Math 3070 § 1. **Parametric Bootstrapping Example:** Name: Example
Treibergs **Find Standard Error and Bias of s.** June 27, 2011

This is an example of Parametric Bootstrapping to find the approximate standard error and bias of a statistic. We use the Bar Soap data encountered before. A QQ-Plot shows that the sample is reasonably normal, so we assume it is taken from the distribution $N(\mu, sigma)$. The sample mean $\bar{x}$ and standard deviation s are estimators for μ and σ. We are interested in the standard error of the sample standard deviation. We know that the sample variance s^2 is an unbiased estimator for σ^2, but because we make a nonlinear transformation, s is not necessarily an unbiased estimator for σ. In principle, we could derive the theoretical standard error knowing that it is the square root of the sample variance which satisfies a chi-squared distribution.

But, now with availability of computing power, it is easier just to simulate and measure! The idea is to take a random sample of size n from $N(\bar{x}, s)$, call it $x_1^*, \ldots, x_n^*$, take its standard deviation and call that the first bootstrapped estimate s_1^*. We repeat B times to get simulated sd's $s_1^*, s_2^*, \ldots, s_B^*$ and then the mean and standard deviations give the bootstrapped estimates

$$\overline{s^*} = \frac{1}{B}\sum_{i=1}^{B} s_i^*; \qquad S_{s^*} = \sqrt{\frac{1}{B-1}\sum_{i=1}^{B}\left(s_i^* - \overline{s^*}\right)^2}.$$

S_{s^*} is an estimate for the standard error σ_s. However, ve have to correct for bias. Denote bias of the sample standard deviation by

$$b_s = \mathrm{E}(s) - \sigma.$$

An estimate for the bias is given by

$$\widehat{b_s} = \overline{s^*} - s$$

which is the difference of the performance of the sample standard deviations against the bootstrap population standard deviation s. If $\widehat{b_s}$ is negative then the sample standard deviation is underestimating the population sd. Since we are simulating a population whose parameters are close to the background $N(\mu, \sigma)$, the bias in the bootstrap sample will be close to the background sample. Thus the bias corrected estimate of the background sd is

$$s^{**} \approx s - \widehat{b_s}.$$

The bias also influences the standard error. Indeed, the square of the standard error is

$$\begin{aligned}
\sigma_s^2 &= \mathrm{E}\left([S-\sigma]^2\right)\\
&= \mathrm{E}\left([S - \mathrm{E}(S) + \mathrm{E}(S) - \sigma]^2\right)\\
&= \mathrm{E}\left([S - \mathrm{E}(S)]^2 + 2[S - \mathrm{E}(S)][\mathrm{E}(S) - \sigma] + [\mathrm{E}(S) - \sigma]^2\right)\\
&= \mathrm{E}\left([S - \mathrm{E}(S)]^2\right) + 2\mathrm{E}\left(S - \mathrm{E}(S)\right)[\mathrm{E}(S) - \sigma] + \mathrm{E}\left([\mathrm{E}(S) - \sigma]^2\right)\\
&= \mathrm{E}\left([S - \mathrm{E}(S)]^2\right) + 0 + b_s^2\\
&\approx S_{s^*}^2 + (\overline{s^*} - s)^2.
\end{aligned}$$

