The Jackknife in Monte Carlo Studies

Dennis D. Boos and Jason Osborne
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203
boos@stat.ncsu.edu
January 30, 2013

NC State Department of Statistics Technical Report # 2648

Summary

The jackknife has a long history as a tool for reducing bias and estimating variances. Here we highlight the use of the jackknife to produce standard errors in Monte Carlo studies and illustrate it with F-statistics also based on the jackknife.

Key words and phrases: Monte Carlo studies; standard errors; k-sample comparisons; Oneway F-statistic.

1 Introduction

The *jackknife* was introduced by Quenouille (1949, 1956) as a general method to remove bias from estimators, and Tukey (1958) suggested its use for variance and interval estimation. Here we want to discuss the use of jackknife variance estimation in Monte Carlo studies and illustrate it using an F-statistic that is also based on jackknife variance estimates.

Our basic points are

1. Any table of Monte Carlo estimates should include a summary of standard errors for each different type of estimate in the table.

2. Jackknife standard errors are so simple and effective for use in Monte Carlo studies that they should automatically be included as part of almost any analysis of simulations.

We hope that 1. is an absolute core tenet for any member of the statistics profession. The fact that journal articles do not always reflect this core tenet brings us to point 2. We want to demonstrate the ease of computing jackknife standard errors in Monte Carlo simulation analyses, therefore encouraging statistical practice to follow core belief.

In a typical Monte Carlo study, we might be comparing several estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ and associated variance estimates based on N independent Monte Carlo (computer-generated) random samples. For each sample, suppose that we compute $(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{V}_1, \widehat{V}_2)$, where $\widehat{V}_1$ and $\widehat{V}_2$ are the variance estimates for $\widehat{\theta}_1$ and $\widehat{\theta}_2$, respectively. Thus we have an N by 4 matrix of estimators and associated variance estimators, where the rows are independent but there is dependence within rows. We shall call this matrix the "Monte Carlo result matrix $\boldsymbol{X}$."

A variety of summaries computed from $\boldsymbol{X}$ could be of interest: bias in the estimators, comparison of Monte Carlo estimated variances of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ (i.e., sample variances $S^2_{1,N}$ and $S^2_{2,N}$ of the $\widehat{\theta}_1$ and $\widehat{\theta}_2$ columns), and comparison of the average of the variance estimator $\widehat{V}_1$ and $\widehat{V}_2$ columns (say $\overline{\widehat{V}}_{1,N}$ and $\overline{\widehat{V}}_{2,N}$) to the Monte Carlo estimated variances $S^2_{1,N}$ and $S^2_{2,N}$, respectively. For the latter two types of summaries, our preference is to use ratios like $S^2_{1,N}/S^2_{2,N}$ and $\overline{\widehat{V}}_{1,N}/S^2_{1,N}$. These ratios are easy to interpret and, coupled with stan-

dard errors, give immediate inference about equality of expectations of the numerators and denominators (H_0 : true ratio = 1.0).

Basic asymptotic analysis via the delta method reveals how to compute the Monte Carlo standard errors of the above ratios. However, most of us do not want to go through those details every time we run a simulation, especially since the dependence within rows of the Monte Carlo output makes the analysis more complicated. However, the jackknife does delta method calculus automatically. The jackknife does not care how complicated the summaries are as long as they are not nonregular summaries like a maximum or nonsmooth summaries like percentiles. (Note, however, that the estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ being studied do not need to be regular.) Virtually any regular Monte Carlo summary is fair game; examples in Section 3 include the ratios above and standard estimates of skewness and kurtosis.

To keep notation and ideas clear, we use the term "Monte Carlo standard error" and denote it by $\mathrm{SE}_{\mathrm{MC}}$ for any estimated standard deviation of a Monte Carlo summary computed from a sample of size N of vectors, each vector computed from a computer-generated random sample. Continuing the earlier example, the vectors were $(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{V}_1, \widehat{V}_2)$, and the obvious $\mathrm{SE}_{\mathrm{MC}}$ for the Monte Carlo mean $\overline{\widehat{V}}_{1,N}$ of variance estimates is just the sample standard deviation of the N values of $\widehat{V}_1$ divided by $\sqrt{N}$. Our main goal is to encourage use of the jackknife for calculation of $\mathrm{SE}_{\mathrm{MC}}$ for more complicated summaries than sample means. Precise language and notation are important because Monte Carlo studies involve two kinds of samples, the original N samples or datasets, and then the resulting sample of N vectors of estimates (the Monte Carlo result matrix $\boldsymbol{X}$), computed from those datasets. $\mathrm{SE}_{\mathrm{MC}}$ is based on this latter sample of size N.

