University of Washington

STATISTICS

STAT 498 B

The Bootstrap

Fritz Scholz

Spring Quarter 2007

Sources & Resources

In this section I make use of some of the material from Tim Hesterberg's web site.

`http://www.insightful.com/Hesterberg/bootstrap/`

There you also find software and instructions for downloading free student versions of Splus.

As background reading I recommend Tim Hesterberg's Chapter 18 on "Bootstrap Methods and Permutation Tests" from *The Practice of Business Statistics* by Hesterberg, Monaghan, Moore, Clipson, and Epstein (2003), W.H. Freeman and Company, New York.

`http://bcs.whfreeman.com/pbs/cat_160/PBS18.pdf`

A Concrete Example

We have a random sample $\mathbf{X} = (X_1, \ldots, X_n)$ from an unknown cdf F with mean μ.

We have an estimator $\hat{\mu} = \bar{X} = \sum_{i=1}^{n} X_i/n$ for μ.

Due to variability from sample to sample there will be variability in $\bar{X}$.

With $\bar{X}$ as estimate for μ we should also quantify the uncertainty in $\bar{X}$,

e.g., its standard error $\mathrm{SE}_F(\bar{X}) = \sigma_F(\bar{X})$, and any possible bias $b_F = E_F(\bar{X}) - \mu$.

If we could produce similar such samples ad infinitum, we could obtain

the sampling distribution of $\bar{X}$, get its $\mathrm{SE}(\bar{X})$ and bias b.

Unfortunately we don't have that luxury. Enter the Bootstrap, (Efron, 1978).

The Sampling Distribution of $\bar{X}$

$$F \longrightarrow \left\{ \begin{array}{cccc} \longrightarrow & \mathbf{X}_1 & \rightarrow & \bar{X}_1 \\ \longrightarrow & \mathbf{X}_2 & \rightarrow & \bar{X}_2 \\ \longrightarrow & \mathbf{X}_3 & \rightarrow & \bar{X}_3 \\ \vdots & \vdots & \vdots & \vdots \\ \longrightarrow & \mathbf{X}_B & \rightarrow & \bar{X}_B \end{array} \right\} \rightarrow$$

For $B = \infty$

(or B very large) we get the

$(\approx)$ sampling distribution of $\bar{X}$

$\mathcal{D}(\bar{X})$

Here F denotes the sampled distribution with $\theta(F) = \mu_F$ as parameter of interest.

$\mathbf{X_i}$ is the i^{th} sample of size n from F.

$\bar{X}_i$ is the estimator $\hat{\theta} = \bar{X}$ computed from the sample $\mathbf{X_i}$.

The Bootstrap Distribution of $\bar{X}$

Use the bootstrap distribution of $\bar{X}$ as proxy/estimate for the sampling distribution.

A bootstrap sample $\mathbf{X}^\star = (X_1^\star, \ldots, X_n^\star)$ is obtained by sampling the original sample $\mathbf{X} = (X_1, \ldots, X_n)$ with replacement n times.

Same as getting a random sample $\mathbf{X}^\star$ of size n from the empirical cdf $\hat{F}_n$ of $\mathbf{X}$.

Calculate the bootstrap sample mean $\bar{X}^\star$ for this bootstrap sample $\mathbf{X}^\star$, and repeat this many times, say $B = 1000$ or 10000 times, getting $\bar{X}_1^\star, \ldots, \bar{X}_B^\star$.

If we did this $B = \infty$ times, we would get the full bootstrap distribution of $\bar{X}^\star$, as generated from $\mathbf{X}$ or $\hat{F}_n$. As it is, for $B = 1000$, we get a good estimate of it, calling it still the bootstrap distribution of the sample mean.

