

Improved jackknife estimates for median variance in equally weighted samples using percentile scale based calculations

John P. D. Martin

Thursday, February 5, 2015

Executive Summary

This paper demonstrates the improvement in accuracy of jackknife estimates of the classical confidence intervals of medians from equally weighted (or unweighted) samples using calculations based on the percentile scale rather than the original scale of the observations. The method used is similar to the Woodruff approach for weighted samples (1) and the improved estimates are consistent with the expectation that suitably transforming the reference frame can improve variance estimation accuracy (2).

As shown in Table 1, the jackknife estimates of median confidence intervals for unweighted samples conducted in the percentile scale, and then backtransformed to the original scale are found to be highly consistent with bootstrap estimates for several common distributions; uniform, normal, log-normal and skewed bivariate normal.

Table 1: Estimated median 95% confidence intervals of unweighted samples for different re-sampling methods

Distribution	popn median	bootstrap	original scale jackknife	percentile scale jackknife
random uniform(0,1)	0.5	(0.4612,0.5218)	(0.4402,0.5513)	(0.4612,0.5218)
normal N(0,1)	0	(-0.0993,0.0576)	(-0.08749,.02576)	(-.09972,.05772)
systematic uniform(0,1)	0.5	(0.4532,0.5494)	(0.4534,0.5492)	(0.4545,0.5505)
log-normal exp(N(0.1,1))	1.09	(1.026,1.18)	(0.9434,1.236)	(1.026,1.181)
bivariate normal	76.65	(74.7,77.73)	(75.98,77.75)	(74.71,77.74)

To highlight the need for a locally linear (smooth) quality to the unweighted sample distribution (2,3,4) in the region of the median point, in order for jackknife median variance calculations to match the performance of bootstrap median variance calculations. The third example in Table 1, is the highly artificial case of evenly spaced observations. In this rare linear example in the original scale, as expected, the regular jackknife calculation approach also closely matches the bootstrap confidence interval estimate.

Various aspects of transforming the median variance estimation calculation to the percentile scale, have been in long use by statistical agencies such as the Australian Bureau of Statistics and Westat for weighted data. From Rogers 2003 (1), in particular, "... the Woodruff method calculates a confidence interval for the sample amount below the estimated percentile. That confidence interval is then transformed to the measurement scale using the inverse of the sample quantile function".

This is basically the same approach as what has been performed in this paper. There have been descriptions of the benefits of transformation for both bootstrap and jackknife variances estimates (2,3) but this report (i) describes this good performance of the jackknife approach for estimating the median variance for the special case of unweighted (or equally weighted) samples or (ii) uses this special case as a limit for convergence of bootstrap resampling calculations.

Introduction

The issue with the jackknife estimator as a linearisation of the bootstrap method, is that it relies on the distribution of sample estimates of mean, median etc being asymptotically linear and smooth.

For estimates of means, as the sample size grows, the central limit theorem is applicable and the distribution of sample means with jackknife subsampling becomes increasingly smooth and locally linear.

For sampling estimates of median variance, the distribution of resampled (bootstrap) or subsampled (jackknife) estimates of median is generally asymmetric in the original scale (2) of observed distributions and the central limit theorem does not apply.

However, in the case of unweighted samples, it is also generally true that the percentile distribution (empirical cumulative density distribution (ecdf)) about the median value 0.5, is linear. This is shown in figure 1b&d, for a N(0,1) unweighted sample of size 50, where each unweighted observation contributes the same amount $1/n$ to the ecdf.

Note that there are some more complex versions of cdfs where the ends of sample distributions are given less weighting via continuity corrections, eg. empirical distribution function with averaging and interpolation (used by Australian Bureau of Statistics). Such modified cdfs, value the sample distribution end points more weakly as the true end points of population distributions are harder to accurately sample.

Performing the jackknife calculations for estimates of median variance for unweighted samples in the percentile scale has the advantage that the percentile distribution of unweighted samples is linear while the jackknife is a linearised bootstrap estimator. Hence, the jackknife estimator could in principle approach the performance of bootstrap calculations.

Sample variance of medians

The jackknife estimation of variance for the median, using the original measurement scale, has been shown to be inconsistent (6).

In particular, the known distribution of the sample median from a population with a density function $f(x)$ is asymptotically normal with mean m and variance (7)

$$\frac{1}{4nf(m)^2} \quad (1)$$

Derived values for the jackknife variance estimate of median, using the original measurement scale, are not consistent with the above value.

However, it has also been shown that “transformation of means” to particular reference frames can improve jackknife variance accuracy (2,3). In this paper, a suitable choice of reference frame for “transformation of median” that results in improved jackknife variance estimates is the percentile scale, where the observations are arranged and labelled in percentile order.

In these next two sub-sections, the variance and bias of unweighted sample medians using jackknife estimation in the percentile scale are calculated. Then the jackknife variance estimate is backtransformed to the original measurement scale, for the standard normal distribution and compared to the asymptotic median variance given in equation 1.

Transformation of median calculations to the percentile scale

For the percentile scale for unweighted samples, the median (and p th-quantile) are points on a linear scale. Each ordered point in this distribution has the following cdf values, (using a basic median definition)

$1/n, 2/n, 3/n, \dots, n/n$

In the drop one unit jackknife variance approach for such a distribution (ignoring continuity corrections), the $(n-1)$ subsampled cdf values assigned to each ordered point are

$$1/(n-1), 2/(n-1), 3/(n-1), \dots, (n-1)/(n-1)$$

Performing the drop one unit jackknife variance calculation about the median point in this reference frame.

If n is even,

For the lower half of the ordered units, the drop one unit jackknife estimate of the median point has the value

$$jk_lower = \left(\frac{n}{2} - 1\right) \cdot \frac{1}{n-1} \quad (2)$$

$$\approx \left(\frac{1}{2} - \frac{1}{n}\right) \cdot \left(1 + \frac{1}{n} + \frac{1}{n^2} + \dots\right) \quad (3)$$

$$\approx \frac{1}{2} - \frac{1}{2n} - \frac{1}{2n^2} + \dots \quad (4)$$

For the upper other half of the ordered units, the drop one jackknife estimate of median point has the value

$$jk_lower = \left(\frac{n}{2}\right) \cdot \frac{1}{n-1} \quad (5)$$

$$\approx \frac{n}{2} \cdot \frac{1}{n} \cdot \left(1 + \frac{1}{n} + \frac{1}{n^2} + \dots\right) \quad (6)$$

$$\approx \frac{1}{2} + \frac{1}{2n} + \frac{1}{2n^2} + \dots \quad (7)$$

The mean of the drop jackknife estimates, for n even, is

$$\frac{1}{n} \cdot \left(\frac{n}{2} \cdot (jk_lower + jk_upper)\right) = \frac{1}{2} \quad (8)$$

which is the full sample median estimate

$$\therefore bias_jk_n_even = 0 \quad (9)$$

If n is odd,

For the lower half of the ordered units plus the median point, the drop one unit jackknife estimate of the median point has the value

$$jk_lower = \left(\text{floor}\left(\frac{n}{2}\right) - 1\right) \cdot \frac{1}{n-1} \quad (10)$$

$$\approx \left(\frac{1}{2} - \frac{1}{n}\right) \cdot \left(1 + \frac{1}{n} + \frac{1}{n^2} + \dots\right) \quad (11)$$

$$\approx \frac{1}{2} - \frac{1}{2n} - \frac{1}{2n^2} + \dots \quad (12)$$

For the points ordered above the median point, the drop one jackknife estimate of median point has the value

$$jk_lower = floor(\frac{n}{2}) \cdot \frac{1}{n-1} \quad (13)$$

$$\approx \frac{1}{2} \cdot (1 + \frac{1}{n} + \frac{1}{n^2} + \dots) \quad (14)$$

$$\approx \frac{1}{2} + \frac{1}{2n} + \frac{1}{2n^2} + \dots \quad (15)$$

The mean of the drop jackknife estimates, for n odd, is

$$\frac{1}{n} \cdot ((floor(\frac{n}{2}) + 1)jk_lower + floor(\frac{n}{2})jk_upper) = \frac{1}{n} \cdot (n \cdot \frac{1}{2} - \frac{1}{2n} - \frac{1}{2n^2}) \quad (16)$$

$$\approx \frac{1}{2} - \frac{1}{2n^2} - \frac{1}{2n^3} + \dots \quad (17)$$

which differs from the full sample median estimate

$$\therefore bias_jk_n_odd \approx -\frac{n-1}{2n^2} + \dots \quad (18)$$

The next step in the jackknife estimation is to calculate the variance of the distribution of the jackknife estimates.

