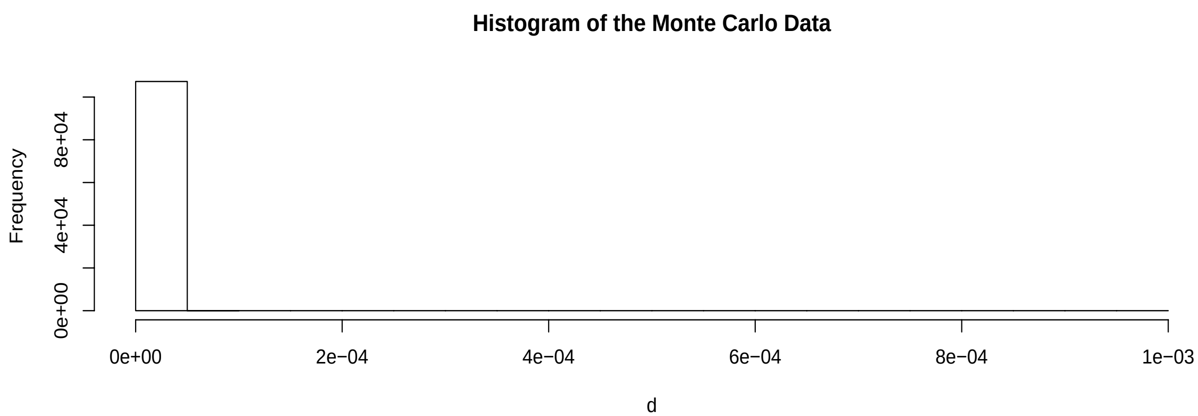
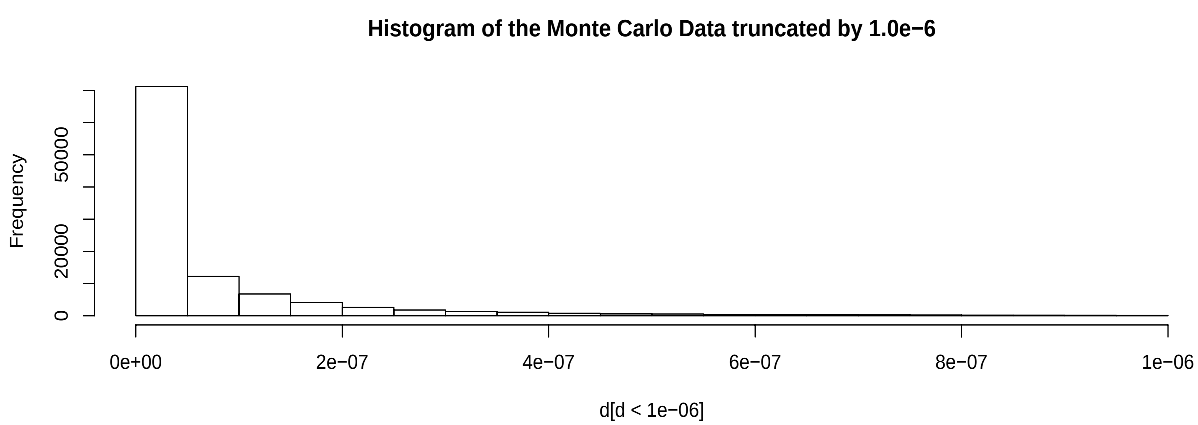
Calculating the Confidence Intervals Using Bootstrap [a]

Longhai Li
Department of Statistics
University of Toronto
October 28 2004

[a]Revised on Nov 12 2004

Looking at the Data





Estimation of the Values Interested

- Mean=1.557816e-07
- 95% Percentile: 4.510736e-07
- 99% Percentile: 1.663400e-06
- Probability that the rate exceeds 1.4e-05: 0.0007828664
- Probability that the rate exceeds 3.0e-4: 9.319838e-06

