well in the $Normal$ population with a small sampling fraction ($f = 0.02$), since the approximation to an infinite population may fit well in this situation (see, for example, [29]), and both Assumptions hold. We consider the values $N = \{500, 1000\}$ for the case of the real population. For this purpose, the size of the *Rods* population ($N_{Rods} = 183$) is duplicated $k = \lfloor N/N_{Rods} \rfloor$ times, and $N_1 = N_{Rods} \times k$ data are thus obtained. We assume that $N = N_1 + N_2$, and the $N_2 = N - N_1$ remaining units are randomly selected from the *Rods* population. For instance, the population size $N = 500$ is obtained by taking the values $k = 2$, $N_1 = 366$ and $N_2 = 134$. The same sample size used in simulated populations is considered in this population, i.e., $n = 200$, and this implies that the sampling fractions are $f = \{20\%, 40\%\}$ in the *Rods* population.

From Table 1 we observe that two important scenarios that may arise in practice are also investigated. First, a common problem in many data sets is the presence of outliers, i.e., values that statistically differ from the data set to which they belong ([1, 48]). Note that outliers significantly affect the estimation of the population variance, and important biases may be thus introduced. The $Ouliers$ population is obtained by randomly selecting N_{out} units from the N elements of the $Normal$ population with mean 10 and variance $\sigma^2 = 1$, and they are replaced by N_{out} data extracted from a Normal distribution with mean 16 and the same variance. Different percentages of outliers (p_{out}) are considered, in particular, $p_{out} = \{20\%, 30\%, 40\%\}$, where $p_{out} = N_{out}/N$.

Second, the problem of spikes in the distribution may arise in practice, for example, when dealing with missing data. The problem of missing data is very common in many disciplines, and various solutions can be adopted in this situation, as can be seen in [20], [23], [28], etc. In practice, one of the most accepted and used techniques are the imputation methods ([13, 45]). However, it is well known that various imputation techniques may create spikes in the distribution of variables ([19]). Some examples are the (groupwise) mean and median imputation methods ([23, 43]). In addition, various imputation methods may also generate spikes in the case of categorical predictors with a small number of categories. Some examples are the Nearest Neighbour Imputation (also named as NNI), and the regression imputation methods ([7, 21]). The *Spikes* population is considered for the purpose of analyzing the problem of spikes in the distribution of the variable of interest. This population is obtained by randomly selecting n_{md} units from the original sample with size n, which in turn is selected from the $Normal$ population. The n_{md} units are treated as missing data, i.e., we assume that the original sample only has $n_r = n - n_{md}$ respondents. The mean imputation method is then applied, and the n_{md} missing values are thus substituted by the mean of the n_r observations. Different percentages of missing data (p_{md}) are considered, where $p_{md} = n_{md}/n$. In particular, $p_{md} = \{10\%, 30\%, 50\%, 70\%\}$.

3.2. Empirical results

The customary confidence interval for the population variance defined in (1) depends on the Chi-square distribution. In this section, QQ plots based on the empirical and theoretical distributions of the statistic

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

are used to measure the impact on the performance of the distribution of the sample variance under various of the populations described in Table 1. Poor results derived from the QQ plots may have an important impact on the performance of the customary confidence interval (1), and this issue justifies the use of alternative confidence intervals under situations where the Assumptions ($A1$) and/or ($A2$) are violated, such as the presence of outliers or spikes in the distribution. QQ plots can be seen in Figures 2, 3 and 4, and they contain the empirical distributions based on a total of 10000 samples, with size $n = 200$, selected under simple random sampling without replacement from various of the populations described in Table 1.

Figure 2 about here

Figure 3 about here

Figure 4 about here

Figure 2 contains QQ plots based on the $Normal$ and $Uniform$ populations (see Table 1). As we expected, both empirical and theoretical distributions are similar for the $Normal$ population with small sampling fractions ($f = 0.02$), since the Assumption ($A1$) holds, and the Assumption ($A2$) is approximately satisfied. The Assumption ($A2$) does not hold in the case of large sampling fractions ($f = 0.2$), and for this reason, the empirical distribution differs from the theoretical distribution in this situation. Important differences between the empirical and theoretical distributions can be observed in the $Uniform$ population.

Figure 3 contains QQ plots based on the *Outliers* population. The Assumption ($A1$) is violated as the percentage of outliers (p_{out}) increases, and for this reason, the distance between the empirical and theoretical distribution is greater as p_{out} increases. Finally, the presence of spikes in the distribution is analyzed in Figure 4, which contains QQ plots based on the *Spikes* population. As we expected, both empirical and theoretical distributions are similar when the percentage of missing data (p_{md}) is small. The Assumption ($A1$) is violated as p_{md} increases, hence the presence of spikes has an important impact on the performance of the distribution of Y. Results are slightly better as the sampling fraction decreases.

Figures 2, 3 and 4 indicate that the empirical and theoretical distributions of Y may differ substantially when the Assumption ($A1$) and ($A2$) are violated. In particular, the presence of outliers or spikes have a relevant impact on the performance of the distribution of Y. Similarly, differences between the empirical and theoretical distributions can be observed in the case of large sampling fractions, since the Assumption ($A2$) is violated in this situation. As can be seen in Section 5, the performance of the customary confidence interval for the population variance is poor when Y does not fit well to a Chi-square distribution. For instance, this is the situation of populations that contain outliers or spikes in the distribution, or samples based on large sampling fractions. For the aforementioned reasons, the use of alternative confidence intervals for the population variance under the commented scenarios is thus justified. Confidence intervals based on the rescaled bootstrap method may be a possible solution, as can be seen in Sections 4 and 5.