The variance of the jackknife estimates, for n even is

$$(n-1) \cdot \frac{1}{n} \sum (jk_est - jk_mean)^2 \quad (19)$$

using the results from eqns(2-9)

$$= \frac{n-1}{n} \cdot \frac{n}{2} \cdot ((jk_lower - \frac{1}{2})^2 + (jk_upper - \frac{1}{2})^2) \quad (20)$$

$$\approx \frac{n-1}{n} \cdot \frac{n}{2} \cdot 2 \cdot (\frac{1}{2n} + \frac{1}{2n^2} + \dots)^2 \quad (21)$$

$$\approx (n-1) \cdot \frac{1}{4} \cdot (\frac{1}{n} + \dots)^2 \quad (22)$$

$$\approx \frac{1}{4n} \cdot \frac{n-1}{n} \quad (23)$$

The variance of the jackknife estimates, for n odd is

$$(n-1) \cdot \frac{1}{n} \sum (jk_est - jk_mean)^2 \quad (24)$$

using the results from eqns(10-18)

$$= \frac{n-1}{n} \cdot ((floor(\frac{n}{2}) + 1) \cdot (jk_lower - \frac{1}{2} + \frac{1}{2n^2})^2 + (floor(\frac{n}{2}) \cdot (jk_upper - \frac{1}{2} + \frac{1}{2n^2})^2)) \quad (25)$$

$$\approx \frac{n-1}{n} \cdot (2 \cdot \text{floor}(\frac{n}{2}) + 1) \cdot (\frac{1}{2n} + \dots)^2 \quad (26)$$

$$\approx \frac{n-1}{n} \cdot n \cdot (\frac{1}{2n} + \dots)^2 \quad (27)$$

$$\approx (n-1) \cdot \frac{1}{4} \cdot (\frac{1}{n} + \dots)^2 \quad (28)$$

$$\approx \frac{1}{4n} \cdot \frac{n-1}{n} \quad (29)$$

Therefore, the first order jackknife variance estimate of the unweighted median, in the percentile scale, is of the common form (for n odd or even)

$$(n-1) \cdot \frac{1}{n} \sum (jk_est - jk_mean)^2 \approx \frac{1}{4n} \cdot \frac{n-1}{n} \quad (30)$$

Backtransformation of jackknife estimates of median variance to original measurement scale - standard normal distribution

Given calculated median estimates in the percentile scale, it is necessary to backtransform the results to the original measurement scale. In practice, this requires an inversion of the cumulative density function which can be algebraically intractable. In Table 1 and the appendix of this paper this backtransformation has been conducted numerically.

In the case of the normal distribution, the cumulative density function does have a known form (8). In particular, the cdf can be expressed as a power series

$$\Phi(x) = 0.5 + \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \left[x + \frac{x^3}{3} + \frac{x^5}{3 \cdot 5} + \dots + \frac{x^{2n+1}}{(2n+1)!!} + \dots \right] \quad (31)$$

Equating the percentile scale standard error points, median point +/- equation 30, to the CDF (equation 31), it is possible to find measurement scale solutions for the variance values given by equation 30.

$$median \pm \sqrt{jk_var_perc_scale} = CDF(x) \quad (32)$$

$$0.5 \pm \text{sqrt}(\frac{1}{4n} \cdot \frac{n-1}{n}) = 0.5 + \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} \left[x + \frac{x^3}{3} + \frac{x^5}{3 \cdot 5} + \dots + \frac{x^{2n+1}}{(2n+1)!!} + \dots \right] \quad (33)$$

In the asymptotic limit of large sample sizes

$$0.5 \pm \text{sqrt}(\frac{1}{4n}) \approx 0.5 + \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} [x + \dots] \quad (34)$$

Therefore,

$$\text{sqrt}(\frac{1}{4n}) \approx \pm \frac{1}{\sqrt{2\pi}} \cdot (1 - \frac{x^2}{2}) [x + \dots] \quad (35)$$

$$(\frac{1}{4n}) \approx \frac{1}{2\pi} \cdot x^2 \quad (36)$$

Therefore, in the original measurement scale for the standard normal distribution

$$x^2 \approx \frac{2\pi}{4n} \quad (37)$$

which can be directly compared to the general asymptotic median variance given in equation 1, which for the standard normal distribution has the value

$$N(0,1)_{asymptotic_median_var} = \frac{1}{4nf(0)^2} \quad (38)$$

$$= \frac{1}{4n(\frac{1}{\sqrt{2\pi}})^2} \quad (39)$$

$$= \frac{2\pi}{4n} \quad (40)$$

The large sample estimates, equations 37 & 40 are in good agreement indicating that to first order, jackknife estimates of median (and potentially p-quantiles) calculated in the percentile scale and backtransformed to the original measurement scale can be consistent with population values.

Numerical calculations for unweighted sample medians of several continuous distributions

In Table 1 and the appendix, jackknife variance estimates of unweighted medians in the percentile scale are compared to standard jackknife calculations and classical bootstrap median confidence intervals (based on 10000 resamples) for five examples of continuous distributions.

Two common symmetric distributions

- (i) uniform distribution unif(0,1) of sample 1000, and
- (ii) the standard normal distribution N(0,1) of sample 1000

an artificial linear smooth symmetric distribution in the measurement scale

- (iii) systematic unif(0,1) of random sample size 415 using start/skip selection

and two common skewed distributions

- (iv) a log-normal distribution exp(N(0.1,1)) of sample 1000, and
- (v) a continuous bivariate normal distribution 1/3(N(55,5.5))+2/3(N(80,6)) of sample 270. This example is a continuous distribution analogue of the old faithful geyser waiting duration data located as data(faithful) on the r library(datasets).

The 2.5th/97.5th percentiles of the jackknife variance estimates are calculated using

$$0.5 \pm 1.96/\sqrt{n}$$

under the normal approximation consistent with comparing to standard CI bootstrap calculations (shown in the R code given in the appendix) and the sample sizes 270, 415 & 1000 used in this paper.