The Bootstrap Sampling Distribution of $\bar{X}^{\star}$

$$\hat{F}_n \longrightarrow \left\{ \begin{array}{cccc} \longrightarrow & \mathbf{X}_1^{\star} & \rightarrow & \bar{X}_1^{\star} \\ \longrightarrow & \mathbf{X}_2^{\star} & \rightarrow & \bar{X}_2^{\star} \\ \longrightarrow & \mathbf{X}_3^{\star} & \rightarrow & \bar{X}_3^{\star} \\ \vdots & \vdots & \vdots & \vdots \\ \longrightarrow & \mathbf{X}_B^{\star} & \rightarrow & \bar{X}_B^{\star} \end{array} \right\} \rightarrow$$

For $B = \infty$

(or B very large) we get the

$(\approx)$ bootstrap sampling distribution of $\bar{X}$

$\mathcal{D}(\bar{X}^{\star})$

$\hat{F}_n =$ the estimated distribution with corresponding parameter $\theta(\hat{F}_n) = \mu_{\hat{F}_n} = \bar{X}$.

$\mathbf{X}_\mathbf{i}^{\star}$ is the i^{th} bootstrap sample of size n from $\hat{F}_n$.

$\bar{X}_i^{\star}$ is the estimator $\hat{\theta}^{\star} = \bar{X}^{\star}$ computed from the bootstrap sample $\mathbf{X}_\mathbf{i}^{\star}$.

The Bootstrap Approximation Step

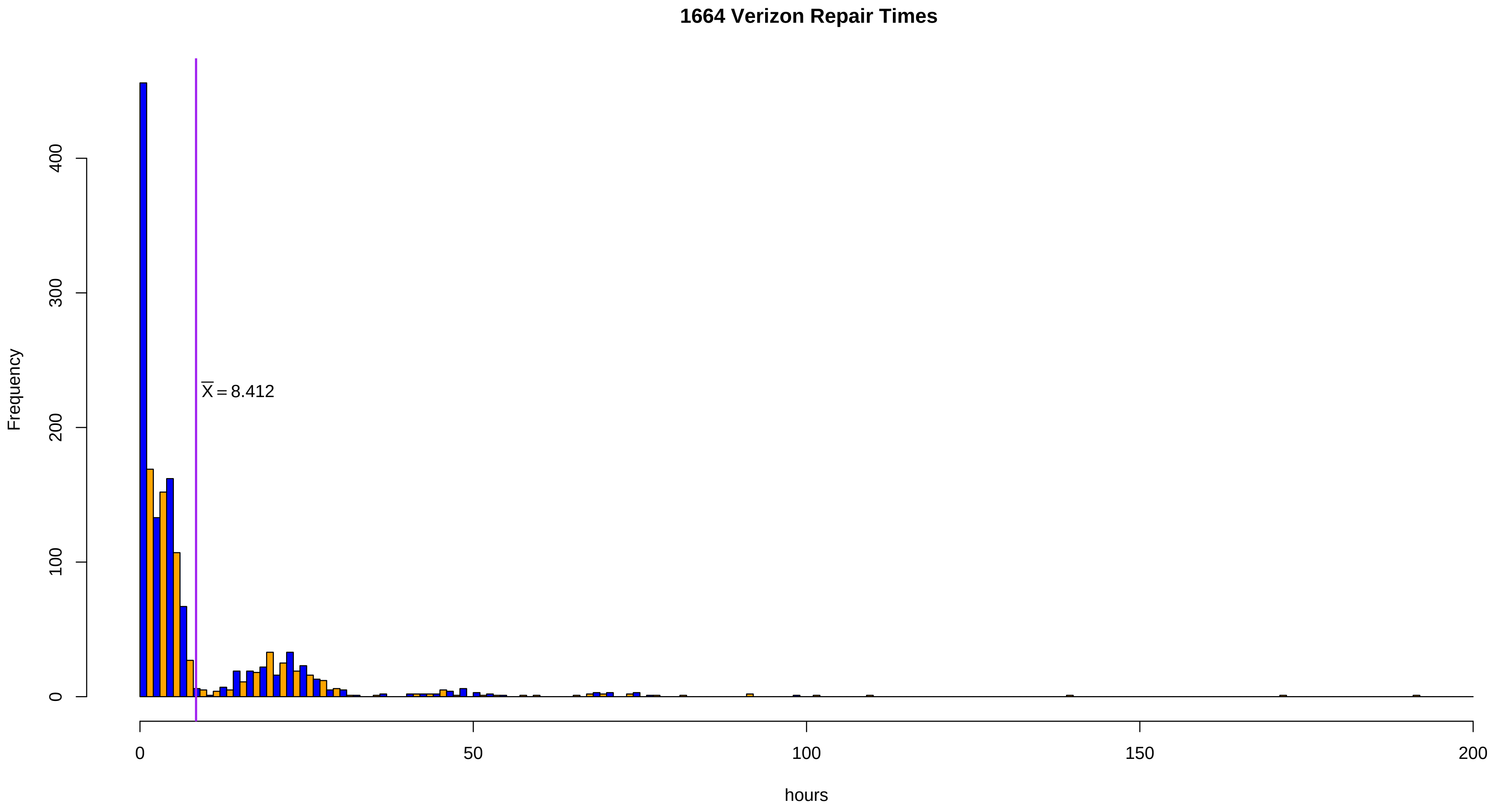
Note the complete parallelism between the sampling distribution concept and the bootstrap sampling distribution.