For illustration we perform a small Monte Carlo study of an F-statistic used to test the equality of coefficients of variation across k independent samples. This F-statistic is also based on the jackknife and is related to the method introduced by Miller (1968) for testing equality of variances in k independent samples. A variety of tests are available in the literature to compare coefficients of variation. However, many of these are based on assuming normality of the samples and are very sensitive to nonnormality. This jackknife F-statistic approach is actually a general method that can be used to quickly test for equality of any

regular parameter in k samples. Here we are mainly using it to illustrate the use of the jackknife for getting standard errors in Monte Carlo studies, but the simulations suggest it can be a competitive method.

A brief outline of the paper is as follows. Section 2 reviews the basics of the jackknife along with a motivating data example. Section 3 gives the illustrative Monte Carlo study along with detailed explanation of how to use the jackknife to obtain SE_{MC} for summaries. R code is sprinkled throughout in order to be clear about practical use. The complete R code to construct Table 2 and Figure 1 are given in the Supplementary Material website. The Appendix gives the background and justification for the F-statistic that motivates the Monte Carlo study. R code Table 2 and Figure 1 are found at `http://www4.stat.ncsu.edu/~boos/software`.

2 Jackknife Basics

For an estimator $\widehat{\theta}$ based on an iid sample $Y_1, \ldots, Y_n$, let $\widehat{\theta}_{[i]}$ be the "leave-1-out" estimator obtained by computing $\widehat{\theta}$ with Y_i deleted from the sample. We denote the average of these "leave-1-out" estimators by $\overline{\theta}_1 = n^{-1} \sum_{i=1}^n \widehat{\theta}_{[i]}$ and define the pseudo-values by

$$\widehat{\theta}_{ps,i} = n\widehat{\theta} - (n-1)\widehat{\theta}_{[i]}. \tag{1}$$

The average of these pseudo-values is the bias-adjusted jackknife estimator

$$\begin{aligned} \widehat{\theta}_{\text{J}} = \frac{1}{n}\sum_{i=1}^n \widehat{\theta}_{ps,i} &= n\widehat{\theta} - (n-1)\overline{\theta}_1 \\ &= \widehat{\theta} - (n-1)(\overline{\theta}_1 - \widehat{\theta}). \end{aligned}$$

The jackknife variance estimator for $\widehat{\theta}$ or $\widehat{\theta}_{\text{J}}$ is

$$\widehat{V}_{\text{J}} = \frac{n-1}{n}\sum_{i=1}^n \left(\widehat{\theta}_{[i]} - \overline{\theta}_1\right)^2 = \frac{1}{n}\,\frac{1}{n-1}\sum_{i=1}^n (\widehat{\theta}_{ps,i} - \widehat{\theta}_{\text{J}})^2. \tag{2}$$

The last expression is in the form of a sample variance of the $\widehat{\theta}_{ps,i}$ divided by n, much like the variance of the sample mean. In fact, when $\widehat{\theta} = \overline{Y}$, then $\widehat{\theta}_{ps,i} = Y_i$ and the jackknife

Table 1: Biomass of a New Variety of Centipede

										$\overline{Y}$	s	CV (SE)
Seed:	1	2	79	5	17	11	2	15	85	24.1	33.4	1.38 (0.34)
Sprig:	37	60	48	14	76	23				43.0	23.2	0.54 (0.16)
Combo:	3	61	7	5	27	25	35	17		22.5	19.4	0.86 (0.24)

reproduces results for the sample mean. More generally, suppose that $\widehat{\theta}$ has the typical approximation by averages representation

$$\widehat{\theta} - \theta = \frac{1}{n}\sum_{i=1}^{n} IC(Y_i, \theta) + R_n, \tag{3}$$