CDF in the (original) measurement scale and the percentile scale – $N(0,1)$

red line – median level

blue lines – 2.5%/97.5% CI bounds

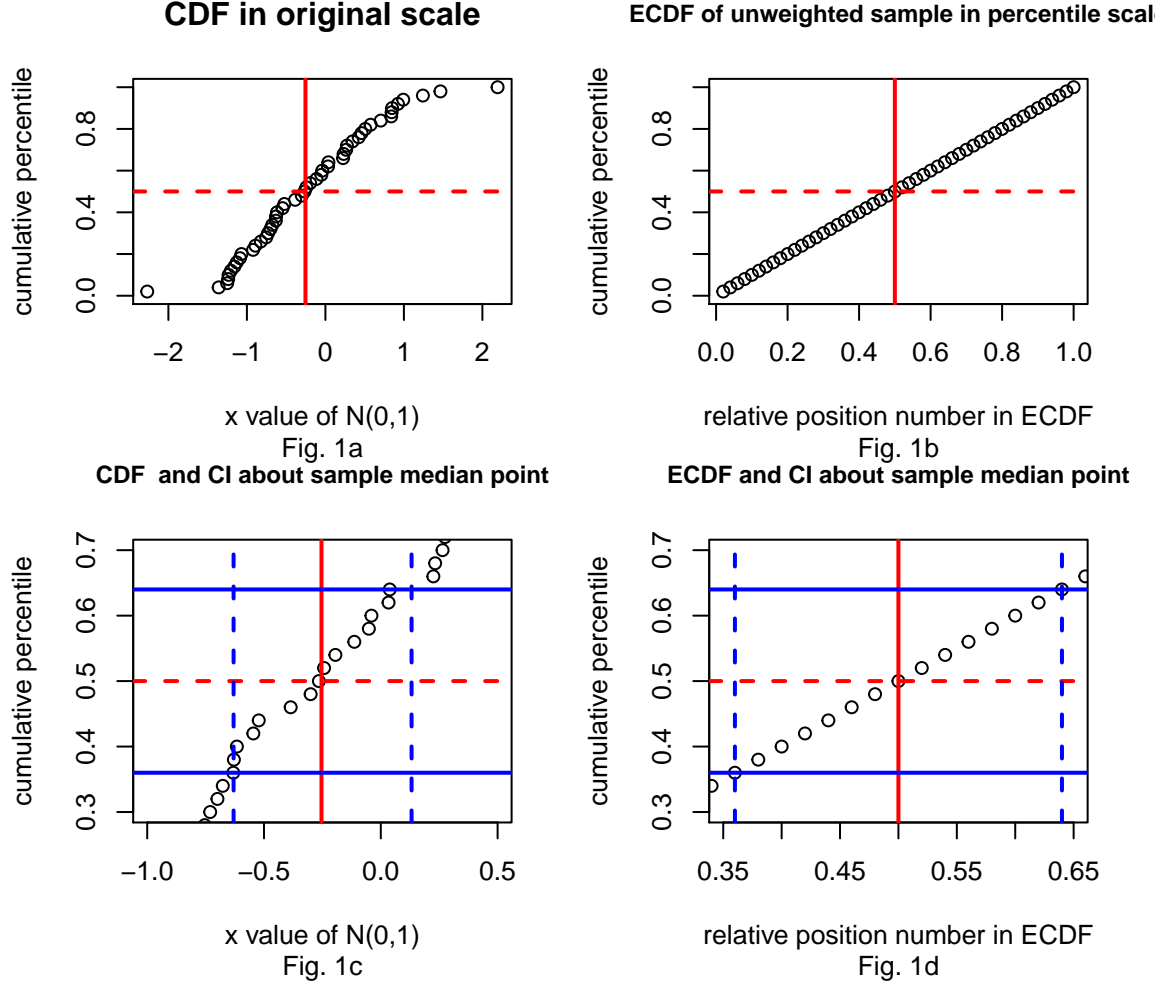


Figure 1

This range of datasets has different smoothness and nonlinearity of the cdf about the median point. As such, the confidence interval results for the regular jackknife calculation conducted in the (original) measurement scale, do not agree with the bootstrap calculation except for the artificial linear distribution example. Of course, for this artificial example (iii), the original scale cdf is linear and the jackknife calculation is expected to be accurate.

However, comparing the “jackknife variance estimates using the percentile scale” to the bootstrap calculations there is good agreement in the confidence intervals of the unweighted medians across all five examples. The transformation of the jackknife calculation to the linear ecdf scale allows the jackknife method to consider the full characteristic of the sample distribution and the backtransformation from the ecdf to the original measurement scale handles the nonlinearity in the observed data distribution.

Table 1: Estimated median 95% confidence intervals of unweighted samples for different re-sampling methods

Distribution	popn median	bootstrap	original scale jackknife	percentile scale jackknife
random uniform(0,1)	0.5	(0.4612,0.5218)	(0.4402,0.5513)	(0.4612,0.5218)

Distribution	popn median	bootstrap	original scale jackknife	percentile scale jackknife
normal $N(0,1)$	0	(-0.0993,0.0576)	(-0.08749,.02576)	(-.09972,.05772)
systematic uniform(0,1)	0.5	(0.4532,0.5494)	(0.4534,0.5492)	(0.4545,0.5505)
log-normal $\exp(N(0.1,1))$	1.09	(1.026,1.18)	(0.9434,1.236)	(1.026,1.181)
bivariate normal	76.65	(74.7,77.73)	(75.98,77.75)	(74.71,77.74)

Conclusions

The transformation of the jackknife calculation to the percentile scale for the special case of median estimates of unweighted samples, allows jackknife variance estimates to be consistent with median confidence interval estimates using first-order bootstrap calculations.

This special case provides (i) a limiting value for bootstrap calculations as the number of resamples approaches infinity, (ii) supports jackknife estimation as a linearisation of classical bootstrap estimation and (iii) supports the paradigm that SRSWR bootstrap resampling is not somehow modifying the information given by the collected distribution since the jackknife estimation is using subsampling which is more clearly thought not to modify the information given by the sample.

The issue of the accuracy of the Woodruff method for unequally weighted data (complex surveys) using replicate weighting and jackknife in the percentile scale, should be assessed by investigation of the linearity and smoothness of the cdf about the estimation point (mean, median etc).

References

1. John W. Rogers (2003), Estimating the variance of percentiles using replicate weights, 2003 Joint Statistical Meetings - Section on Survey Research Methods, p3525-3532, <http://www.amstat.org/sections/SRMS/Proceedings/y2003/Files/JSM2003-000742.pdf>
2. Efron, B. (1979) Bootstrap Methods: Another look at the Jackknife, The Annals of Statistics 7, pp1-26
3. Efron, B. (2003) Second thoughts on the bootstrap, Statistical Science 18, pp135-140
4. Miller, R.G. (1974) The Jackknife—a review, Biometrika 61, pp1-15
5. Caffo, B. (2007) Mathematical Biostatistics Boot Camp Lecture 12 on bootstrapping and resampling, <http://www.biostat.jhsph.edu/~bcaffo/651/files/lecture12.pdf>
6. Efron, B. (1982). The Jackknife, the Bootstrap and other Resampling Plans. Philadelphia: SIAM. ISBN 0898711797.
7. Rider, Paul R. (1960). “Variance of the median of small samples from several special populations”. J. Amer. Statist. Assoc. 55 (289): 148-150. [doi:10.1080/01621459.1960.10482056](https://doi.org/10.1080/01621459.1960.10482056)
8. https://en.wikipedia.org/wiki/Normal_distribution

Appendix A - Adaption of R code to jackknife calculations in percentile scale for continuous distributions

With the John Hopkins University, Data Science and Biostatistics online courses. The drop one unit jackknife variance estimation for median (5) was presented with the following R code

```
n <- length(gmVol)
theta <- median(gmVol)
```



```

jk <- sapply(1 : n,
function(i) median(gmVol[-i]) )
thetaBar <- mean(jk)
biasEst <- (n - 1) * (thetaBar - theta)
seEst <- sqrt((n - 1) * mean((jk - thetaBar)^2))

```

where gmVol is the dataset under consideration and the loop “sapply” with argument median(gmVol[-i]) drops one unit at a time and recalculates the median estimate as per the jackknife prescription. The distribution of the jackknife replicate estimates are then collated to estimate the mean, bias and standard deviation of the sample.