If the estimated distribution $\hat{F}_n$ is close to the originally sampled distribution F, we expect these two sampling distributions to be reasonably close to each other.

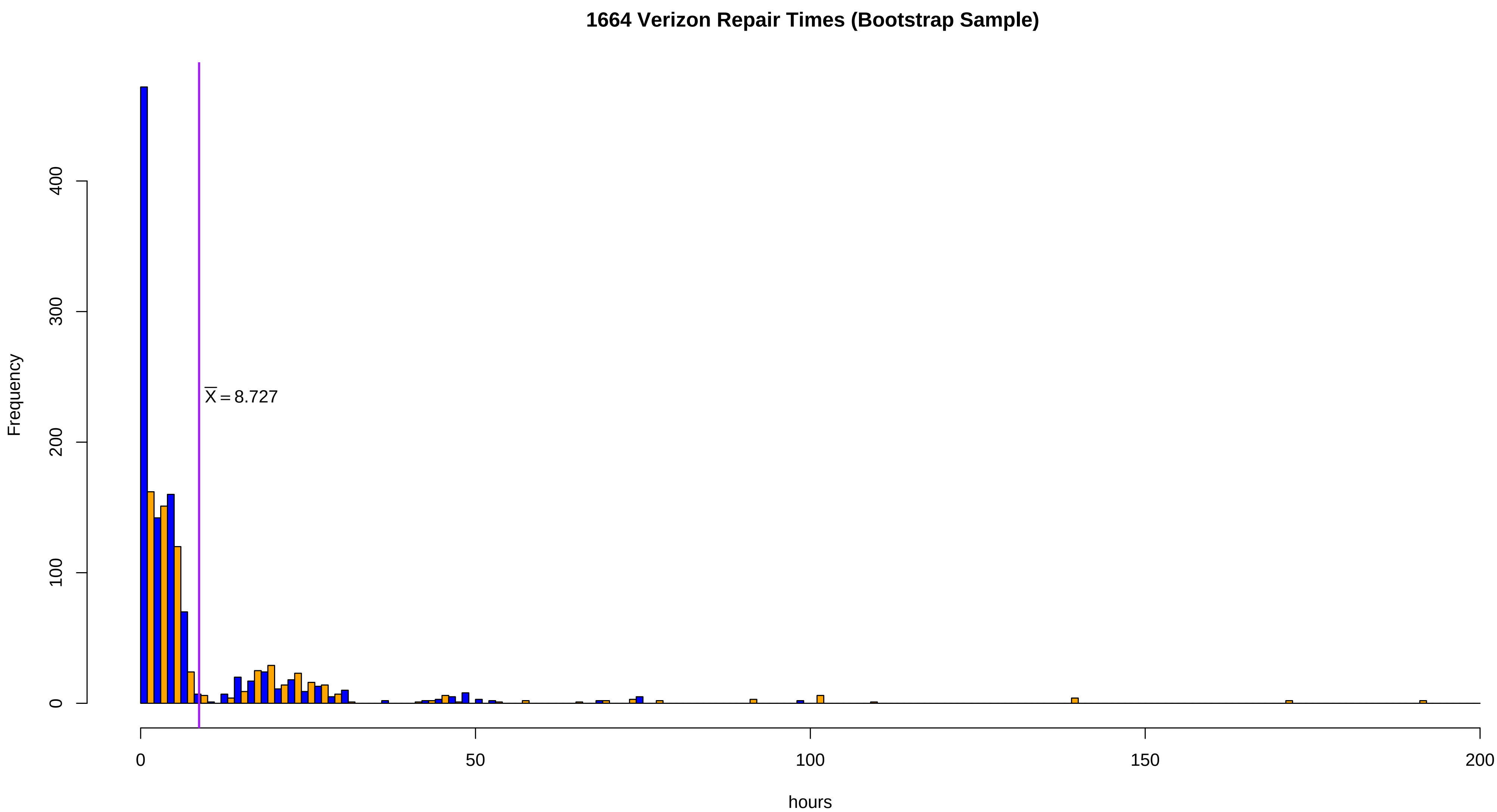
Thus take one as approximation for the other, i.e.,