Thus the bias corrected approximate standard error is

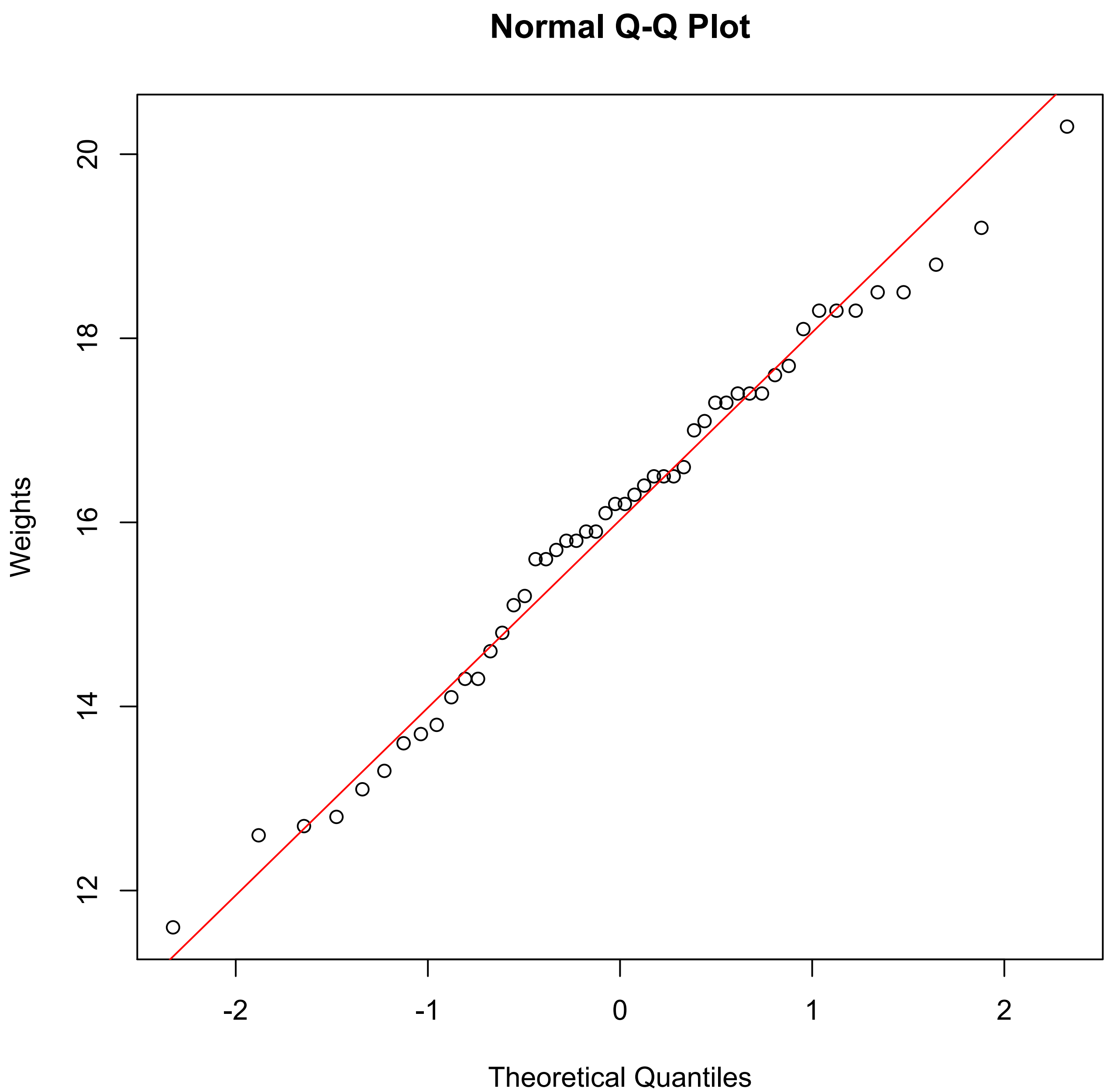
$$se^{**} = \sqrt{S_{s^*}^2 + \widehat{b_s}^{\,2}}.$$

R Session:

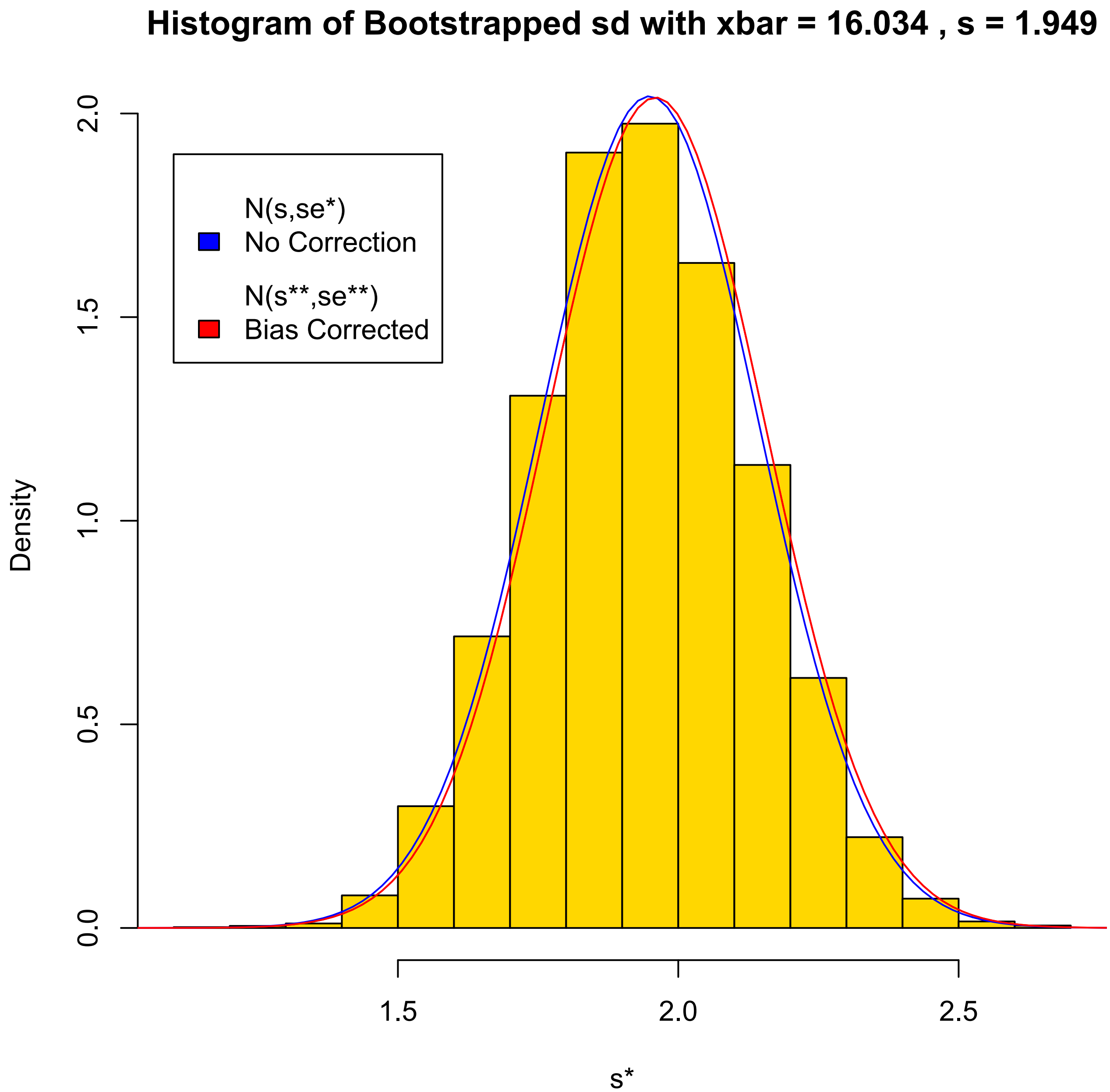
```
R version 2.11.1 (2010-05-31)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.34 (5589) i386-apple-darwin9.8.0]

[Workspace restored from /home/1004/ma/treibergs/.RData]

> tt <- read.table("M3073BarSoapData.txt",header=TRUE)
> attach (tt)
> tt
   Weights
1     11.6
2     12.6
3     12.7
4     12.8
5     13.1
6     13.3
7     13.6
8     13.7
9     13.8
10    14.1
11    14.3
12    14.3
13    14.6
14    14.8
15    15.1
16    15.2
17    15.6
18    15.6
19    15.7
20    15.8
21    15.8
22    15.9
```