The Woodruff method for percentile variance estimation uses (i) the whole sample to estimate the percentile boundaries and then (ii) calculates the uncertainty in the amount of sample within each percentile boundary by subsampling or other methods. The above example code for jackknife variance estimation, in the (original) scale can be converted to the woodruff method with only a few modifications (for continuous distributions)

```

n <- length(gmVol)
theta <- median(gmVol)
jk <- sapply(1 : n,
function(i) {
Fn <- ecdf(gmVol[-i]); # empirical cumulative density function
Fn(theta) # calculates the percent of sample < percentile boundary theta for each jackknife subsample
} )
thetaBarinv <- mean(jkinv) # using inv to indicate the jackknife calculations are in percentile space
jkexpinv <- (jkinv-thetaBarinv)*sqrt(n)+thetaBarinv # a scaled version of jackknife estimate close to 1 sd
jkexpandinv <- quantile(samp,jkexpinv)
biasEstinv <- (n - 1) * (thetaBarinv - theta)
seEstinv <- sqrt((n - 1) * mean((jkinv - thetaBarinv)^2))
jackknife__median <- quantile(gmVol,0.5) # the sample median back again
jackknife__median <- quantile(gmVol,0.5+1.96*seEstinv) # 2.5th percentile jackknife in original scale
jackknife__median <- quantile(gmVol,0.5+1.96*seEstinv) # 97.5th percentile jackknife in original scale

```

My appreciation to Profs Caffo/Peng/Leeto, John Hopkins University and Coursera for their open learning approach and content availability on the world wide web as it greatly facilitated this research to be developed.

Appendix B - Results and calculations of bootstrap and jackknife resampling for several continuous distributions Using continuous distribution examples, allows the empirical cumulative density function (ecdf) to be as smooth as feasible allowing for a clear demonstration of percentile based jackknife calculations.

With drop one unit jackknife median estimation it is well known that the median estimates of jackknife subsamples have only two values. This is because the full sample median point will only move one ordered place as the number of data points in the subsamples is (n-1) compared to the full sample n. Hence in the graphs of jackknife median estimates shown below, you will only see two peaks compared to the multiple discrete distribution of bootstrap median resample estimates.

Figure 2

```

## Warning: bootstrap variances needed for studentized intervals

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 10000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = boot.out)
##
## Intervals :
## Level      Normal          Basic
## 95%   ( 0.4667,  0.5277 )   ( 0.4697,  0.5303 )
##
## Level      Percentile      BCa
## 95%   ( 0.4612,  0.5218 )   ( 0.4605,  0.5214 )
## Calculations and Intervals on Original Scale

## [1] "mean of bootstrap samples"

## [1] 0.4943

## [1] "standard error of bootstrap samples"

## [1] 0.01556

```

Symmetric distribution case 1 – unif(0,1)

red line – population median, black line – sample median
blue lines – 2.5%/97.5% percentile bounds

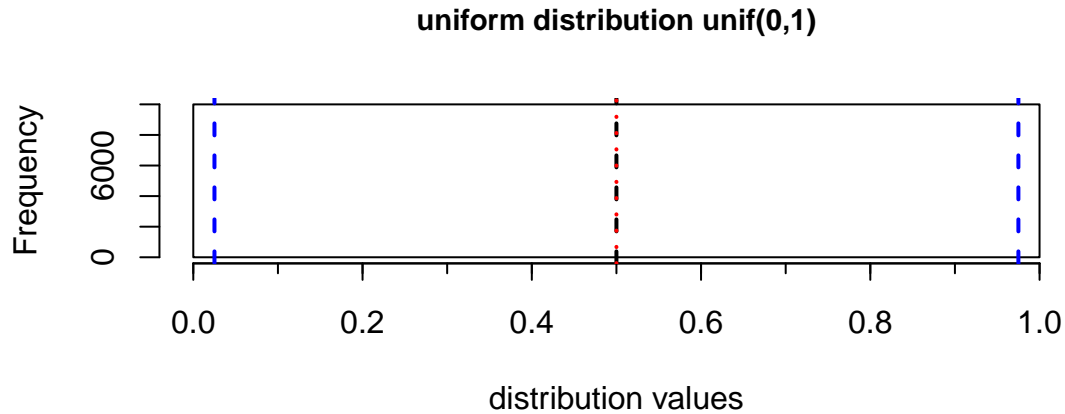


Fig. 2a

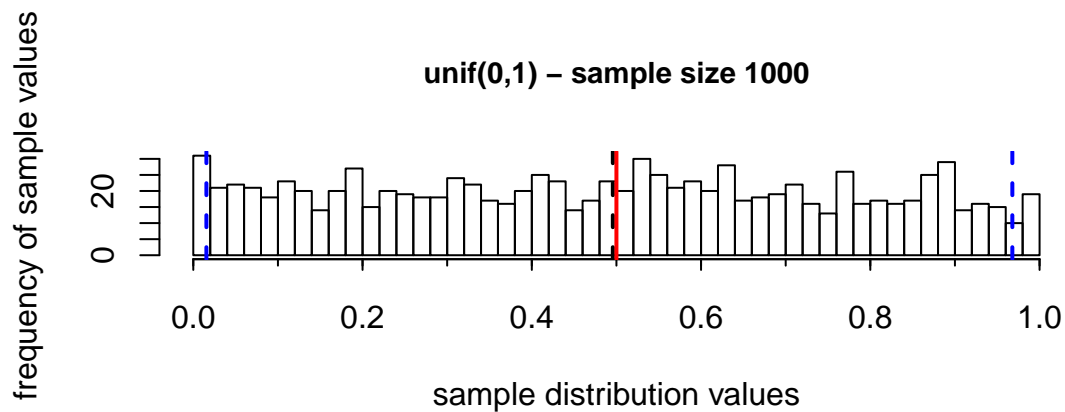


Fig. 2b

```
## [1] "jackknife median estimate using calculations in original scale"

## [1] 0.4957

## [1] "jackknife median CI lower bound using calculations in original scale"

## [1] 0.4402

## [1] "jackknife median CI upper bound using calculations in original scale"

## [1] 0.5513

## [1] ""

## [1] "jackknife median estimate using calculations in percentile scale"
```

```
## [1] "and then backtransformed to original scale"

##      50%
## 0.4957

## [1] "jackknife median estimate lower bound using calculations in "

## [1] "percentile scale and then backtransformed to original scale"

##      46.9%
## 0.4612

## [1] "jackknife median estimate upper bound using calculations in "

## [1] "percentile scale and then backtransformed to original scale"

##      53.1%
## 0.5218
```