$\mathcal{D}(\bar{X}^\star)$ (known) $\quad\approx\quad$ $\mathcal{D}(\bar{X})$ (unknown).

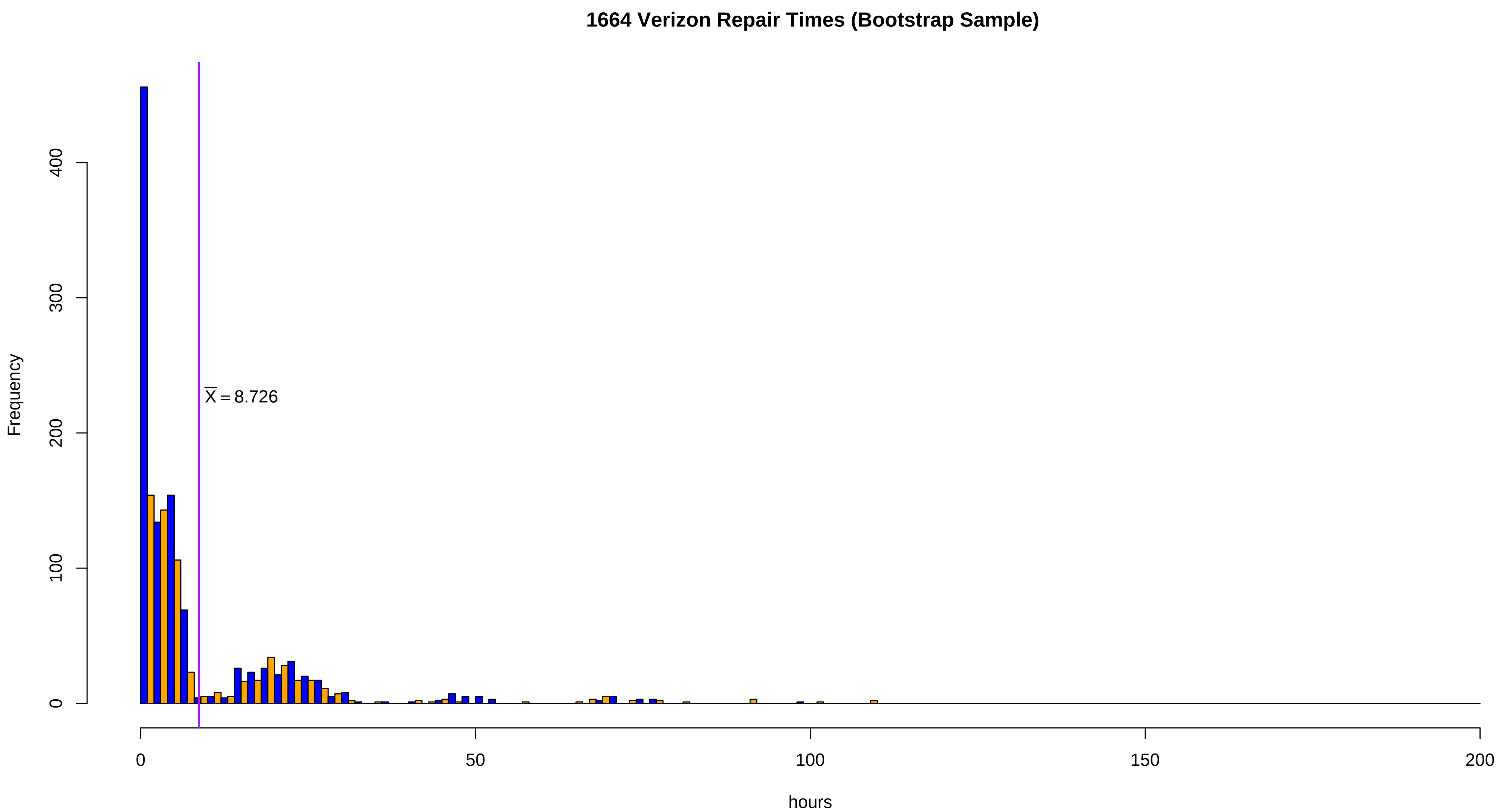
Verizon Repair Times (Not Normal!)