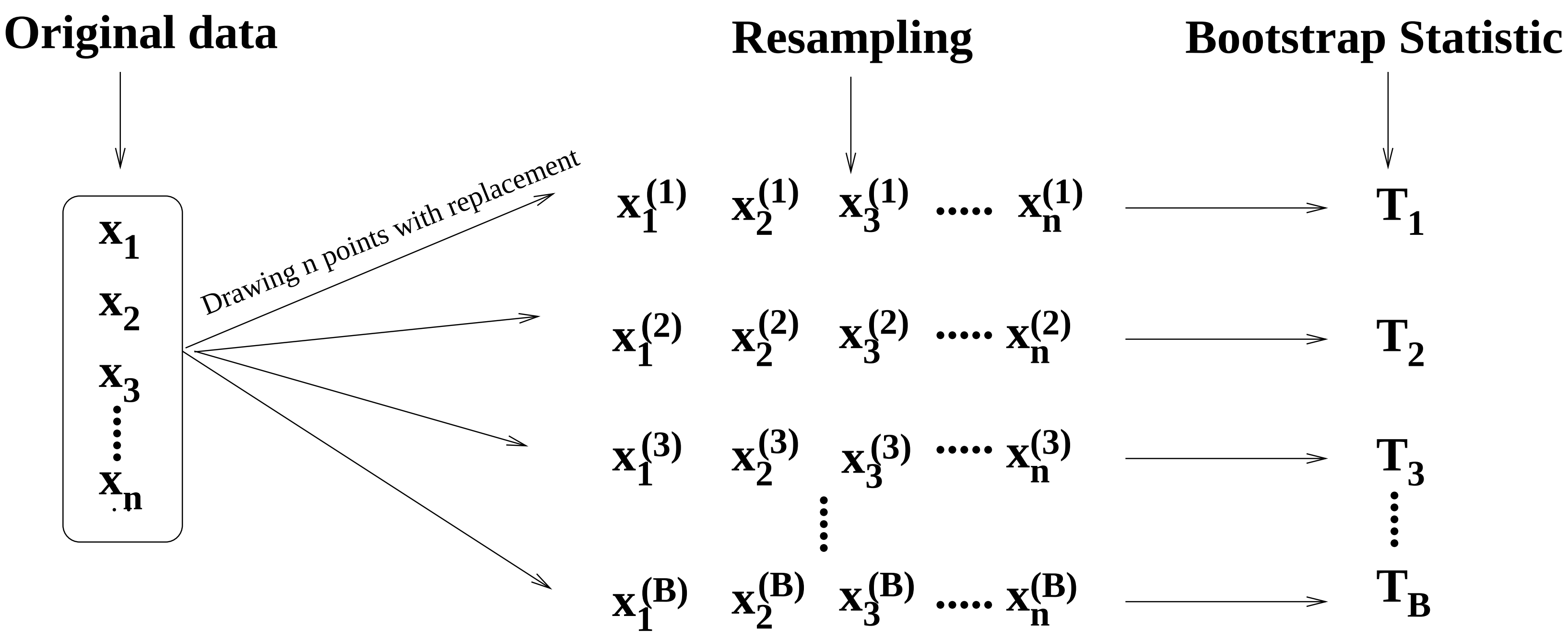
Confidence Intervals

- To construct the confidence interval of a value interested using statistic T we need the sampling distributions of T.
- Under most circumstances normal distributions are used to approximate the sampling distributions, as justified by CLT.
- For our problems, normal distributions may not good enough since they are distributed asymmetric and may have two modes(referring to the plots given later).
- Bootstrap is another class of general methods for constructing Confidence Intervals.

Introduction to Bootstrap

Bootstrap draws samples from the *Empirical Distribution* of data $\{x_1, x_2, \cdots, x_n\}$ to replicate statistic T to obtain its sampling distribution. The *Empirical Distribution* is just a *Uniform* distribution over $\{x_1, x_2, \cdots, x_n\}$. Therefore Bootstrap is just drawing i.i.d samples from $\{x_1, x_2, \cdots, x_n\}$. The procedure is illustrated by the following graph.