Median variance and 95% Confidence Interval estimates
symmetric distribution case 1 – uniform – $\text{unif}(0,1)$

red – pop'n value, black – sample median, blue – 2.5%/97.5% CI bounds

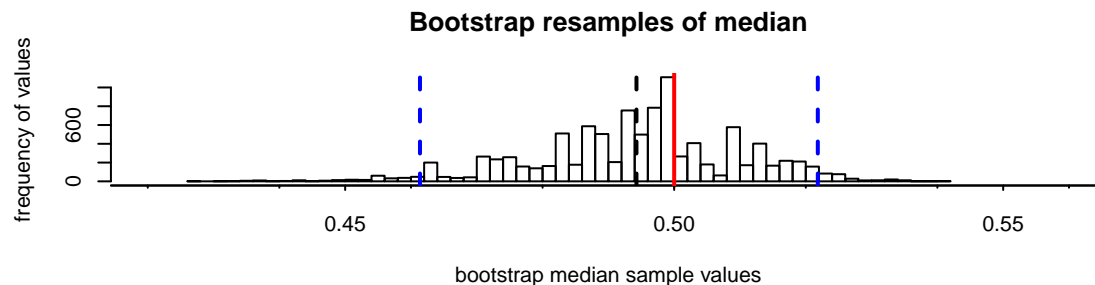


Fig. 3a

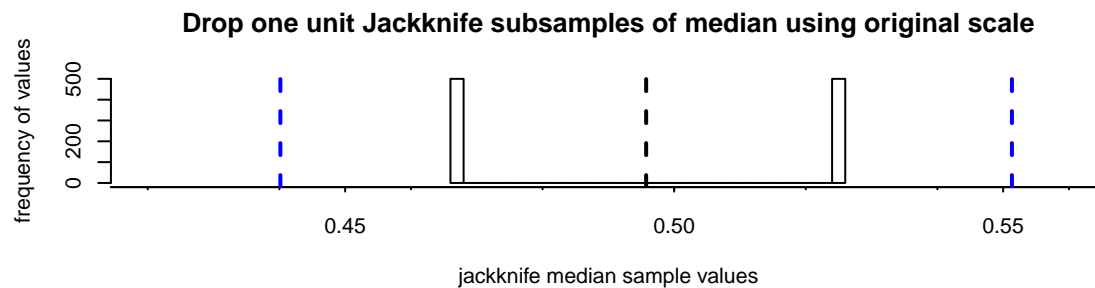


Fig. 3b

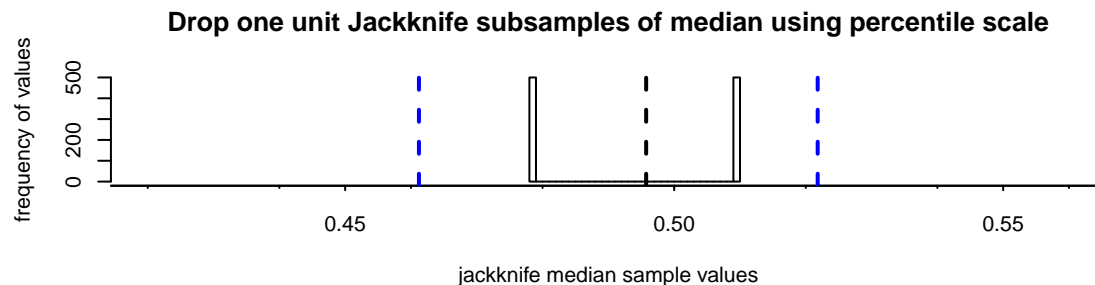
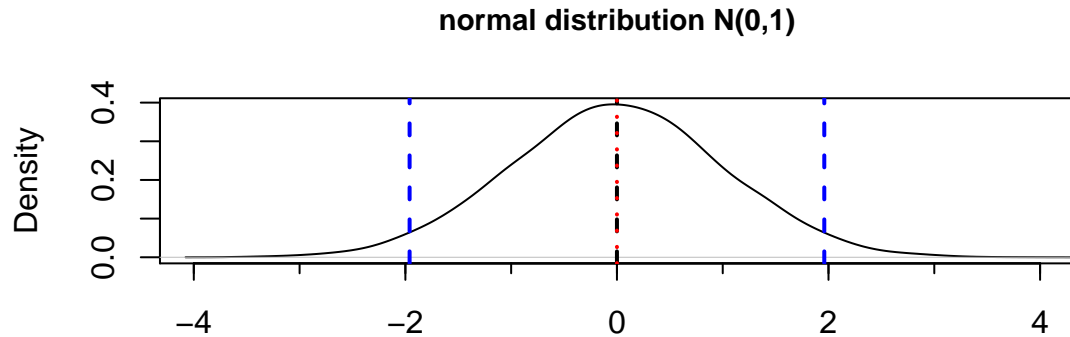


Fig. 3c

Symmetric distribution case 2 – normal – $N(0,1)$

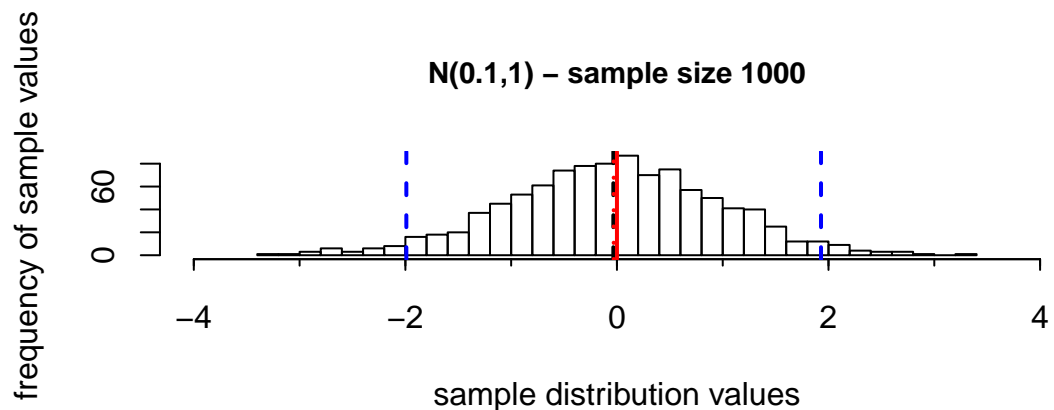
red line – population median, black line – sample median

blue lines – 2.5%/97.5% percentile bounds



distribution values

Fig. 4a



sample distribution values

Fig. 4b

Figure 4

Figure 5

```
## Warning: bootstrap variances needed for studentized intervals
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
```

```
## Based on 10000 bootstrap replicates
```

```
##
```

```
## CALL :
```

```
## boot.ci(boot.out = boot.out)
```

```
##
```

```
## Intervals :
```

```
## Level      Normal      Basic
```

```
## 95%  (-0.1192, 0.0455 )  (-0.1193, 0.0376 )
```

```
##
```

```
## Level      Percentile
```

```
## 95%  (-0.0993, 0.0576 )  (-0.1025, 0.0563 )
```

```
## Calculations and Intervals on Original Scale
```

```
## [1] -0.02484

## [1] 0.04202

## [1] "jackknife estimate based on original scale"

## [1] -0.03087

## [1] -0.08749

## [1] 0.02576

## [1] "jackknife estimates based on percentile scale"

##      50%
## -0.03087

##      46.9%
## -0.09972

##      53.1%
## 0.05772
```