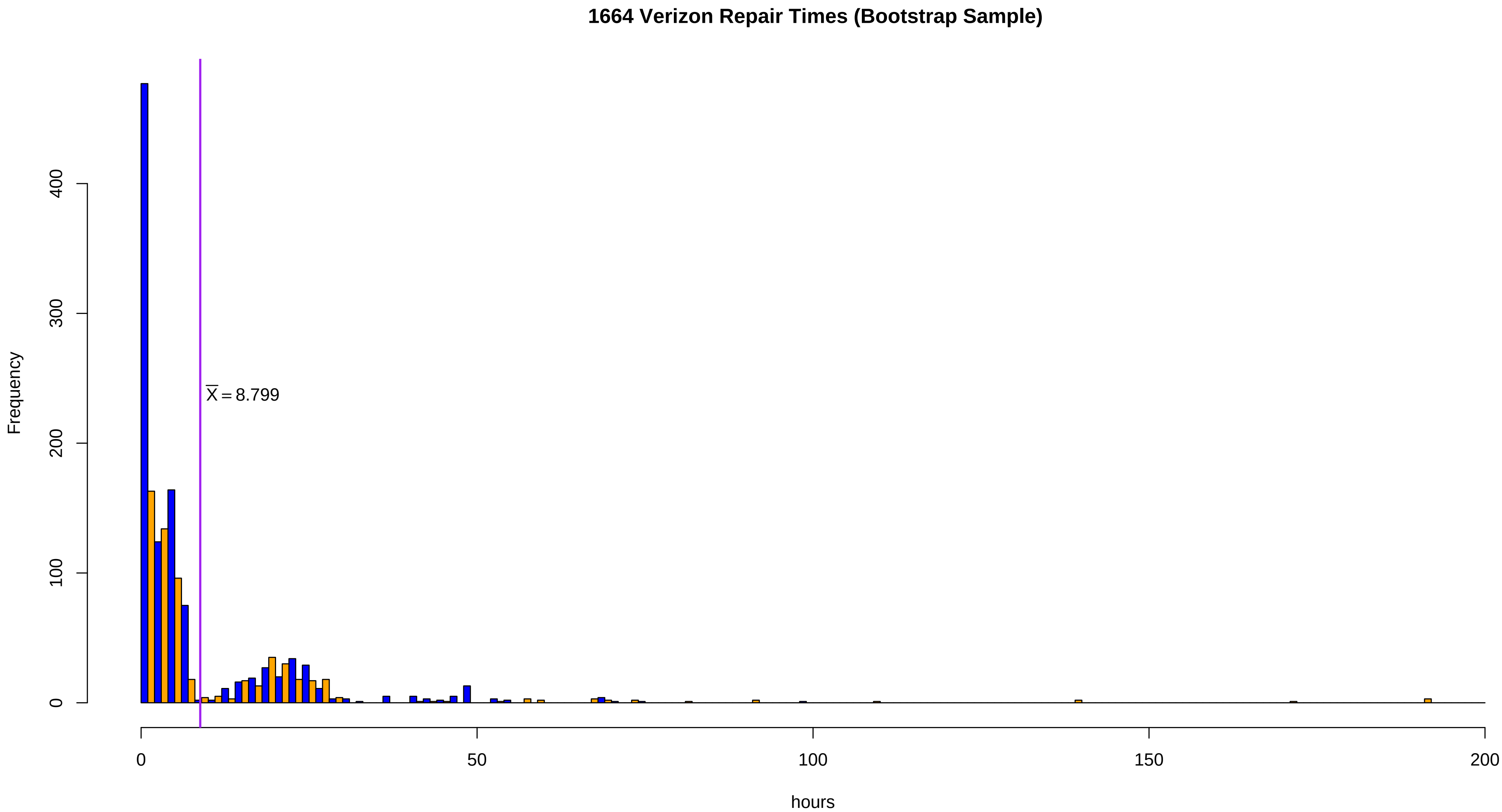
A Bootstrap Sample of Verizon Repair Times