where $\sqrt{n}R_n \xrightarrow{p} 0$ as $n \to \infty$, θ is the probability limit of $\widehat{\theta}$, and $IC(Y_i, \theta)$ is the influence curve (Hampel, 1974). If $\mathrm{E}\{IC(Y_1, \theta)\} = 0$ and $\mathrm{E}_F\{IC(Y_1, \theta)\}^2 = \sigma^2_{\mathrm{IC}}$ exists, then by the Central Limit Theorem, $\widehat{\theta}$ is $\mathrm{AN}(\theta, \sigma^2_{\mathrm{IC}}/n)$. Delta method and influence curve methods estimate σ^2_{IC} by a sample variance of $\widehat{IC}(Y_i, \widehat{\theta})$, where $\widehat{IC}$ means that the influence curve has been estimated in addition to replacing θ by $\widehat{\theta}$. Using (3), one can show that $\widehat{\theta}_{ps,i} - \widehat{\theta} \approx \widehat{IC}(Y_i, \widehat{\boldsymbol{\theta}})$ and thus the jackknife variance estimate $\widehat{V}_{\mathrm{J}}$ in (2) is just an approximate version of the delta and influence curve methods. An appealing feature of $\widehat{V}_{\mathrm{J}}$ is that you do not need to know calculus or $IC(Y_i, \theta)$ in order to calculate $\widehat{V}_{\mathrm{J}}$. You just need a computer program.

To illustrate the calculations, consider the data in Table 1 on the biomass from small plots at the end of one year for three types of planting methods for a new variety of centipede grass (from seed, sprigs, or a combination of seed and sprigs). There is interest in comparing the means of the three planting methods as well as the standard deviations, but it might make more sense to compare the relative standard deviations, i.e., the coefficients of variation (CV). Using seed is much cheaper than using sprigs, but the relative standard deviation 1.38 appears high compared to the 0.54 for sprigs or to the 0.86 for a combination of sprigs and seeds. We have added SE $= (\widehat{V}_{\mathrm{J}})^{1/2}$ in parentheses, and now explain their R computation.

To get the standard errors in Table 1, we first give R code for the jackknife variance estimate adapted from the Appendix of Efron and Tibshirani (1993):

```
jack.var <- function(x, theta, ...){
       n <- length(x)
       u <- rep(0, n)
       for(i in 1:n) {
           u[i] <- theta(x[ - i], ...)   # leave-1-out estimators
       }
       jack.var <- ((n - 1)/n) * sum((u - mean(u))^2)
       return(jack.var)
}
```

For those less familiar with R, the ... in the first line of code allows extra arguments of the function `theta` to be included, as in `jack.var(x,theta=mean,trim=.10)` to obtain $\widehat{V}_{\text{J}}$ for the 10% trimmed mean. Putting the the first row of Table 1 in data vector `y`, the call producing $\widehat{V}_{\text{J}}$ is `jack.var(y,theta=cv)`, where `cv=function(y){sd(y)/mean(y)}`. Repeating for rows two and three and taking square roots gives standard errors 0.34, 0.16, and 0.24, respectively.

To make a formal comparison of the coefficients of variation, we consider using an F-statistic given in (5) of the Appendix. Applied to the estimated CV's in Table 1 coupled with their associated $\widehat{V}_{\text{J}}$, F =2.20 with associated p-value = 0.14. Thus, there is a suggestion of real differences in CV but not clear evidence.

This data example motivates us to consider several questions:

1. Is $\widehat{V}_{\text{J}}$ approximately unbiased for estimating Var($\widehat{\text{CV}}$)?

2. Is the p-value associated with F from (5) approximately valid?

In the next Section we use Monte Carlo simulation to investigate these questions. The main focus, though, is to illustrate use of $\text{SE}_{\text{MC}} = (\widehat{V}_{\text{J}})^{1/2}$ for Monte Carlo study summaries.

3 Monte Carlo Study

To answer the questions posed above, we first study the sample coefficient of variation for a sample $Y_1, \ldots, Y_n$ from a single population, $\widehat{\theta} = s/\overline{Y}$, where $\overline{Y} = n^{-1}\sum_{i=1}^{n} Y_i$ and $s^2 =$

$(n-1)^{-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$. We also study the natural logarithm $\log(\widehat{\theta})$ and the corresponding jackknife estimators $\widehat{\theta}_{\text{J}}$ and $\log(\widehat{\theta})_{\text{J}}$. We want to see which estimators are approximately normal and whether the associated jackknife variance estimators are close in expectation to the true variances.