4. The suggested rescaled bootstrap confidence intervals

The customary confidence interval for the population variance (1) assumes that the Assumptions $(A1)$ and $(A2)$ hold. In this section, we propose using the rescaled bootstrap method ([3, 4, 6, 31, 34]) to construct confidence intervals for the population variance, since this method has some desirable properties. First, confidence intervals based on the rescaled bootstrap method are free of probabilistic distributions, hence the Assumption $(A1)$ is not required. Second, the suggested confidence intervals are based on sampling weights, which implies that they can be easily applied to the case of finite populations. In addition, the rescaled bootstrap method can be easily applied to the case of samples selected from complex sampling designs. Finally, the rescaled bootstrap method is more simple than alternative resampling methods, since it only requires a new set of sampling weights.

We suggest two different confidence intervals for the population variance. Assuming a general sampling design, the common implementation steps for both confidence interval based on the rescaled bootstrap method are as follows:

Step 1. Draw the sample s, with size n, from a finite population and using a general sampling design with inclusion probabilities given by π_i.

Step 2. Then calculate the original sampling weights, which are defined as $d_i = \pi_i^{-1}$, with $i = \{1, \ldots, n\}$. See also [35].

Step 3. Calculate the Hájek-type estimator ([18, 35]) for the population variance σ^2, which is defined as

$$\hat{\sigma}_H^2 = \frac{1}{\hat{N}} \sum_{i=1}^{n} d_i (x_i - \bar{x}_H)^2, \tag{3}$$

where $\hat{N} = \sum_{i=1}^{n} d_i$, and

$$\bar{x}_H = \frac{1}{\hat{N}} \sum_{i=1}^{n} d_i x_i$$

is the sample mean.

Step 4. Then calculate the bootstrap weights d_i^*, which are obtained after using the scale adjustment on d_i suggested by [34].

Step 5. Set the number of bootstrap samples, which will be denoted as B.

Step 6. Calculate the Hájek-type estimator of σ^2 for each bootstrap sample, $b = \{1, ..., B\}$, and which is obtained by replacing d_i by d_i^* into the expression (3), i.e.:

$$\hat{\sigma}^2_{H(b)} = \frac{1}{\hat{N}^*_{(b)}} \sum_{i=1}^{n} d^*_{i(b)} (x_i - \bar{x}_{H(b)})^2,$$

where $\hat{N}^*_{(b)} = \sum_{i=1}^{n} d^*_{i(b)}$, and

$$\bar{x}_{H(b)} = \frac{1}{\hat{N}^*_{(b)}} \sum_{i=1}^{n} d^*_{i(b)} x_i.$$

The first confidence interval is based on the empirical distribution of the bootstrapped values $\hat{\sigma}^2_{H(b)}$, i.e., we consider the percentile bootstrap approximation. The additional implementation step of this method is:

Step 7 (CI_P). Calculate the $100 \times (1-\alpha)\%$ confidence interval, which is defined as

$$CI_P = [L_P, U_P],$$

where the lower and upper limits are defined, respectively, as

$$L_P = \hat{\sigma}^2_H[\alpha/2]$$

and

$$U_P = \hat{\sigma}^2_H[1 - \alpha/2],$$

and where $\hat{\sigma}^2_H[a]$ denotes the ath quantile of the bootstrapped values $\hat{\sigma}^2_{H(b)}$.

The second confidence interval is based on bootstrap distribution of the Student's t-test, which is commonly named as the bootstrap-t method or the studentized bootstrap ([10, 14, 30]). The additional implementation steps of this second method are:

Step 7 (CI_t). Calculate the rescaled bootstrap variance estimator, which is defined as

$$\hat{V}_{boot}(\hat{\sigma}_H^2) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\sigma}_{H(b)}^2 - \hat{\sigma}_H^2)^2.$$

Step 8 (CI_t). Then calculate the empirical t-values, which are given by

$$t_b^* = \frac{\hat{\sigma}_{H(b)}^2 - \hat{\sigma}_H^2}{\hat{V}_{boot}(\hat{\sigma}_H^2)},$$

with $b = 1, \ldots, B$.

Step 9 (CI_t). Then calculate the $100 \times (1-\alpha)\%$ confidence interval, which is defined as

$$CI_t = [L_t, U_t],$$

where the lower and upper limits are given, respectively, by

$$L_t = \hat{\sigma}_H^2 + t_{\alpha/2}^* \sqrt{\hat{V}_{boot}(\hat{\sigma}_H^2)},$$

$$U_t = \hat{\sigma}_H^2 + t_{1-\alpha/2}^* \sqrt{\hat{V}_{boot}(\hat{\sigma}_H^2)},$$

and t_a^* is the ath quantile of the values t_b^*.

5. Monte Carlo simulation studies

In this section, we empirically evaluate the different confidence intervals discussed in this paper. In addition, we observe that both proposed confidence intervals are based on the Hájek-type estimator $\hat{\sigma}_H^2$, and for this reason, the error estimates of $\hat{\sigma}_H^2$ are also calculated. We consider the populations and scenarios described in Section 3.1. It is expected that the customary confidence interval has a poor performance in the scenarios where its underlying assumptions are violated, but an additional aim is to measure the impact on the empirical performance of this confidence interval under the investigated scenarios.