A Bootstrap Sample of Verizon Repair Times



A Bootstrap Sample of Verizon Repair Times



What Do the Last 3 Bootstrap Samples Suggest?

The last 3 bootstrap samples show histograms very similar in character to the originally sampled histogram of Verizon repair times.

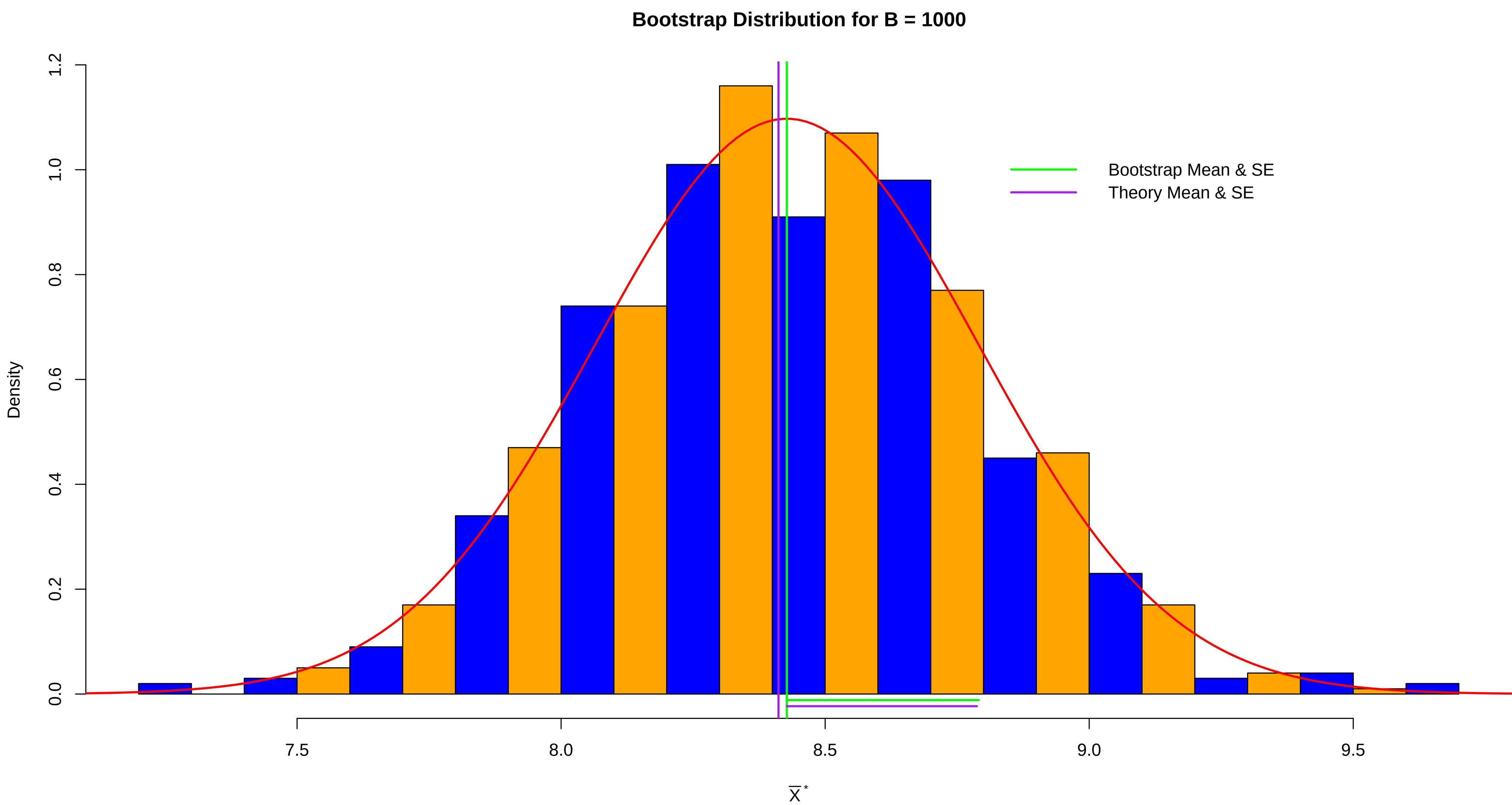
The bootstrap sample histograms don't stray far afield, at least not for large n $(n = 1664)$. $\hat{F}_n^\star \approx \hat{F}_n$.

