# REVISITING JACKKNIFE CONFIDENCE INTERVALS

**Janete F. R. Aguirre**
Departamento de Estatística, ICEx, Universidade Federal de Minas Gerais
31270-901 – Belo Horizonte – MG, Brasil
janete.aguirre@gmail.com

**F. R. B. Cruz**
Departamento de Estatística, ICEx, Universidade Federal de Minas Gerais
31270-901 – Belo Horizonte – MG, Brasil
fcruz@est.ufmg.br

**RESUMO**

O jackknife é revisitado nesse artigo, sendo uma de suas finalidades a divulgação do método para estimação intervalar de parâmetros em modelos estatísticos de interesse prático. É mostrado como o jackknife pode ser facilmente implementável e apresentar um desempenho satisfatório conforme evidenciado pelas simulações exemplificadas. Além disso, apesar de ser um método computacionalmente intensivo, seu uso é atualmente de fácil acesso e viável em computadores pessoais, conforme apresentado neste artigo.

**PALAVRAS CHAVE. Jackknife, Intervalos de Confiança, Reamostragem.**

**Tópicos: EST&MP - Estatística e Modelos Probabilísticos; SIM - Simulação**

**ABSTRACT**

The jackknife is revisited in this article, in which the method is showcased as being still useful and resourceful while estimating statistical intervals for parameters in models of practical interest. It is shown that the jackknife is easily implemented and can perform fairly well, as shown in some exemplified situations. Furthermore, despite being a computationally intensive method, its use is perfectly feasible on common PCs, as evidenced by the simulations presented.

**KEYWORDS. Jackknife. Confidence Intervals. Resampling Method.**

**Paper topics: EST&MP - Statistics and Probabilistic Models; SIM - Simulation**

## 1. Introduction

It is well known that statistical inference plays a crucial role in operations research as part of its decision-making tools. Interval estimation of parameters is one of these tools and the focus of this article. More specifically, jackknife confidence intervals are studied and exemplified here. Jackknifing is a commonly employed statistical technique for estimation of variances of sample statistics. Quenouille introduced the method in 1949 for the limited purpose of correcting possible bias in the estimates [Quenouille, 1956]. In 1958, Tukey noticed that the procedure could be used to construct reasonably reliable confidence interval for a wide variety of estimators as well [Miller, 1974].

This article aims to investigate coverage of jackknife confidence intervals by evaluating its performance to estimate the standard deviation for a few distributions. By making use of computer intensive re-sampling techniques, the idea was to build a computational routine using the software R [R Core Team, 2020] as a tool to access and compare confidence intervals calculated by the jackknife method to classically calculated ones.

The remaining of this article is organized as follows. Sections 2 and 3 detail the methodology. Section 4 presents and discusses the experimental results. Finally, Section 5 concludes with closing remarks.

## 2. The Jackknife Method

As explained by Severiano et al., the delete-one jackknife relies on re-samples that leave out one entity of the sample at a time, where $n$ entities are those individuals that are randomly sampled from the population [Severiano et al., 2011]. A pseudo-value approach is used to calculate the jackknife confidence interval (CIs). For an estimator $S$, the $i$-th pseudo-value of $S$ was calculated as

$$S_i' = nS - (n-1)S_i, \tag{1}$$

where $S_i$ is the estimator value for the sample with the $i$-th data value deleted.

The jackknife CI at the $(1-\alpha)100\%$ level is then calculated as

$$\mathrm{CI_J}(1-\alpha) = \bar{S}' \pm t_{\alpha/2;\, n-1}(s/\sqrt{n}), \tag{2}$$

being $\bar{S}' = \dfrac{1}{n}\sum_{i=1}^{n} S_i'$, and $t_{\alpha/2;\, n-1}$ the $(1-\alpha/2)$ $t$-student percentile, with $(n-1)$ degrees of freedom.

The jackknife can be thought of as a method for converting the problem of estimating any population parameter into the problem of estimating a population mean [Manly, 2007].