It is quite common to use the empirical relative bias (RB) and the empirical relative root mean square error ($RRMSE$) to evaluate the performance of a given estimator (see [3]). For the Hájek-type estimator, such empirical measures are defined as

$$RB = 100 \times \frac{E[\hat{\sigma}_H^2 - \sigma^2]}{\sigma^2}$$

and

$$RRMSE = 100 \times \frac{MSE[\hat{\sigma}_H^2]}{\sigma^2},$$

where

$$E[\hat{\sigma}_H^2] = \frac{1}{R}\sum_{i=1}^{R} \hat{\sigma}_{Hi}^2$$

and

$$MSE[\hat{\sigma}_H^2] = \frac{1}{R}\sum_{i=1}^{R} (\hat{\sigma}_{Hi}^2 - \sigma^2)^2$$

are, respectively, the expectation and the mean square error based on $R = 10000$ simulation runs, and $\hat{\sigma}_{Hi}^2$ denotes the value of $\hat{\sigma}_H^2$ at the ith simulation run. On the other hand, the empirical performances of the various confidence intervals, with a 95% for the confidence level, are compared in terms of the empirical coverage rate (CR) and the empirical average width (AW). For a given confidence interval with lower and upper limits denoted as L and U, respectively, the CR is defined as

$$CR = 100 \times \frac{1}{R}\sum_{i=1}^{R} \delta\left(L_i \leq \sigma^2 \leq U_i\right),$$

where L_i and U_i are, respectively, the lower and upper limits for the ith simulation run, and $\delta(\cdot)$ is the indicator variable, which takes the value 1 if its argument is true, and $\delta(\cdot) = 0$ otherwise. The AW is the average width of the R confidence intervals calculated in the simulation study, i.e.,

$$AW = \frac{1}{R}\sum_{i=1}^{R} \left(U_i - L_i\right).$$

Note that a method for the construction of confidence intervals with desirable properties should have values of CR close to the nominal level of 95%, and values of AW smaller than the empirical average width of alternative methods. The error estimates of the Hájek-type estimator can be seen in Table 2, whereas Table 3 reports the empirical results derived from the various confidence intervals. The algorithmic efficiency for the problem of calculating the proposed confidence intervals was also measured, and the time efficiency can be approximated by $\mathcal{O}(n^2)$.

Table 2 about here

Table 3 about here

First, we analyze the empirical performance of the Hájek-type estimator (see Table 2). We observe that the empirical biases of $\hat{\sigma}_H^2$ are negligible, with values of RB smaller than 1% for the various populations. As we expected, $\hat{\sigma}_H^2$ is more efficient as the sampling fraction increases, that is, as the sample size increases. The values of $RRMSE$ range from 12.5% (Spikes population with $p_{md} = 70\%$ and $f = 20\%$) to 69.6% (Spikes population with $p_{md} = 10\%$ and $f = 2\%$).

From Table 3, we first use the $Outliers$ population to analyze the impact of the presence of outliers on the various confidence intervals. We observe that the customary confidence interval is extremely conservative, since the empirical coverage rates are 100% when the percentage of outliers is large ($p_{out} = 40\%$), and they are close to this upper bound when $p_{out} = 30\%$. However, the customary confidence interval gives reasonable empirical coverage rates when $p_{out} = 20\%$. The confidence interval proposed by [5] is also very conservative in the presence of outliers, with values of CR close to 100%. In addition, both alternative confidence intervals (CI_{χ^2} and CI_B) are wider than the suggested rescaled bootstrap confidence intervals, which in turn have reasonable coverage rates (close to the nominal level of 95%). In particular, the values of CR are between 94.5% and 96.5% in the presence of outliers.

The presence of spikes in the distribution (see the $Spikes$ population) also has an important impact on the customary confidence interval, since we observe values of CR between 65.8% and 82.8% for large values of p_{md}, which are considerably smaller than the nominal level. Reasonable empirical coverage rates are observed for the suggested confidence intervals, with values of CR between 92.9% (observed when the percentage of missing data is extremely large, $p_{md} = 70\%$) and 96.2% for the various scenarios of the *Spikes* population. The values of AW are similar for the Bonett and suggested confidence intervals, although the Bonett confidence interval is slightly worse than the suggested confidence intervals in terms of empirical coverage rates.

As we expected, desirable results are observed for the customary confidence interval in the $Normal$ population and when the sampling fraction is small ($f = 2\%$), since both Assumptions ($A1$) and ($A2$) hold. However, this method is conservative as the sampling fraction increases, i.e., when the Assumption ($A2$) is violated. Reasonable empirical coverage rates are also observed for the suggested confidence intervals (values of CR are between 93.7% and 96.7%), and the corresponding values of AW are also smaller than the empirical average widths of alternative confidence intervals.

The customary and Bonett confidence intervals are generally very conservative in the $Uniform$ population, and the values of AW are larger than the empirical average widths of the suggested confidence intervals, which in turn have empirical coverage rates between 93.7% and 96.8%.

Finally, we also observe desirable empirical coverage rates for the suggested confidence intervals in the population based on the real data set (the *Rods* population). The confidence interval CI_B is very conservative, whereas the empirical coverage rates of CI_{χ^2} are considerably smaller than the nominal level in the *Rods* population. Note that similar results are observed for the suggested confidence intervals (CI_P and CI_t) and for the various populations considered in this study.