Graphical Illustration of Bootstrap



Bootstrap Confidence Intervals

- Simple Method
  To obtain the 95% Confidence Interval, the simple method is by taking 2.5% and 97.5% quantiles of the B replication $T_1, T_1, \cdots, T_B$ as the lower and upper bound respectively.

- More Sophisticated Method
  When the distributions are skewed we need do some adjustment. One method which is proved to be reliable is BCa method(BCa stands for *Bias-corrected and accelerated*). For the details please refer to DiCiccio, T.J. and Efron B. (1996) [2]

Bootstrap Confidence Intervals Given by BCa

When the distribution of T is skewed, we instead use the $q.low$ and $q.up$ percentiles of the bootstrap replicates of T to calculate the lower bound and upper bound of the confidence intervals. Formally, for confidence level 95%,

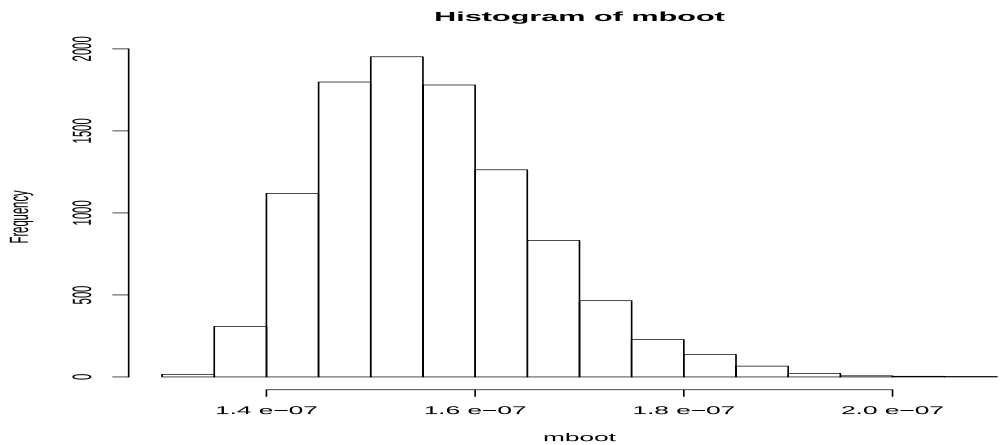
$$q.low = \Phi(z_0 + \frac{z_0 + z^{0.025}}{1 - a(z_0 + az^{0.025})}) \quad (1)$$

$$q.up = \Phi(z_0 + \frac{z_0 + z^{0.975}}{1 - a(z_0 + az^{0.975})}) \quad (2)$$

where z^α is the α quantile of standard normal distribution, z_0 and a ,namely *bias-correction* and *alleleration*, are two parameters to be estimated, by (2.8) and (6.6) in DiCiccio and Efron[2].

Confidence Interval of the *Average*

The plot of 10,000 replicates of the sample *average*:



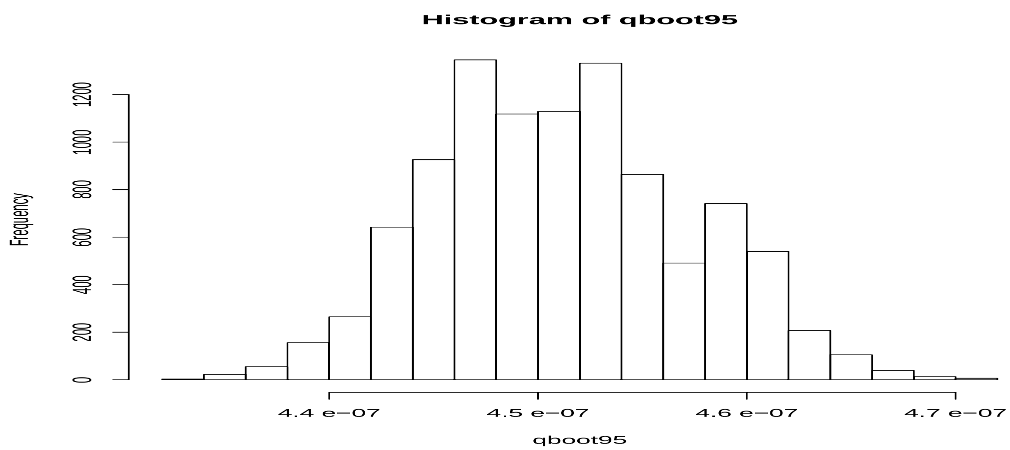
The 95% confidence interval of the *Average* calculated by BCa is:

$$\begin{array}{cc} 3.196408\% & 99.32855\% \\ 1.399666e-07 & 1.871455e-07 \end{array}$$

where 3.196408% and 99.32855% is the $q.low$ and $q.up$ given by (1) and (2), similarly for other C.I.s given henceforth.

Confidence Interval of 95% Percentile

The plot of 10,000 bootstrap replicates of the 95% sample quantile:

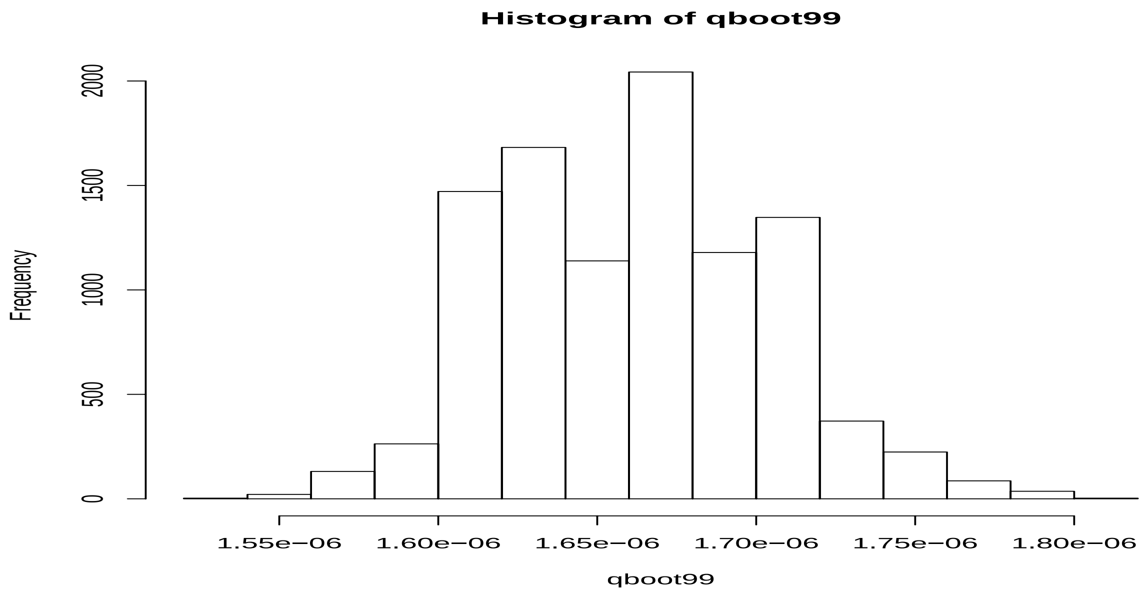


95% Confidence Interval:

$$\begin{array}{cc} 2.381881\% & 97.41008\% \\ 4.400038e-07 & 4.626755e-07 \end{array}$$

Confidence Interval of 99% Percentile

The plot of 10,000 bootstrap replicates of the 99% sample quantile:

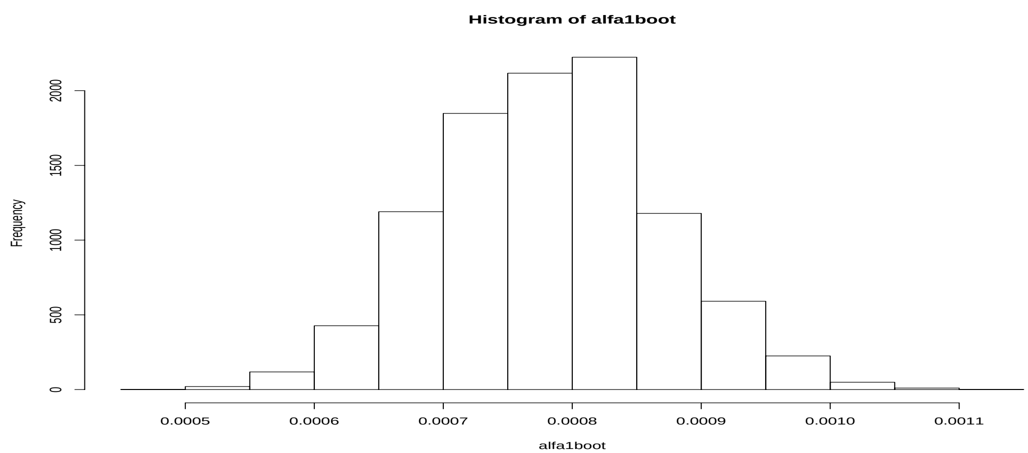


95% Confidence Interval:

$$\begin{array}{cc} 2.514489\% & 97.59096\% \\ 1.589317e-06 & 1.751419e-06 \end{array}$$

Confidence Interval of the probability that the rate exceeds 1.4E-5

The plot:

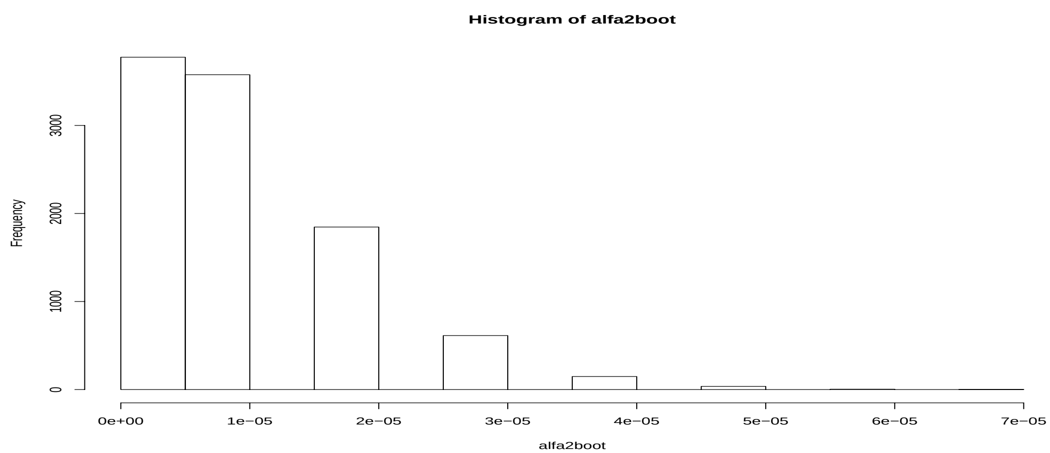


95% Confidence Interval:

1.99185%	97.19283%
0.0006151093	0.0009506235

Confidence Interval of the probability that the rate exceeds 3E-4

The plot:



95% Confidence Interval:

$$\begin{array}{cc} 0.3888835\% & 93.23166\% \\ 0.000000e+00 & 2.795951e-05 \end{array}$$

Summary of Results

Par.	Estimation	95% Confidence Interval
Average	1.557816e-07	(1.399666e-07,1.871455e-07)
95% Perc.	4.510736e-07	(4.400038e-07,4.626755e-07)
99% Perc.	1.663400e-06	(1.589317e-06,1.751419e-06)
p_1	0.0007828664	(0.000615109,0.000950623)
p_2	9.319838e-06	(0.00000e+00,2.79595e-05)

p_1—Probability that the rate exceeds 1.4e-05

p_2—Probability that the rate exceeds 3.0e-4

Reference

[1] Davison, A.C. and Hinkley, D.V. (1997) Bootstrap Methods and Their Application, Chapter 5. Cambridge University Press.

[2] DiCiccio, T.J. and Efron B. (1996) Bootstrap confidence intervals (with Discussion). Statistical Science, 11, 189-228.

[3] Efron, B. (1987) Better bootstrap confidence intervals (with Discussion). Journal of the American Statistical Association, 82, 171-200.