## 3. Development

Three distributions were chosen to access and compare the coverage of such calculated intervals for the standard deviation: Normal, Exponential, and Poisson. Considering these mentioned distributions, traditional form of interval calculation for the standard deviation were performed as according to the formulae found in the literature [Casella e Berger, 2002]. That is, for a normal population,

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2;\, n-1}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2;\, n-1}}}, \tag{3}$$

being $s^2$ the variance of the sample, defined as

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

in which $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\chi^2_{\alpha/2;\, n-1}$ and $\chi^2_{1-\alpha/2;\, n-1}$ are, respectively, the $\alpha/2$ and $(1-\alpha)/2$ $\chi^2$ percentiles with $(n-1)$ degrees of freedom; for an exponential population,

$$\frac{2y}{\chi^2_{\alpha/2;\, 2n}} \leq \sigma \leq \frac{2y}{\chi^2_{1-\alpha/2;\, 2n}}, \tag{4}$$

being $y = \sum_{i=1}^{n} x_i$; for a Poisson population,

$$\sqrt{\frac{\frac{1}{2n}}{\chi^2_{1-\alpha/2;\, 2y}}} \leq \sigma \leq \sqrt{\frac{\frac{1}{2n}}{\chi^2_{\alpha/2;\, 2(y+1)}}}. \tag{5}$$

Following Eqs. (3)–(5), the traditional CIs can be calculated in R [R Core Team, 2020] as shown in Table 1.

Table 1: Traditional CI Lower and Upper Bounds

| Distribution | R Code for Traditional CIs |
|---|---|
| Normal | `CIstdNo <- function(smp,alpha) {`<br>`  n <- length(smp)`<br>`  df <- n-1`<br>`  varsmp <- var(smp)`<br>`  lower2 <- (varsmp*df) /`<br>`    qchisq((alpha/2), df, lower.tail = FALSE)`<br>`  upper2 <- (varsmp*df) /`<br>`    qchisq((1 - alpha/2), df, lower.tail = FALSE)`<br>`  return(list(lower=sqrt(lower2),upper=sqrt(upper2)))`<br>`}` |
| Exponential | `CIstdEx <- function(smp,alpha) {`<br>`  n <- length(smp)`<br>`  df <- 2*n`<br>`  lower <-((2*sum(smp) /`<br>`    (qchisq((alpha/2), df, lower.tail = FALSE))))`<br>`  upper <-((2*sum(smp) /`<br>`    (qchisq((1 - alpha/2), df, lower.tail = FALSE))))`<br>`  return(list(lower=lower,upper=upper))`<br>`}` |
| Poisson | `CIstdPo <- function(smp,alpha) {`<br>`  n <- length(smp)`<br>`  dfL <- 2*sum(smp)`<br>`  dfU <- 2*(sum(smp)+1)`<br>`  lower2 <- (1/(2*n)) *`<br>`    qchisq((1-alpha/2), dfL, lower.tail = FALSE)`<br>`  upper2 <- (1/(2*n)) *`<br>`    qchisq((alpha/2), dfU, lower.tail = FALSE)`<br>`  return(list(lower=sqrt(lower2),upper=sqrt(upper2)))`<br>`}` |

```
CIstdJk <- function(smp,alpha) {
  n <- length(smp)
  pseudov <- numeric(n)
  smpjack <- numeric(n-1)
  sdsmp <- StdEst(smp)
  for (i in 1:n) {
    # remove i-th sample element
    smpjack <- smp[-i]
    sdjack <- StdEst(smpjack)
    # pseudo-value
    pseudov[i] <- (n * sdsmp) - ((n-1)*sdjack)
  }
  # point estimate and standard jackknife error
  epj <- mean(pseudov)
  erj <- sd(pseudov)/sqrt(n)
  alphaux <- 1 - alpha
  lower <- epj-qt(1-alpha/2,df=n-1)*erj
  upper <- epj+qt(1-alpha/2,df=n-1)*erj
  return(list(lower=lower,upper=upper))
}
```

Listing 1: R code to jackknife CIs

```
StdEst<-function(smp) {
  n <- length(smp)
  return(sqrt(sum((smp - mean(smp))^2)/n))
}
```

Listing 2: Point estimator for population standard deviation

A computational routine was developed and written to calculate the jackknife confidence intervals as it can be seen in Listing 1, noticing that the function in R that implements the jackknife uses the function `StdEst`, which is the point estimator whose confidence interval is sought. In particular, Listing 2 presents the implementation of a point estimator for standard deviation.