6. Concluding remarks

The customary confidence interval for the population variance (see equation (1)) performs well when the underlying Assumptions hold. However, this confidence interval is very sensitive when such assumptions are violated, i.e., the interval CI_{χ^2} may have coverage rates extremely far from the required nominal level under certain scenarios. For instance, the values of CR for CI_{χ^2} are 32.5% and 36.5% in the real population considered in this study (the *Rods* population), and many of them are close to the upper bound (100%) in various of the remaining populations analyzed in Section 5. Note that such results are supported by the conclusions derived from [26]. [5] proposed an alternative confidence interval for the population variance (see equation (2)), which is based on the estimation of the coefficient of kurtosis. Both confidence intervals (CI_{χ^2} and CI_B) were empirically compared under populations generated from various probabilistic distributions.

Monte Carlo simulation studies can be seen in Sections 3 and 5. First, we identify the limitations of the customary confidence interval under various of the scenarios investigated in this paper: (*i*) populations based on the Normal and Uniform probabilistic distributions; (*ii*) large and small sampling fractions; (*iii*) presence of outliers, and; (*iv*) presence of spikes in the distribution. The second Monte Carlo simulation study is used to analyze the empirical performance of the suggested confidence intervals, and results are compared to the customary (CI_{χ^2}) and Bonett (CI_B) confidence intervals. The suggested rescaled bootstrap confidence intervals have reasonable empirical coverage rates, with values close to the required nominal level, and the values of AW are generally smaller than the average widths of alternative confidence intervals. The interval CI_B is very conservative in various of the populations considered in Section 5. As far as the presence of outliers is concerned, beyond the fact that the empirical coverage of the proposed confidence intervals are close to the required nominal level and their competitors are very conservative, a relevant advantage is that the average widths of the suggested confidence intervals are almost half of the values AW of the customary and bonett confidence intervals. The presence of spikes in the distribution (see the *Spike* and *Rods* populations) has an important impact on the customary confidence interval, since excessively small coverage rates are observed.

Various limitations of this study and some topics for further research are now explained. As noted by [5], the estimator $\widehat{\gamma}_4$ of the coefficient of kurtosis can be substituted by a pooled estimator based on information collected from a previous study ($\widehat{\gamma}_4^*$), i.e., additional valuable prior information can be introduced at the estimation stage. Rescaled bootstrap confidence intervals for the population variance and based on prior information could be an interesting topic for further research in the near future. One-sided confidence intervals are very common in the problem of estimating the population variance, hence some research efforts on this topic are also welcome. Following [8], it could be interesting to analyze the rescaled bootstrap confidence interval for the population variance with bootstrap samples selected without replacement. The extension of the suggested technique to stratified multistage designs is also a relevant topic for further research in the near future, since this sampling design is quite common in large scaled sample surveys, and it could be used to a wide range of reweighting methods. Finally, it is well known that the use of auxiliary information at the estimation stage may report more accurate estimates, and for this reason, the construction of rescaled bootstrap confidence intervals for the population variance and based on auxiliary variables is a topic that may provide confidence interval with better properties.

References

[1] Aggarwal, C. C. 2017. *Outlier analysis*. 2nd ed. Cham: Springer.

[2] Ahmad, T. 1997. A resampling technique for complex survey data. *Journal of the Indian Society of Agricultural Statistics* 50 (3):364–79.

[3] Berger, Y. G., and J. F. Muñoz. 2015. On estimating quantiles using auxiliary information. *Journal of Official Statistics* 31 (1):101–19.

[4] Biswas, A., A. Rai, and T. Ahmad. 2020. Rescaling bootstrap technique for variance estimation for ranked set samples in finite population. *Communications in Statistics-Simulation and Computation* 49 (10):2704–2718.

[5] Bonett, D. G. 2006. Approximate confidence interval for standard deviation of nonnormal distributions. *Computational Statistics & Data Analysis* 50 (3):775–82.

[6] Bruch, C. 2019. Applying the rescaling bootstrap under imputation: A simulation study. *Journal of Statistical Computation and Simulation* 89 (4):641–59.

[7] Chen, J., and J. Shao. 2000. Nearest neighbor imputation for survey data. *Journal of Official Statistics* 16 (2):113–31.

[8] Chipperfield, J., and J. Preston. 2007. Efficient bootstrap for business surveys. *Survey Methodology* 33 (2):167–72.

[9] Čiginas, A., and D. Pumputis. 2014. A note on the upper bound to variance of the sample extreme from a finite population. *Communications in Statistics-Theory and Methods* 43 (8):1793–99.

[10] Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

[11] Deville, J. C., and C. E. Särndal. 1992. Calibration estimators in survey sampling. *Journal of the American Statistical Association* 87 (418):376–82.

[12] Diana, G., and M. Giordan. 2012. Finite population variance estimation in presence of measurement errors. *Communications in Statistics-Theory and Methods* 41 (23):4302–14.

[13] Dong, Y., and C. Y. J. Peng. 2013. Principled missing data methods for researchers. *SpringerPlus* 2 (1):1–17.

[14] Efron, B., and R. J. Tibshirani. 1994. *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.

[15] European Commission. 2018. *European Union Statistics on Income and Living Conditions (EU-SILC)*. Luxembourg: European Commission.

[16] Ghalamfarsa Mostofi, A., and M. Kharrati-Kopaei. 2017. Ratio-and difference-type estimators for estimating finite population variance. *Communications in Statistics-Theory and Methods* 46 (7):3542–55.

[17] Ghosh, S. 2020. A note on using the empirical moment generating function to estimate the variance of nonparametric trend estimates from independent time series replicates. *Communications in Statistics-Simulation and Computation* 49 (9): 2287–2301.