To this end, we generated $N = 4,000$ samples of size $n = 10$ from normal and gamma($\alpha = 2$) distributions having coefficients of variation equal to 0.1, 0.2, and 0.5. For each of these six distributions, we calculated the four estimators described above for each sample and also the associated jackknife estimators of variance. Note that the same $\widehat{V}_{\text{J}}$ is associated with both $\widehat{\theta}$ and $\widehat{\theta}_{\text{J}}$, and $\widehat{V}_{\text{J}}$ based on $\log(\widehat{\theta})$ is associated with both $\log(\widehat{\theta})$ and $\log(\widehat{\theta})_{\text{J}}$. So our Monte Carlo result matrix $\boldsymbol{X}$ consists of N vectors of length 6, where the first 3 elements are $(\widehat{\theta}, \widehat{\theta}_{\text{J}}, \widehat{V}_{\text{J}})$, and the second three elements are the same quantities for $\log(\widehat{\theta})$. Summaries will be based on averages over N of the two $\widehat{V}_{\text{J}}$, denoted generically by $\overline{\widehat{V}_{\text{J}}}$, and by the sample variances of the four basic estimators, denoted generically by S^2 for the $\widehat{\theta}$ and $\log(\widehat{\theta})$ and by S^2_{J} for the $\widehat{\theta}_{\text{J}}$ versions.

Table 2 contains the results. The first column of each group, RB = Relative Bias $= \overline{\widehat{V}_{\text{J}}}/S^2$, is estimating $\text{E}(\widehat{V}_{\text{J}})$ divided by the estimated true variance for each of the four estimators. "Relative Bias" may be an abuse of statistical language, but we prefer to note whether this ratio is near 1 rather than subtract 1 and compare to 0. For normal data, because RB > 1 in Table 2, the jackknife variance estimator $\widehat{V}_{\text{J}}$ is overestimating the true variance of $\widehat{\theta}$ and $\widehat{\theta}_{\text{J}}$ in every case. For gamma data, however, $\widehat{V}_{\text{J}}$ is underestimating the variance of the jackknife estimator $\widehat{\theta}_{\text{J}}$, likely because reducing bias adds variance to $\widehat{\theta}_{\text{J}}$. The size of the standard errors (SE_{MC}) in the footnote allow us to make these statements with confidence.

The second columns in Table 2 are RV $= S^2/S^2_{\text{J}}$, where S^2 is the Monte Carlo sample variance of $\widehat{\theta}$ and S^2_{J} is the Monte Carlo sample variance of $\widehat{\theta}_{\text{J}}$. These ratios show that $\widehat{\theta}_{\text{J}}$ has higher variance than $\widehat{\theta}$ for both $\widehat{\text{CV}}$ and $\log(\widehat{\text{CV}})$ in all but one case. Columns 3 and 4 report the Monte Carlo estimates of the skewness and kurtosis coefficients Skew $= \text{E}\{X-\text{E}(X)\}^3/\left[\text{E}\{X-\text{E}(X)\}^2\right]^{3/2}$ and Kurt $= \text{E}\{X-\text{E}(X)\}^4/\left[\text{E}\{X-\text{E}(X)\}^2\right]^2$. Because normal data have Skew $= 0$ and Kurt $= 3$, we see that the sampling distribution of $\widehat{\text{CV}}$ is right-skewed and $\log(\widehat{\text{CV}})$ tends to be left-skewed, and only $\widehat{\text{CV}}_{\text{J}}$ tends to have longer tails

Table 2: Relative bias (RB) of jackknife variance estimators and ratio of variances (RV) of $\widehat{\theta}$ to $\widehat{\theta}_{\text{J}}$, where $\widehat{\theta}$ is the sample coefficient of variation ($\widehat{\text{CV}}$) and $\log(\widehat{\text{CV}})$. Based on $N = 4,000$ samples of size $n = 10$ from normal and gamma populations. RB = ratio of $\overline{\widehat{V}_{\text{J}}}$ = Monte Carlo average of $\widehat{V}_{\text{J}}$ to S^2 = Monte Carlo sample variance of estimators. RV = S^2/S^2_{J}.