## 4. Results

The scripts developed are available upon request, directly from the authors, for practitioners and researchers in the field. The experiments were run on a common PC running Microsoft Windows(c) 10 with an Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30 GHz, and 8.00 GB RAM. Samples with six different sizes, *i.e.*, $n = \{10, 20, 50, 100, 200, 500\}$, were randomly generated and, for each size, at the 95% level, the traditional CI (Traditional) and the intensive computational confidence interval (Jackknife) were calculated alongside. The procedure was repeated 1,000 times for each size group as a short form of Monte Carlo method and the fraction of these comprised within the interval for the parameter used in the generation ($\sigma = 2$) is registered in Table 2.

From the displayed output for the coverage performance achieved (Table 2) it can be noticed that for the three sampled distributions, the number of intervals that included the parameter was considerably smaller in the jackknife counts. Besides, from the same table data, Figure 1 was constructed from the mean coverage of both kinds of studied intervals considering the chosen sample sizes. From that, it can be seen, in general, that jackknife intervals are smaller in coverage than the traditional ones, even though they converge unto the nominal value (95%) as the sample size is enlarged, what can be considered rather encouraging.

## 5. Conclusions

In this article the problem of estimating confidence intervals by the jackknife method was addressed, along with an analysis via Monte Carlo simulation of the coverage of confidence intervals

Table 2: Confidence Interval Coverage for a confidence level of 95%

| | Normal | | Exponential | | Poisson | |
|---|---|---|---|---|---|---|
| Sample Size ($n$) | Traditional | Jackknife | Traditional | Jackknife | Traditional | Jackknife |
| 10 | 0.945 | 0.907 | 0.955 | 0.770 | 0.956 | 0.864 |
| 20 | 0.947 | 0.911 | 0.937 | 0.791 | 0.969 | 0.909 |
| 50 | 0.948 | 0.931 | 0.950 | 0.849 | 0.959 | 0.927 |
| 100 | 0.954 | 0.947 | 0.961 | 0.890 | 0.948 | 0.922 |
| 200 | 0.956 | 0.950 | 0.947 | 0.903 | 0.958 | 0.935 |
| 500 | 0.961 | 0.956 | 0.948 | 0.931 | 0.954 | 0.948 |