[18] Hájek, J. 1971. Comment on a paper by D. Basu. In *Foundations of statistical inference*, ed. V. R. Godambe and D. A. Sprott, 236. Toronto: Holt, Rinehart and Winston.

[19] Hu, M. X., and S. Salvucci. 2001. *A study of imputation algorithms*. Washington, DC: National Center for Education Statistics.

[20] Kalton, G., and D. Kasprzyk. 1986. The treatment of missing survey data. *Survey Methodology* 12 (1):1–16.

[21] Landerman, L. R., K. C. Land, and C. F. Pieper. 1997. An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociological Methods & Research* 26 (1):3–33.

[22] Lehmann, E. L., and J. P. Romano. 2005. *Testing statistical hypotheses*. 3rd ed. New York: Springer.

[23] Little, R. J. A., and D. B. Rubin. 2019. *Statistical analysis with missing data*. 3rd ed. Hoboken: Wiley.

[24] Maji, R., G. N. Singh, and A. Bandyopadhyay. 2017. Effective estimation strategy of finite population variance using multi-auxiliary variables in double sampling. *Journal of Modern Applied Statistical Methods* 16 (1):158–78.

[25] Martínez, S., M. Rueda, A. Arcos, H. Martínez, and J. F. Muñoz. 2012. On determining the calibration equations to construct model-calibration

estimators of the distribution function. *Revista Matemática Complutense* 25 (1):87–95.

[26] Miller, R. G. 1986. *Beyond ANOVA. Basics of applied statistics*. New York: Wiley.

[27] Mitra, A. 2016. *Fundamentals of quality control and improvement*. 4th ed. Hoboken: John Wiley & Sons, Inc.

[28] Molenberghs, G., G. Fitzmaurice, M. G. Kenward, A. Tsiatis, and G. Verbeke. 2014. *Handbook of missing data methodology*. Boca Raton: Chapman & Hall/CRC.

[29] Montgomery, D. C. 2019. *Introduction to statistical quality control*. 8th ed. New York: John Wiley & Sons, Inc.

[30] Moran, T. P. 2006. Statistical inference for measures of inequality with a cross-national bootstrap application. *Sociological Methods & Research* 34 (3):296–333.

[31] Muñoz, J. F., E. Álvarez-Verdejo, and R. M. García-Fernández. 2018. On estimating the poverty gap and the poverty severity indices with auxiliary information. *Sociological Methods & Research* 47 (3):598–625.

[32] Preston, J. 2009. Rescaled bootstrap for stratified multistage sampling. *Survey Methodology* 35 (2):227–34.

[33] Rao, J. N. K., and C. F. J. Wu. 1988. Resampling inference with complex survey data. *Journal of the American Statistical Association* 83 (401):231–41.

[34] Rao, J. N. K., C. F. J. Wu, and K. Yue. 1992. Some recent work on resampling methods for complex surveys. *Survey Methodology* 18 (2):209–17.

[35] Särndal, C. E., B. Swensson, and J. H. Wretman. 2003. *Model Assisted Survey Sampling*. Springer Science & Business Media.

[36] Sengupta, S. 2018. Admissible unbiased estimation of finite population variance under a randomized response model. *Communications in Statistics-Theory and Methods* 47 (20):5077–82.

[37] Shao, J., and D. Tu. 1995. *The jackknife and bootstrap.* New York: Springer-Verlag.

[38] Singh, H. P., and S. K. Pal. 2016. An efficient class of estimators of finite population variance using quartiles. *Journal of Applied Statistics* 43 (10):1945–58.

[39] Singh, H. P., S. K. Pal, and R. S. Solanki. 2014. A new procedure for estimation of finite population variance using auxiliary information. *Journal of Reliability and Statistical Studies* 7 (2):149–60.

[40] Singh, H. P., S. K. Pal, and A. Yadav. 2018. A study on the chain ratio-ratio-type exponential estimator for finite population variance. *Communications in Statistics-Theory and Methods* 47 (6):1442–58.

[41] Song, Z., Y. Liu, Z. Li, and J. Zhang. 2018. A comparative study of memory-type control charts based on robust scale estimators. *Quality and Reliability Engineering International* 34 (6):1079–102.

[42] Song, Z., Y. Liu, Z. Li, and J. Zhang. 2018. A weighted likelihood ratio test-based chart for monitoring process mean and variability. *Journal of Statistical Computation and Simulation* 88 (7):1415–36.

[43] Tabachnick, B. G., and L. S. Fidell. 2019. *Using multivariate statistics.* 7th ed. Boston: Pearson.

[44] Tate, R. F., and G. W. Klett. 1959. Optimal confidence intervals for the variance of a normal distribution. *Journal of the American Statistical Association* 54 (287):674–82.

[45] Van Buuren, S. 2018. *Flexible imputation of missing data.* 2nd ed. Boca Raton: Chapman & Hall/CRC.

[46] Wolter, K. 2007. *Introduction to variance estimation.* 2nd ed. New York: Springer.

[47] Wu, J. C., A. F. Martin, and R. N. Kacker. 2019. Monte Carlo studies of bootstrap variability in ROC analysis with data dependency. *Communications in Statistics-Simulation and Computation* 48 (2):317–33.

[48] Zhang, S., H. Cao, Z. Ye, Y. Zhang, and X. Hei. 2019. An outlier detection scheme for dynamical sequential datasets. *Communications in Statistics-Simulation and Computation* 48 (5):1450–502.