					cv=0.1				
	N(10,1)					Gamma(2,1)+12.1			
	RB	RV	Skew	Kurt		RB	RV	Skew	Kurt
$\widehat{\text{CV}}$	1.13	.94	0.3	3.3		1.10	.75	0.8	4.3
$\widehat{\text{CV}}_{\text{J}}$	1.06		0.3	3.3		0.83		1.2	6.1
$\log(\widehat{\text{CV}})$	1.24	.96	−0.5	3.5		1.23	.73	−0.2	3.4
$\log(\widehat{\text{CV}})_{\text{J}}$	1.19		−0.4	3.5		0.90		0.3	4.2
					cv=0.2				
	N(5,1)					Gamma(2,1)+5.1			
$\widehat{\text{CV}}$	1.14	.96	0.3	3.1		1.14	.72	0.7	4.1
$\widehat{\text{CV}}_{\text{J}}$	1.09		0.3	3.1		0.82		1.2	6.4
$\log(\widehat{\text{CV}})$	1.26	.97	−0.5	3.4		1.30	.70	−0.2	3.1
$\log(\widehat{\text{CV}})_{\text{J}}$	1.22		−0.4	3.4		0.91		0.4	4.1
					cv=0.5				
	N(2,1)					Gamma(2,1)+0.8			
$\widehat{\text{CV}}$	1.09	1.12	0.8	4.2		1.22	.63	0.6	3.7
$\widehat{\text{CV}}_{\text{J}}$	1.22		0.6	3.7		0.77		1.3	6.8
$\log(\widehat{\text{CV}})$	1.15	1.02	−0.2	3.2		1.29	.67	−0.3	3.3
$\log(\widehat{\text{CV}})_{\text{J}}$	1.17		−0.3	3.2		0.86		0.5	4.5

Range of standard errors (SE_{MC}): (.02,.04) for RB, (.003,.01) for RV, (.04,.16) for Skew, (.08,1.5) for Kurt.

(Kurt > 3) than normal distributions for the gamma cases.

Now we turn our attention to the standard errors in the note at the bottom of Table 2, called $\mathrm{SE}_{\mathrm{MC}}$ in the Introduction. All of the 4 Monte Carlo summaries (RB, RV, Skew, Kurt) are ratios of correlated sample moments computed from the Monte Carlo result matrix $\boldsymbol{X}$ of $N = 4,000$ independent replications. As mentioned in the Introduction, we could derive delta method asymptotic approximations to get standard errors for (RB, RV, Skew, Kurt), but using the jackknife is so much easier.

For simple statistics like the mean of the first column of $\boldsymbol{X}$ that estimates $\mathrm{E}(\widehat{\theta})$ (not displayed in Table 2), the R code using `jack.var` from Section 2 to obtain $\widehat{V}_{\mathrm{J}}$ is

```
var.J <- jack.var(x=X[,1],theta=mean)
```

In this case we do not need the jackknife because the exact same result is obtained by `var(X[,1])/N`. The jackknife variance estimate for the skewness coefficent estimate is a similar trivial R calculation,

```
skew <- function(x){mean((x-mean(x))^3)/(mean((x-mean(x))^2)^(1.5))}
var.J <- jack.var(x=X[,1],theta=skew)
```

but an asymptotic approach based on the delta method would be very time-consuming to derive and program.

For statistics that are functions of a vector like a ratio of two sample variances, the jackknife computation is slightly more complicated. R code to compute RV in Table 2 and its jackknife variance estimate is

```
ratio.var <- function(index,xdata){var(xdata[index,1])/var(xdata[index,2])}
var.J <- jack.var(1:N,theta=ratio.var,xdata=X[,c(1,2)])
```

The salient feature here is that `jack.var` needs to receive "data" `1:N` so that it can drop out one row at a time to compute the leave-1-out estimators. Thus the function `ratio.var`

also needs to receive data `1:N`, and the real data in the first two columns of $\boldsymbol{X}$ comes in as the additional argument `xdata=X[,c(1,2)]`. In the supplementary website material, we give the full R code to produce Table 2.

Once standard errors for every table entry have been computed, it is easy to report a summary of these standard errors. In Table 2 we have given ranges of standard errors in the footnote. Other summaries such as the mean or median or maximum (see caption to Figure 1) might also be used. Such summaries could be placed at the bottom of columns or in the table caption. There should be freedom to make the standard error summaries informative but not overly obtrusive. Students sometimes go to the extreme and put a standard error in parentheses beside each entry. This often clutters the table and detracts from the main message in the table.