Median variance and 95% Confidence Interval estimates
symmetric distribution case 2 – $N(0,1)$

red – pop'n value, black – sample median, blue – 2.5%/97.5% CI bounds

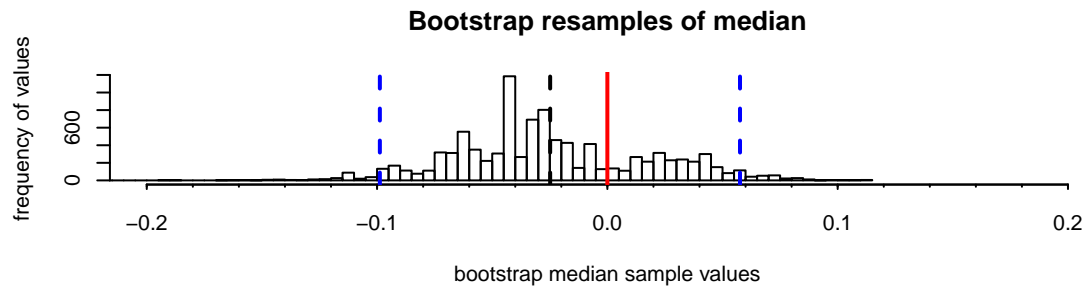


Fig. 5a

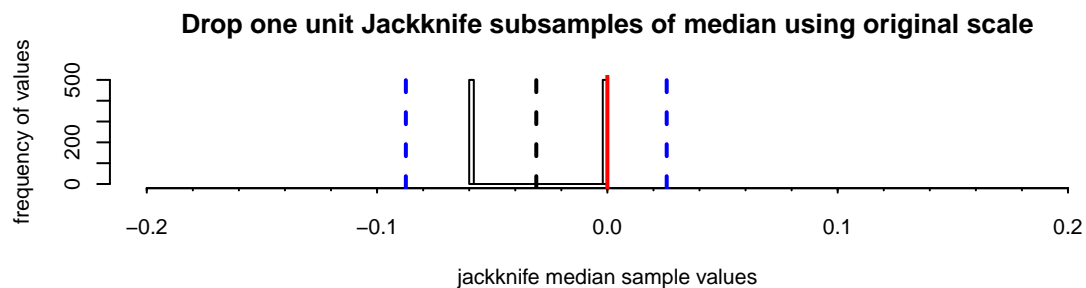


Fig. 5b

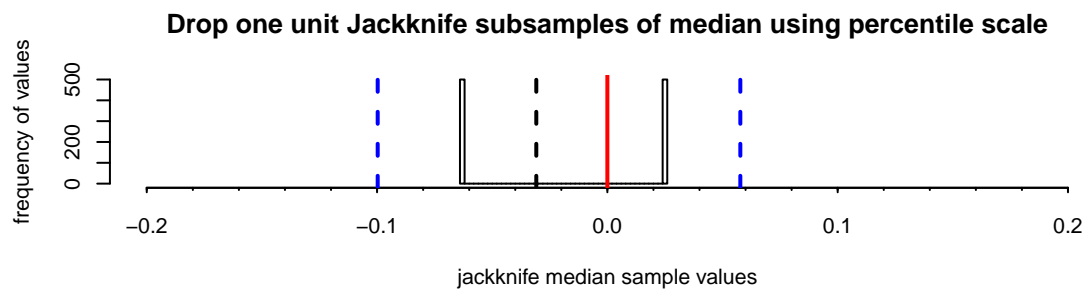


Fig. 5c

Figure 6

Warning: bootstrap variances needed for studentized intervals

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS

Based on 10000 bootstrap replicates

##

CALL :

boot.ci(boot.out = boot.out)

##

Intervals :

Level Normal Basic

95% (0.4533, 0.5497) (0.4532, 0.5494)

##

Level Percentile BCa

95% (0.4532, 0.5494) (0.4532, 0.5494)

Calculations and Intervals on Original Scale

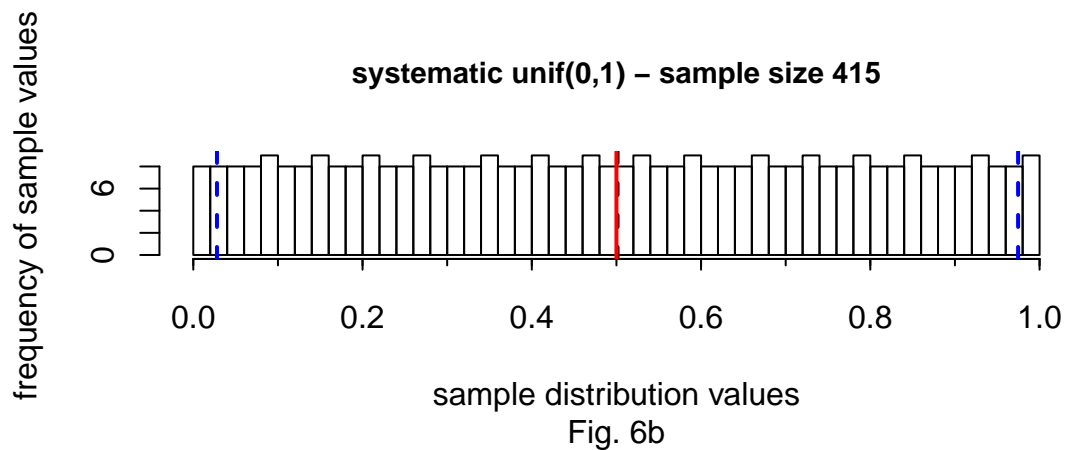
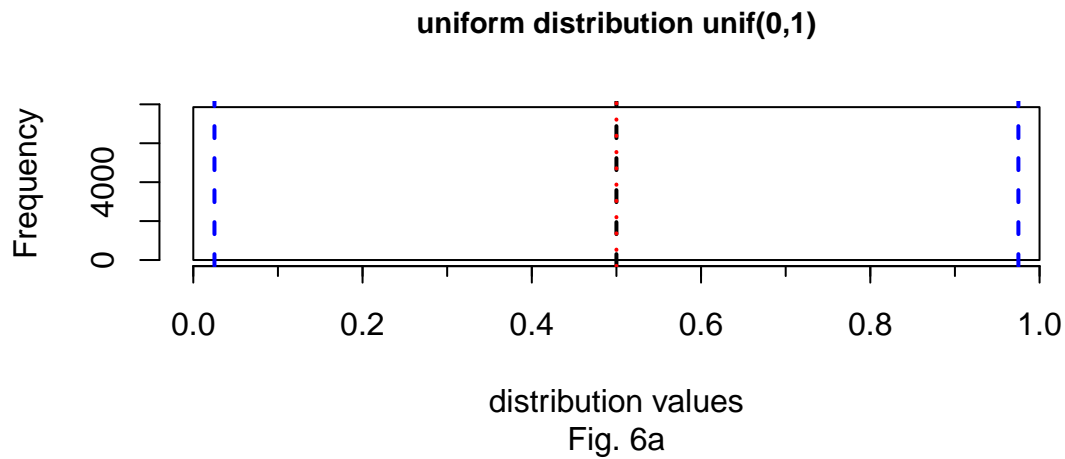
```
## [1] "mean of bootstrap samples"

## [1] 0.5011

## [1] "standard error of bootstrap samples"

## [1] 0.0246
```