Similarly, histograms for original samples should not stray far afield either, at least not for the same large n $(n = 1664)$. $\hat{F}_n \approx F$.

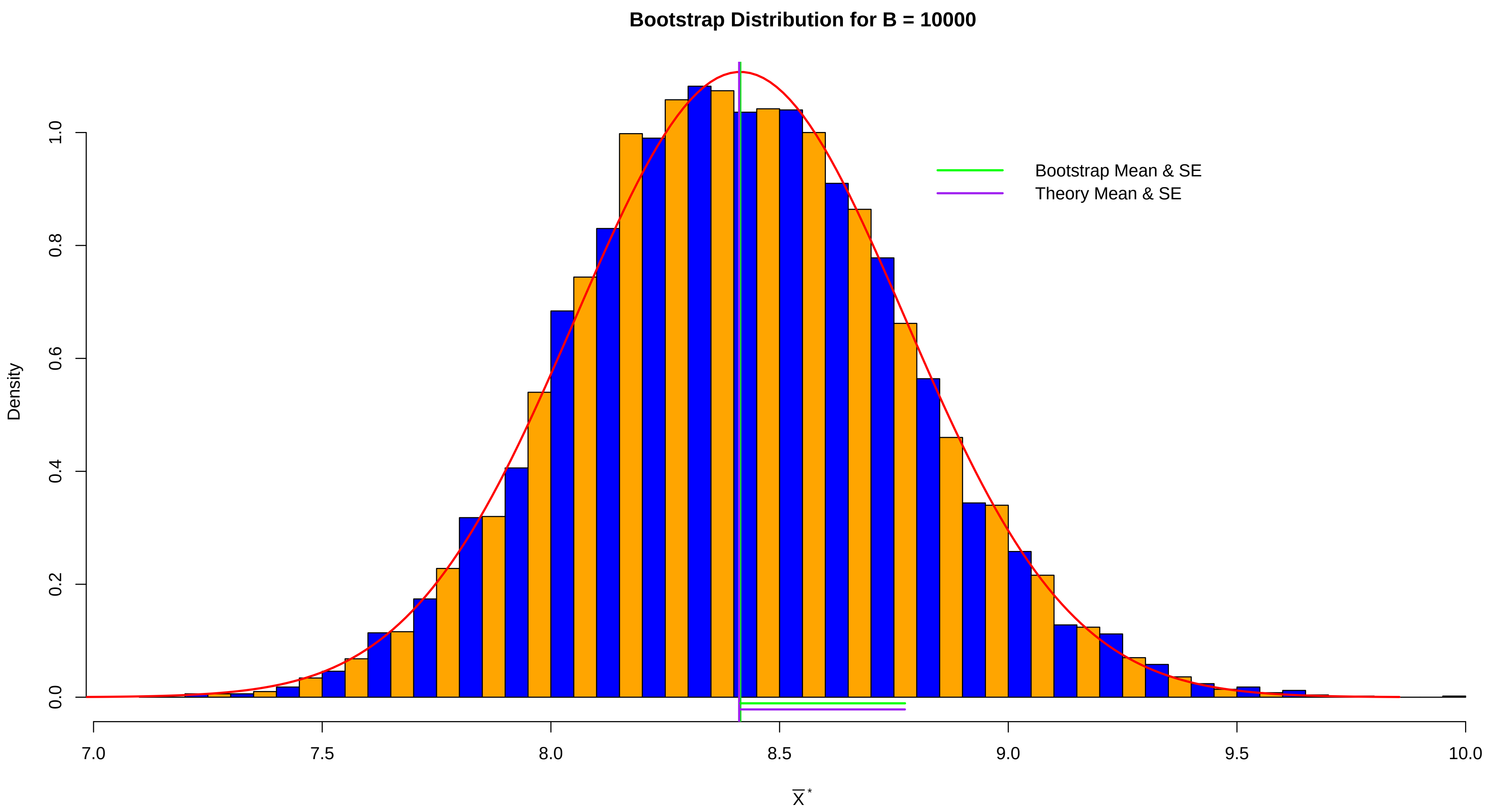
The amount of stray is mainly a function of n.