Monte Carlo Study Part II

Theorem 1 of the Appendix supports use of all four estimators studied above to test equality of k coefficients of variation using the F-statistic (5). The Table 2 simulations indicate that $\widehat{V}_{\text{J}}$ tends to overestimate the variance of $\widehat{\text{CV}}$ and $\log(\widehat{\text{CV}})$ in all situations considered there. For this reason, we anticipate that F-statistics using them might control test size better than F-statistics based on $\widehat{\text{CV}}_{\text{J}}$ and $\log(\widehat{\text{CV}})_{\text{J}}$ where $\widehat{V}_{\text{J}}$ underestimates variance for gamma populations. However, it makes sense to carry out a Monte Carlo study on test size and power for all four F-statistics. Here we only study size.

Thus we generated $N = 10,000$ samples with equal coefficients of variation from the same normal and gamma populations as in Table 2 and for the following designs:

Design 1: $k = 2$, $n_1 = n_2 = 10$
Design 2: $k = 4$, $n_1 = n_2 = n_3 = n_4 = 10$
Design 3: $k = 4$, $n_1 = n_2 = 5, n_3 = n_4 = 10$
Design 4: $k = 8$, $n_i = 5, i = 1, \ldots, 8$

For each design and estimators $\widehat{\text{CV}}$, $\widehat{\text{CV}}_{\text{J}}$, $\log(\widehat{\text{CV}})$, and $\log(\widehat{\text{CV}})_{\text{J}}$ along with their associated $\widehat{V}_{\text{J}}$, we calculated the F-statistic (5) and estimated the probability of exceeding the 95th percentile of an $F(k-1, \sum n_i - k)$ distribution. Figure 1 displays all 48 estimated rejection

rates with ($\widehat{\text{CV}}$, $\widehat{\text{CV}}_{\text{J}}$) in the left panel and (log($\widehat{\text{CV}}$), log($\widehat{\text{CV}}$)$_{\text{J}}$) in the right panel. Within each panel, the normal distribution results are plotted on the left half of the panel, and the gamma distribution results are on the right half. The open circles are for $\widehat{\text{CV}}$ and log($\widehat{\text{CV}}$), and the related filled circles are for $\widehat{\text{CV}}_{\text{J}}$ and log($\widehat{\text{CV}}$)$_{\text{J}}$ and are directly above or below the open circles. Results are ordered from left to right for the four designs (labeled D1-D4), and CVs (0.1,0.2,0.5) are ordered within each design from left to right and connected by lines.

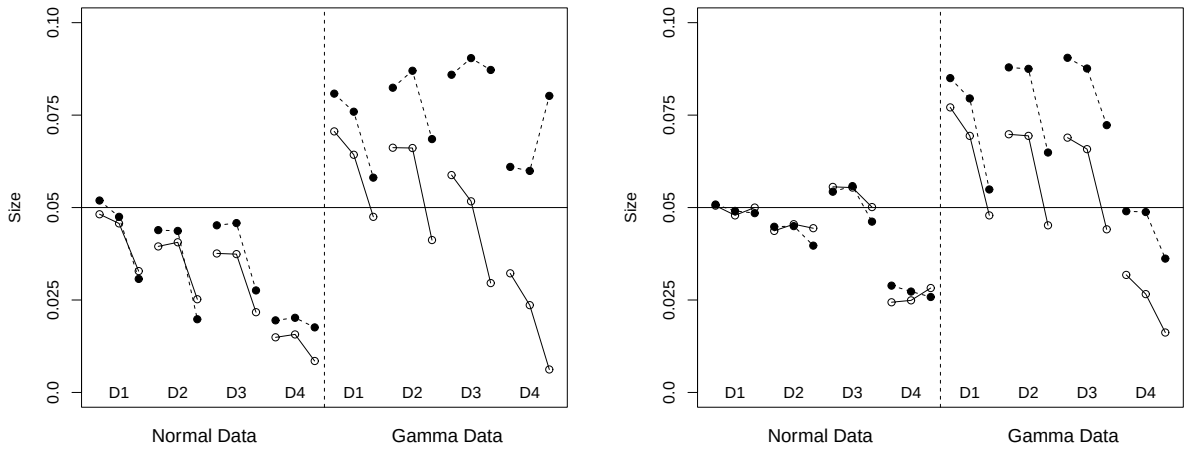

Figure 1: Empirical rejection rates of nominal 0.05 level F-tests for equality of coefficient of variation. The left panel is based on $\widehat{\text{CV}}$ (open circles) and $\widehat{\text{CV}}_{\text{J}}$ (filled circles), and the right panel is similar, based on log($\widehat{\text{CV}}$) and log($\widehat{\text{CV}}$)$_{\text{J}}$. The left half of each panel is for normal data, the right half for gamma data. Lines connect the CVs (0.1,0.2,0.5) within each design (D1-D4). Each point has standard error bounded by $\{.1(.9)/10000\}^{1/2} = 0.003$.