Symmetric distribution case 3 – systematic unif(0,1)
red line – population median, black line – sample median
blue lines – 2.5%/97.5% percentile bounds



```
## [1] "jackknife median estimate using calculations in original scale"

## [1] 0.5013

## [1] "jackknife median CI lower bound using calculations in original scale"

## [1] 0.4534
```



```

## [1] "jackknife median CI upper bound using calculations in original scale"

## [1] 0.5492

## [1] ""

## [1] "jackknife median estimate using calculations in percentile scale "

## [1] "and then backtransformed to original scale"

## 50.12%
## 0.5025

## [1] "jackknife median estimate lower bound using calculations in "

## [1] "percentile scale and then backtransformed to original scale"

## 45.3%
## 0.4545

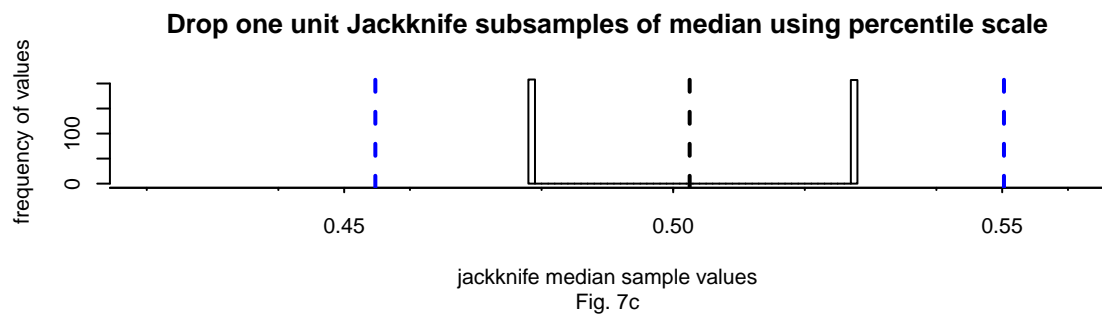
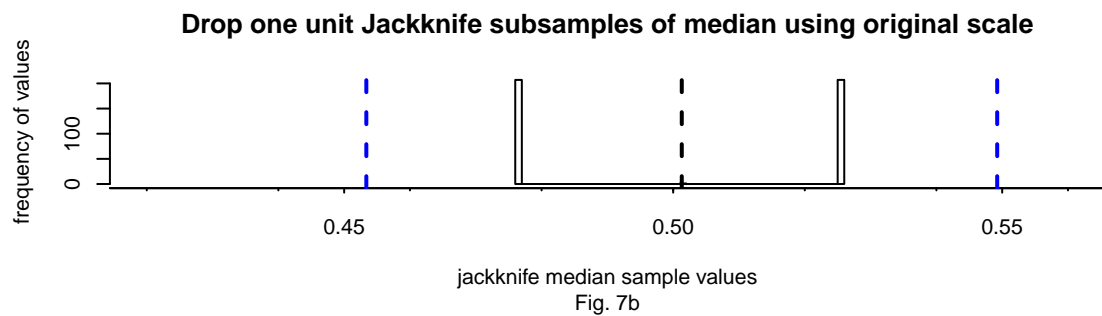
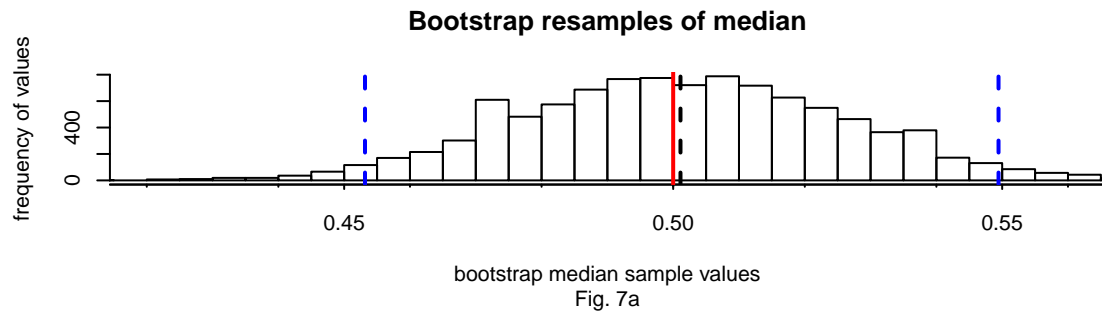
## [1] "jackknife median estimate upper bound using calculations in "

## [1] "percentile scale and then backtransformed to original scale"

## 54.94%
## 0.5505

```

Median variance and 95% Confidence Interval estimates
symmetric distribution case 3 – systematic uniform – evenly spaced points (0,1)
red – pop'n value, black – sample median, blue – 2.5%/97.5% CI bounds



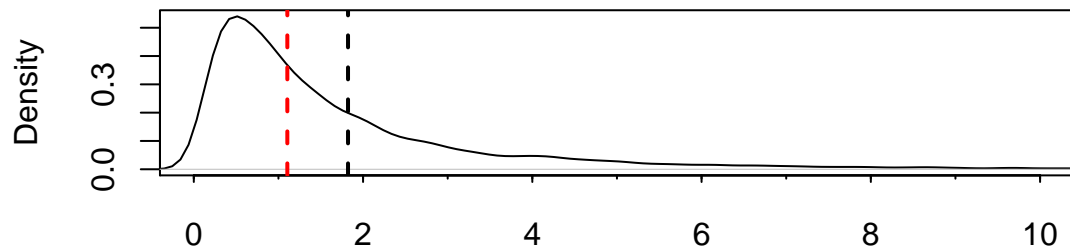
A simulated log-normal dataset is created to have a continuous right skewed distribution. This allows examination of the asymmetric nature of the median confidence interval in the original scale.

Skewed distribution case 1 – log-normal $\exp(N(0.1,1))$

red line – population/sample median

black line – population/sample mean

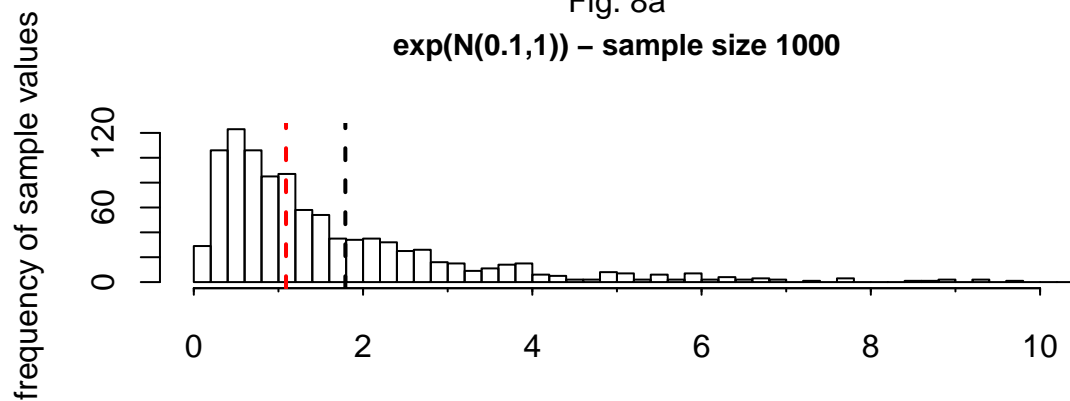
log-normal distribution $\exp(N(0.1,1))$



distribution values

Fig. 8a

$\exp(N(0.1,1))$ – sample size 1000



sample distribution values

Fig. 8b

Figure 8

Figure 9

Warning: bootstrap variances needed for studentized intervals

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS

Based on 10000 bootstrap replicates

##

CALL :

boot.ci(boot.out = boot.out)

##

Intervals :

Level Normal Basic

95% (1.011, 1.166) (0.999, 1.153)

##

Level Percentile BCa

95% (1.026, 1.180) (1.026, 1.179)