23	15.9
24	16.1
25	16.2
26	16.2
27	16.3
28	16.4
29	16.5
30	16.5
31	16.5
32	16.6
33	17.0
34	17.1
35	17.3
36	17.3
37	17.4
38	17.4
39	17.4
40	17.6
41	17.7
42	18.1
43	18.3
44	18.3
45	18.3
46	18.5
47	18.5
48	18.8
49	19.2
50	20.3

```
> ################ CHECK NORMALITY OF THE BACKGROUND POPULATION #################
> qqnorm(Weights, ylab="Weights")
> qqline(Weights, col=2)
> # M3074SoapBoot1.pdf
> # Looks Pretty normal
```



```
> ############# FIND DISTRIBUTION PARAMETERS ####################################
> xbar <- mean(Weights); xbar
[1] 16.034
> s <- sd(Weights); s
[1] 1.948585
>
> ############## BOOTSTRAP SIMULATION FROM APPROXIMATE BACKGROUND POP. ##########
>
> # Sample size  n = 50,   No. Bootstrap Samples  B = 10,000
>
> sstar <- replicate(10000,sd(rnorm(50,xbar,s)))
>
> # The bootstrap sample mean gives biased samples.
> sstarbar <- mean(sstar); sstarbar
[1] 1.937850
>
> # The bootstrap sd gives the biased approximate se
> sestar <- sd(sstar); sestar
[1] 0.1953213
>
> # The approximate bias from the bootstrap sample.
> bias <- sstarbar-s; bias
[1] -0.01073481
>
> # Correct the estimated sd of the background population
> s-bias
[1] 1.95932
>
> # Correct the estimated se of the bootstrapped sample
>
> sestarstar <- sqrt(sestar^2 + bias^2); sestarstar
[1] 0.1956161
>
>
> ss <- seq(.7,3.8,1/57)
> sigmahathat <- s-bias
> lines(ss, dnorm(ss,xbar,s),col=4)
> lines(ss, dnorm(ss,s,sestar),col=4)
> lines(ss, dnorm(ss,sigmahathat,sestarstar),col=3)
>
> ######## HISTOGRAM OF BOOTSTRAPPED SAMPLE s* WITH NORMAL SUPERIMPOSED ##########
>
> hist(sstar, xlab= "s*", main = paste("Histogram of Bootstrapped sd with xbar =",
+ xb, ", s =", sb), col = "gold", freq = F)
> # Plot normals with bootstrapped mean and sd
> # Uncorrected N(s,sestar)   Corrected N(sigmahathat,sestarstar)
> lines(ss, dnorm(ss,s,sestar),col=4)
> lines(ss, dnorm(ss,sigmahathat,sestarstar),col=2)
> legend(1.1, 1.9, legend = c("N(s,se*)\nNo Correction\n",
+ "N(s**,se**)\nBias Corrected\n"), fill = c(4,2), bg = "white")
> # M3074SoapBoot2.pdf
```


Histogram of Bootstrapped sd with xbar = 16.034 , s = 1.949
N(s,se*)
No Correction
N(s**,se**)
Bias Corrected
Density
0.0
0.5
1.0
1.5
2.0
1.5
2.0
2.5
s*