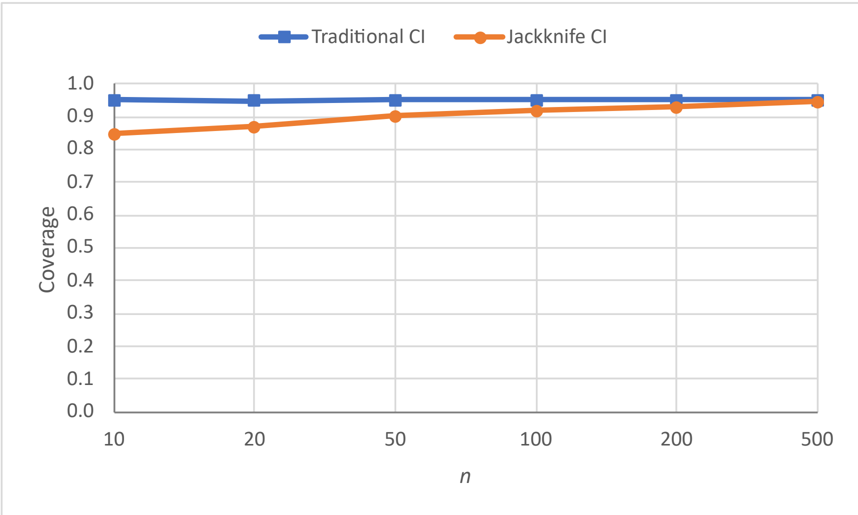


Figure 1: Mean coverage as function of the sample size $n$

for the standard deviation. The following example may illustrate the usefulness of the presented technique. Let us consider the data available in Triola [Triola, 2018], in which the waiting times in a single queue for service in three cashiers, $\mathrm{x}_a = \{6.5, 6.6, 6.7, 6.8, 7.1, 7.3, 7.4, 7.7, 7.7, 7.7\}$, and waiting times in three different queues for service at each of three cashiers, $\mathrm{x}_b = \{4.2, 5.4, 5.8, 6.2, 6.7, 7.7, 7.7, 8.5, 9.3, 10.0\}$, are compared. The average handling time is the same for both instances (that is, $\bar{x} = 7.15$). It should be clear, though, that the determination of confidence intervals for the standard deviations in the two situations is crucial to define whether there is a statistically significant difference between the variability of the two options. If one of them is the best, that is, if there is a difference, which one has the lowest handling time variability and, hence, providing the best customer experience.

Thus simulations exploring the parameter space showed that the jackknife 95% CI has, in several instances, a lower coverage for a range of sample sizes. Despite the rather lower results for the exponential parameter, overall jackknife results complied well with the classic coverage. It is a relatively easy method to use despite its slightly poor performance. Such study clarifies the limitations when the jacknife method is used to estimate CIs. Nevertheless it produced consistent and rather similar measurements. It brings to light how procedures like this were resourceful during a period which lacked computational resources for intensive and repetitive calculations.

Taking into account the above stated example from Triola, when the jackknife technique was applied, the following confidence intervals were obtained for the standard deviation of the samples, $\text{CI}_{\text{Jck-a}}(95\%) = [0.3537179, 0.6081586]$ and $\text{CI}_{\text{Jck-b}}(95\%) = [1.096171, 2.618754]$, evidencing that the single-queue configuration (option $a$) has a significantly lower standard deviation than an individual queue for each of the three cashiers (option $b$), that is, there is no intersection between the CIs at the 95% confidence level. The jackknife interval, Eq. (2), may be preferable since no distribution is assumed for the sample service times. On the other hand, considering that the data can approximately follow a normal distribution and using the traditional estimator, Eq. (3), the following estimates are obtained, $\text{CI}_{\text{Trd-a}}(95\%) = [0.3278761, 0.8702288]$ and $\text{CI}_{\text{Trd-b}}(95\%) = [1.252981, 3.325585]$, leading to the same decision, although the traditional intervals are wider than the jackknife ones. As a final remark, being an intensive computational method, jackknife CIs can take a greater processing time than to calculate the traditional ones. Nevertheless, this time is not high enough to make its use unfeasible. For example, for the above calculations, less than 2 ms were taken for both jackknife and traditional.

Topics for future research in this area include the investigation of second-order jackknife based procedures [Quenouille, 1956], that is, with pseudo-values computed as:

$$S''_{i,j} = \frac{n^2 S'_i - (n-1)^2 S'_{i,j}}{n^2 - (n-1)^2}, \tag{6}$$

where $S'_{i,j}$ is the estimator value for the sample with the $i$-th and $j$-th data points deleted, for all $i$ and $j$, such that $i \neq j$.

## Acknowledgments

FRBC acknowledges CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*, grant 305515/2018-7) and FAPEMIG (*Fundação de Amparo à Pesquisa do Estado de Minas Gerais*, grant CEX-PPM-00564-17) for partial financial support. J. Aguirre acknowledges CAPES for financial support (88887.62654/2021-00 under *Programa de Demanda Social* at UFMG).

## References

Casella, G. e Berger, R. L. (2002). *Statistical inference*. Duxbury, Thomson Learning Inc., Pacific Grove, CA, second edition.

Manly, B. F. J. (2007). *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman & Hall/CRC, Boca Raton, FL, third edition.

Miller, R. G. (1974). The jackknife – A review. *Biometrika*, 61(1):1–15.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43(3/4):353–360.

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

Severiano, A., Carriço, J. A., Robinson, D. A., Ramirez, M., e Pinto, F. R. (2011). Evaluation of jackknife and bootstrap for defining confidence intervals for pairwise agreement measures. *PLoS One*, 6(5):e19539.

Triola, M. F. (2018). *Elementary Statistics*. Pearson, Boston, MA, 13th edition.