Calculations and Intervals on Original Scale

```

## [1] 1.091

## [1] 0.03959

## 2.5% 15.87% 50% 68.27% 97.5%
## 1.026 1.049 1.090 1.100 1.180

## [1] "Jackknife estimates using original scale"

## [1] 1.09

## [1] 0.9434

## [1] 1.236

## [1] "Jackknife estimates using percentile scale and backtransformed"

## 50%
## 1.09

## 46.9%
## 1.026

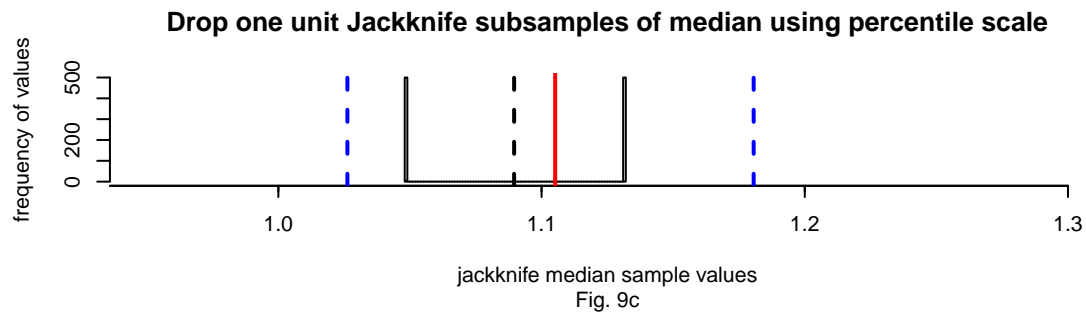
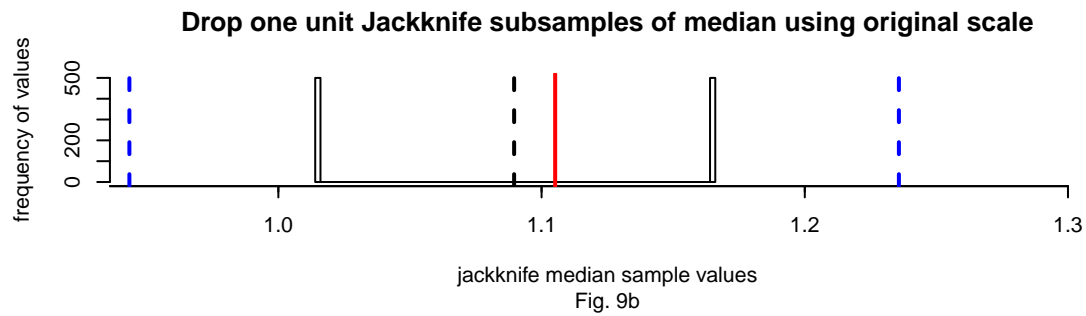
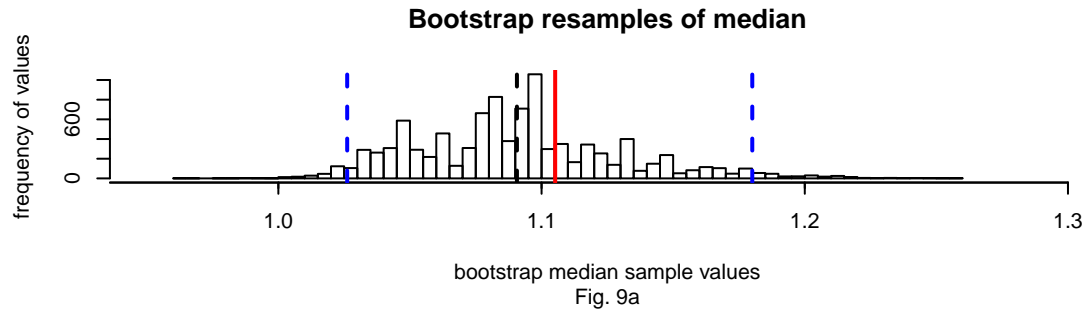
## 53.1%
## 1.181

```

Median variance and 95% Confidence Interval estimates

Skewed distribution case 1 – log-normal $\exp(N(0.1,1))$

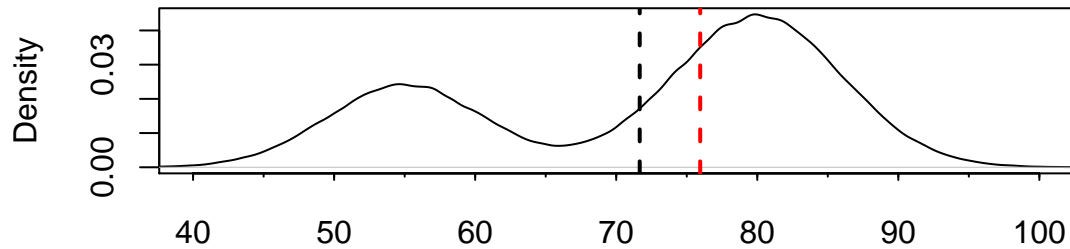
red – pop'n value, black – sample median, blue – 2.5%/97.5% CI bounds



This final example is a continuous distribution analogue of the waiting duration data for the “old faithful geyser” on `r library(datasets)` as `data(waiting)`. The sample size of the simulation (270) and the asymmetry of the bivariate distribution is closely matched to the real dataset.

Skewed distribution case 2 – weighted bivariate normal
approximation of old faithful geyser wait duration
red line – population/sample median
black line – population/sample mean

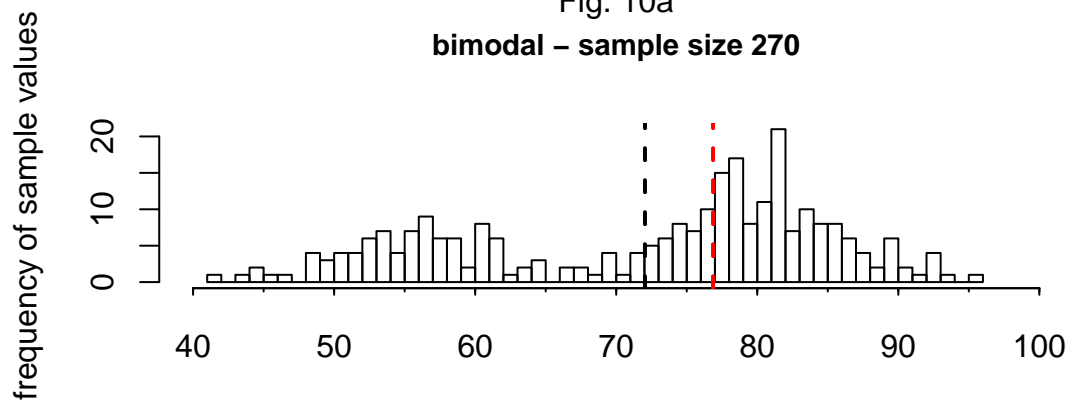
bivariate normal – $1/3 \cdot \exp(N(55, 5.5)) + 2/3 \cdot \exp(N(80, 6))$



distribution values (seconds)

Fig. 10a

bimodal – sample size 270



sample distribution values (seconds)

Fig. 10b

Figure 10

Figure 11

Warning: bootstrap variances needed for studentized intervals

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS

Based on 40000 bootstrap replicates

##

CALL :

boot.ci(boot.out = boot.out)

##

Intervals :

Level Normal Basic

95% (75.56, 78.60) (75.99, 79.02)

##

Level Percentile BCa

95% (74.70, 77.73) (74.70, 77.73)

Calculations and Intervals on Original Scale

```

## [1] 76.65

## [1] 0.7763

## 2.5% 15.87% 50% 68.27% 97.5%
## 74.70 75.91 76.86 77.11 77.73

## [1] "Jackknife estimates using the original scale"

## [1] 76.86

## [1] 75.98

## [1] 77.75

## [1] "Jackknife estimates using percentile scale and backtransformed"

## 50%
## 76.86

## 44.02%
## 74.71

## 55.98%
## 77.74

```

Median variance and 95% Confidence Interval estimates
 skewed distribution case 2 – weighted bivariate normal
 red – pop'n value, black – sample median, blue – 2.5%/97.5% CI bounds