Some conclusions:

1. Rejection rates for normal data are lower than than those for gamma data.

2. Because open circles tend to be lower than filled circles, rejection rates for tests based on the original estimators are lower than those based on $\widehat{\text{CV}}_{\text{J}}$ and log($\widehat{\text{CV}}$)$_{\text{J}}$. This corresponds to the fact that most RV values in Table 2 are less than 1.

3. Rejection rates for Design 4 with eight samples of size 5 tend to be lower than the other designs.

4. For gamma distributions and possibly skewed distributions in general, the rejection rates based on $\widehat{\mathrm{CV}}_{\mathrm{J}}$ and $\log(\widehat{\mathrm{CV}})_{\mathrm{J}}$ are mostly above 0.05. This reflects the fact their corrresponding RB values in Table 2 are less than 1 and suggests a preference for using the tests based on the original estimators.

This small study suggests that using the F-statistic based on $\widehat{\mathrm{CV}} = s/\overline{Y}$ (or on $\log(\widehat{\mathrm{CV}})$) may be useful for testing the equality of coefficients of variation without making distributional assumptions. Certainly more simulations and power comparison with existing procedures would be needed to fully justify the approach.

4 Summary

We think that jackknife standard errors should be routinely used in reporting Monte Carlo results. They are much easier to obtain than those from the standard delta method approach, especially when using the Monte Carlo descriptive ratio statistics discussed in Section 3. The F-statistic studied in Section 3, based on individual estimators and their jackknife variance estimate in each sample, is a potentially useful method for testing equality of parameters from k populations.

ACKNOWLEDGEMENTS

This work was supported by NSF grant DMS-0906421 and NIH grant P01 CA142538-01.

Appendix

Here we discuss briefly the F-test for comparing coefficients of variation and analyzed in Section 3. It is based on a general method introduced by Miller (1968) for comparing parameters from k populations using jackknife pseudo-values.

Suppose that the data consists of k iid samples $Y_{i1}, \ldots, Y_{in_i}$, $i =, \ldots, k$, where $\mathrm{E}(Y_{ij}) = \mu_i$ and $\mathrm{Var}(Y_{ij}) = \sigma_i^2$, and all samples are independent. Miller (1968) was focused on comparing variances, $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$, but used the pseudo-values $\widehat{\theta}_{ps,ij}$ of the natural logarithm of the sample variances $s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_i)^2$, $i = 1, \ldots, k$, where $\overline{Y}_i = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$, because $\log(s_i^2)$ and the associated pseudo-values tend to be more normally distributed than the sample variances and their pseudo-values. These $\log(s_i^2)$ pseudo-values were then used in place of the Y_{ij} in the oneway ANOVA F-statistic for comparing means

$$F = \frac{\sum_{i=1}^{k} n_i (\overline{Y}_i - \overline{\overline{Y}})^2}{(k-1) s_p^2}, \tag{4}$$

where $s_p^2 = \sum_{i=1}^{k} (n_i - 1) s_i^2 / (n - k)$ is the pooled sample variance, $\overline{\overline{Y}} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij} / n$, and $n = \sum_{i=1}^{k} n_i$. Then F is compared to the percentiles of an F distribution with $k-1$ and $n-k$ degrees fo freedom. Because the above F-statistic is comparing sample means, when applied to pesudo-values, the resulting F-statistic is comparing $\widehat{\theta}_{\mathrm{J}}$ from the k samples, and the sample variances are replaced by $\widehat{V}_{\mathrm{J}}$ times n_i.

Here we present a theorem to justify this approach for any set of parameter estimates with associated variance estimates, not just jackknife versions. In particular the following theorem also applies to using the original estimators $\widehat{\theta}_i$ for each sample along with jackknife variance estimators, in the F-statistic. Note that the assumed asymptotic results imply that Theorem 1 is for the null hypothesis $H_0 : \theta_1 = \theta_2 = \cdots = \theta_k = \theta$.