Table 1: Description of the real and simulated populations considered in the various Monte Carlo Simulation Studies. Samples, with size $n = 200$, are selected under simple random sampling without replacement. N is the population size, and $f = n/N$ is the sampling fraction.

Type of population		Code	N	f
Simulated	Normal and outliers	$Outliers$	1000	0.20
			10000	0.02
	Normal and spikes	$Spikes$	1000	0.20
			10000	0.02
	Normal	$Normal$	1000	0.20
			10000	0.02
	Uniform	$Uniform$	1000	0.20
			10000	0.02
Real	Rods	$Rods$	500	0.40
			1000	0.20

Table 2: Values of the empirical measures RB (relative bias) and $RRMSE$ (Relative root mean square error) for the various populations (see Table 1) and the Hájek-type estimator $\hat{\sigma}^2_H$. $\theta = \{p_{out}, p_{md}\}$ for the Outliers and Spikes populations, respectively, and $\theta = \sigma^2$ for the Normal and Uniform populations. CV is the coefficient of variation (in percentage).

Population	θ	CV	f	RB	$RRMSE$
Outliers	0.4	25.05	0.20	-0.4	40.3
			0.02	-0.6	56.7
	0.3	24.80	0.20	-0.4	37.6
			0.02	-0.5	45.6
	0.2	23.22	0.20	-0.5	54.2
			0.02	-0.4	43.9
Spikes	0.7	5.47	0.20	-0.1	12.5
			0.02	-0.5	51.5
	0.5	7.07	0.20	-0.5	48.6
			0.02	-0.6	62.3
	0.3	8.36	0.20	-0.3	32.3
			0.02	-0.4	36.2
	0.1	9.47	0.20	-0.4	36.5
			0.02	-0.7	69.6
Normal	5.0	22.35	0.20	-0.2	17.6
			0.02	-0.4	39.8
	1.0	10.00	0.20	-0.5	51.5
			0.02	-0.5	53.3
Uniform	5.0	22.38	0.20	-0.4	42.6
			0.02	-0.5	54.5
	1.0	10.00	0.20	-0.6	55.3
			0.02	-0.5	45.2
**Rods*	–	0.02	0.4	-0.5	49.8
			0.2	-0.5	53.8

Table 3: Values of the empirical measures AW (average width) and CR (coverage rate) for the various populations (see Table 1) and confidence intervals. $\theta = \{p_{out}, p_{md}\}$ for the Outliers and Spikes populations, respectively, and $\theta = \sigma^2$ for the Normal and Uniform populations. * The values of AW are multiplied by 1000 in the Rods population.

			Customary		Bonett		Bootstrap proposed			
			CI_{χ^2}		CI_B		CI_P		CI_t	
Population	θ	f	AW	CR	AW	CR	AW	CR	AW	CR
Outliers	0.4	0.20	3.9	100.0	3.3	100.0	2.0	96.5	2.0	96.5
		0.02	3.8	100.0	3.2	99.8	1.9	94.6	1.9	94.6
	0.3	0.20	3.4	99.7	3.4	99.7	2.4	96.5	2.4	96.5
		0.02	3.4	99.4	3.4	99.5	2.4	94.8	2.4	94.8
	0.2	0.20	2.7	96.2	3.4	99.3	2.7	96.1	2.7	96.1
		0.02	2.7	93.8	3.4	98.5	2.7	94.5	2.7	94.5
Spikes	0.7	0.20	0.1	70.0	0.3	96.6	0.3	94.8	0.3	94.8
		0.02	0.1	65.8	0.2	95.0	0.2	92.9	0.2	92.9
	0.5	0.20	0.2	82.8	0.3	97.4	0.3	95.7	0.3	95.7
		0.02	0.2	78.4	0.3	95.9	0.3	93.1	0.3	93.1
	0.3	0.20	0.3	91.4	0.4	98.2	0.3	96.2	0.3	96.2
		0.02	0.3	87.2	0.4	96.7	0.3	94.1	0.3	94.1
	0.1	0.20	0.4	96.1	0.5	98.8	0.4	96.0	0.4	96.0
		0.02	0.4	92.8	0.4	97.4	0.4	93.7	0.4	93.7
Normal	5.0	0.20	2.0	98.4	2.3	99.5	1.8	96.7	1.8	96.7
		0.02	2.0	95.7	2.3	98.0	1.9	94.4	1.9	94.4
	1.0	0.20	0.4	97.8	0.5	99.2	0.4	96.4	0.4	96.4
		0.02	0.4	95.0	0.5	97.3	0.4	93.7	0.4	93.7
Uniform	5.0	0.20	2.0	100.0	1.9	100.0	1.3	96.8	1.3	96.8
		0.02	2.0	99.9	1.8	99.6	1.2	94.1	1.2	94.1
	1.0	0.20	2.1	92.0	2.8	97.5	2.3	93.7	2.3	93.7
		0.02	0.4	100.0	0.4	99.5	0.3	96.1	0.3	96.1
**Rods*	–	0.4	0.2	36.5	1.4	98.3	1.1	95.4	1.1	95.4
		0.2	0.2	32.5	1.4	96.3	1.1	93.2	1.1	93.2

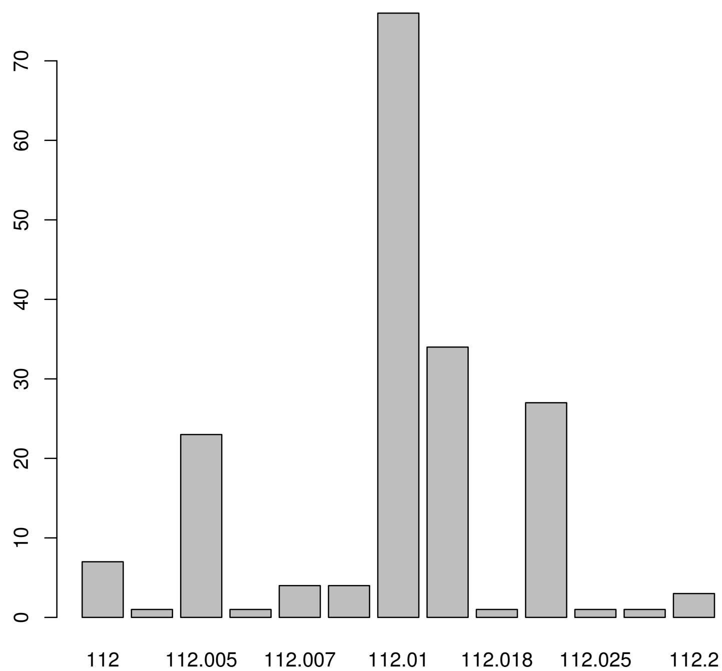

Figure 1: Bar chart for the *Rods* population (183 measurements of the inside diameter of connecting rods manufactured for large diesel engines).

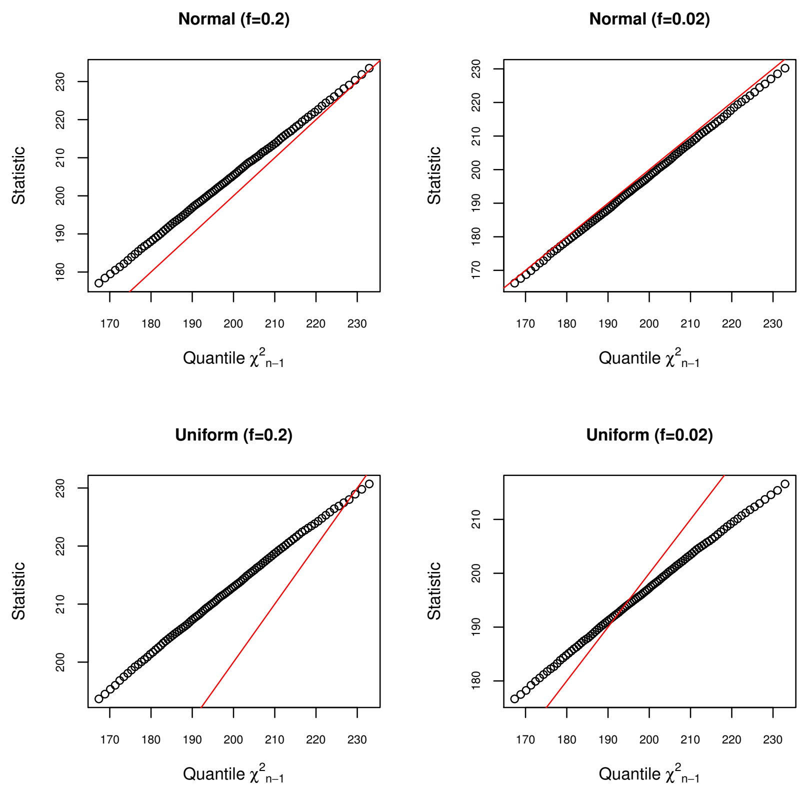

Figure 2: QQ plots for quantiles of the theoretical distribution χ^2_{n-1} and quantiles of the empirical distribution of the statistic $Y = (n-1)S^2/\sigma^2$, where $\sigma^2 = 1$. Samples, with size $n = 200$, are selected under simple random sampling without replacement from the $Normal$ and $Uniform$ populations. The population sizes $N = \{1000, 10000\}$ are considered, which implies that the sampling fractions take the values $f = \{0.02, 0.2\}$.