Theorem 1 *Suppose that we have k independent pairs $(\widehat{\theta}_1, \widehat{V}_1), \ldots, (\widehat{\theta}_k, \widehat{V}_k)$ satisfying*

1. $\sqrt{n_i}(\widehat{\theta}_i - \theta) \xrightarrow{d} N(\theta, AV) \quad i = 1, \ldots, k;$
2. $n_i \widehat{V}_i \xrightarrow{p} AV \quad i = 1, \ldots, k,$

as $min(n_1, \ldots, n_k) \longrightarrow \infty$ with $\lambda_{in} = n_i / n \longrightarrow \lambda_i$, $0 < \lambda_i < 1$, $i = 1, \ldots, k$. Then for

$$F = \frac{\sum_{i=1}^{k} n_i (\widehat{\theta}_i - \overline{\widehat{\theta}})^2}{(k-1) \widehat{V}_p}, \quad \widehat{V}_p = \sum_{i=1}^{k} (n_i - 1) \widehat{V}_i n_i / (n-k), \quad \overline{\widehat{\theta}} = \sum_{i=1}^{k} n_i \widehat{\theta}_i / n, \tag{5}$$

we have

$$(k-1)F \xrightarrow{d} \chi^2_{k-1}.$$

Proof. It is easy to show that the numerator of F is a quadratic form $\boldsymbol{Z}_n\boldsymbol{C}_n\boldsymbol{Z}_n^T$, where $\boldsymbol{Z}_n$ is the asymptotically normal vector $\left\{\sqrt{n}(\widehat{\theta}_i-\theta),\ldots,\sqrt{n}(\widehat{\theta}_k-\theta)\right\}$, and $\boldsymbol{C}_n = \boldsymbol{I}_k - \boldsymbol{\lambda}_n\boldsymbol{\lambda}_n^T$ with $\boldsymbol{\lambda}_n = (n_1/n,\ldots,n_k/n)^T$. Thus $\boldsymbol{Z}_n\boldsymbol{C}_n\boldsymbol{Z}_n^T$ converges in distribution to AV times a χ^2_{k-1} random variable. Because of the assumed convergences of the individual $n_i\widehat{V}_i$, the same convergence holds for the pooled version, and Slutsky's Theorem completes the proof.

For applying Theorem 1 to the sample coefficient of variation $\widehat{\theta} = s/\overline{Y}$, we note that the delta theorem, along with finite 4th moments and a nonzero mean, yields Condition 1. of Theorem 1 with $AV = \theta^4 - (\text{Skew})\theta^3 + (1/4)(\text{Kurt}-1)\theta^2$ (see Serfling, 1980, p. 137). For $\log(\widehat{\theta})$, we similarly have $AV = \theta^2 - (\text{Skew})\theta + (1/4)(\text{Kurt}-1)$. Shao and Tu (1996, p. 65) also mention that the jackknife estimators $\widehat{\theta}_{\text{J}}$ and $\log(\widehat{\theta}_{\text{J}})$ have the same asymptotic normal distributions. Note that an implied assumption is that AV is the same in each of the k populations, thus requiring Skew and Kurt to be the same in each population. Likely in practice the distributions of the samples being compared would not be all that different in shape; so equal Skew and Kurt in the k populations seems like a mild assumption. Condition 2. of Theorem 1 is satisfied for $\widehat{V}_{\text{J}}$ by Theorem 2.1 of Shao and Tu (1996, p. 25).

References

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall: New York.

Hampel, F. G. (1968), "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association* , **69**, 567-582.

Miller, R. R. (1974), "Jackknifing Variances," *The Annals of Mathematical Statistics*, **39**, 383-393.

Quenouille, M. H. (1949), "Approximate Use of Correlation in Time Series," *Journal of the Royal Statistical Society, Series B*, **11**,18-84.

Quenouille, M. H. (1956), "Notes on Bias in Estimation," *Biometrika*, **43**, 353-360.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics* Wiley, New York.

Shao, J, and Tu, D. (1996), *The Jackknife and Bootstrap*, Springer: New York.

Tukey, J. (1958), "Bias and Confidence in not Quite Large Samples (abstract)," *The Annals of Mathematical Statistics*, **29**, 614-614.