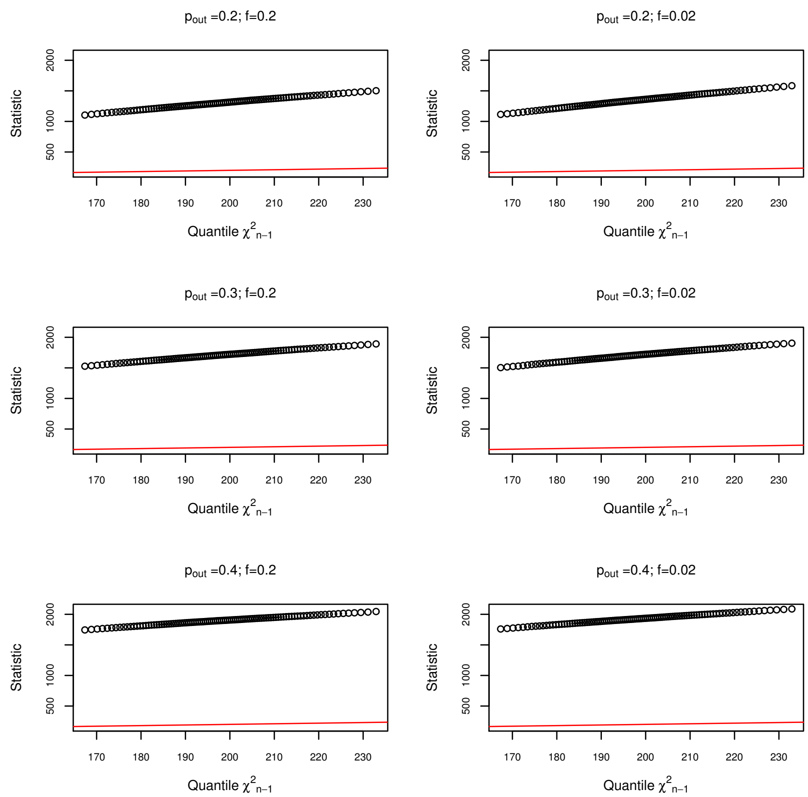

Figure 3: QQ plots for quantiles of probabilistic distribution χ^2_{n-1} and quantiles of the empirical distribution of the statistic $Y = (n-1)S^2/\sigma^2$, where $\sigma^2 = 1$. Samples, with size $n = 200$, are selected under simple random sampling without replacement from the *Outliers* population. Different percentages of outliers are considered ($p_{out} = \{0.2, 0.3, 0.4\}$). The population sizes $N = \{1000, 10000\}$ are considered, which implies that the sampling fractions take the values $f = \{0.02, 0.2\}$.

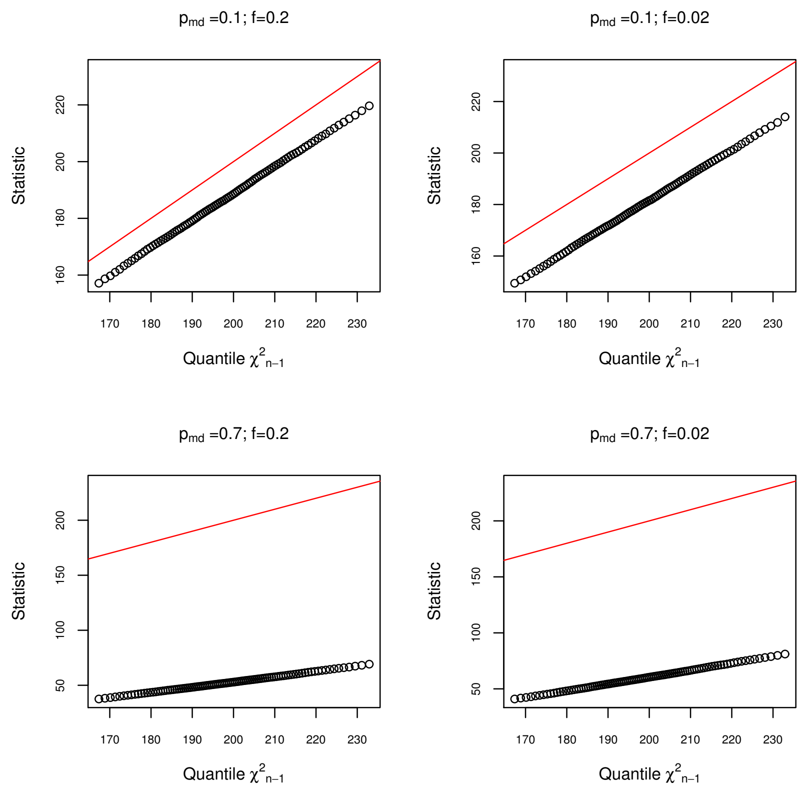

Figure 4: QQ plots for quantiles of the theoretical distribution χ^2_{n-1} and quantiles of the empirical distribution of the statistic $Y = (n-1)S^2/\sigma^2$, where $\sigma^2 = 1$. Samples, with size $n = 200$, are selected under simple random sampling without replacement from the *Spikes* population. Different percentages of missing data are considered ($p_{md} = \{0.1, 0.7\}$). The population sizes $N = \{1000, 10000\}$ are considered, which implies that the sampling fractions take the values $f = \{0.02, 0.2\}$.

Figure captions

- Figure 1: Bar chart for the *Rods* population (183 measurements of the inside diameter of connecting rods manufactured for large diesel engines).

- Figure 2: QQ plots for quantiles of the theoretical distribution χ^2_{n-1} and quantiles of the empirical distribution of the statistic $Y = (n-1)S^2/\sigma^2$, where $\sigma^2 = 1$. Samples, with size $n = 200$, are selected under simple random sampling without replacement from the $Normal$ and $Uniform$ populations. The population sizes $N = \{1000, 10000\}$ are considered, which implies that the sampling fractions take the values $f = \{0.02, 0.2\}$.

- Figure 3: QQ plots for quantiles of probabilistic distribution χ^2_{n-1} and quantiles of the empirical distribution of the statistic $Y = (n-1)S^2/\sigma^2$, where $\sigma^2 = 1$. Samples, with size $n = 200$, are selected under simple random sampling without replacement from the $Outliers$ population. Different percentages of outliers are considered ($p_{out} = \{0.2, 0.3, 0.4\}$). The population sizes $N = \{1000, 10000\}$ are considered, which implies that the sampling fractions take the values $f = \{0.02, 0.2\}$.

- Figure 4: QQ plots for quantiles of the theoretical distribution χ^2_{n-1} and quantiles of the empirical distribution of the statistic $Y = (n-1)S^2/\sigma^2$, where $\sigma^2 = 1$. Samples, with size $n = 200$, are selected under simple random sampling without replacement from the *Spikes* population. Different percentages of missing data are considered ($p_{md} = \{0.1, 0.7\}$). The population sizes $N = \{1000, 10000\}$ are considered, which implies that the sampling fractions take the values $f = \{0.02, 0.2\}$.