Since we use the same n in either case, the induced variation in $\bar{X}^\star$ should serve as good approximation to the induced variation in $\bar{X}$.

Bootstrap Distribution of Means ($\approx$ Normal!)



Bootstrap Distribution of Means



The R Code for Previous Slides

```
verizon.boot.mean=function (dat=verizon.dat,B=1000){
n=length(dat)
Xbar=mean(dat)
out0=hist(dat,breaks=seq(0,200,1),main=paste(n,
"Verizon Repair Times"),
xlab="hours",col=c("blue","orange"))
abline(v=Xbar,lwd=2,col="purple")
text(1.1*Xbar,.5*max(out0$counts),substitute(bar(X)==xbar,
list(xbar=format(signif(Xbar,4)))),adj=0)
readline("hit return\n")
boot.mean=NULL
for(i in 1:B){
boot.mean=c(boot.mean,mean(sample(dat,n,replace=T)))}
out=hist(boot.mean,xlab=expression(bar(X)^" *"),
probability=T,nclass=round(sqrt(B),0),col=c("blue","orange"),
main=paste("Bootstrap Distribution for B =",B))
mu.boot=mean(boot.mean)
```

The R Code for Previous Slides (cont.)

```
mu.theoryFn=Xbar
SE.bootXbar=sqrt(((B-1)/B)*var(boot.mean))
SE.theoryXbar=sqrt(((n-1)/n)*var(dat)/n)
x=seq(mu.boot-4*SE.bootXbar,mu.boot+4*SE.bootXbar,length.out=200)
y=dnorm(x,mu.boot,SE.bootXbar)
lines(x,y,lwd=2,col="red")
abline(v=mu.boot,lwd=2,col="green")
abline(v=mu.theoryFn,lwd=2,col="purple")
segments(mu.boot,-.01*max(out$density),mu.boot+SE.bootXbar,
-.01*max(out$density),col="green",lwd=2)
segments(mu.boot,-.02*max(out$density),mu.boot+SE.theoryXbar,
-.02*max(out$density),col="purple",lwd=2)
legend(mu.boot+SE.bootXbar,.9*max(out$density),
c("Bootstrap Mean & SE","Theory Mean & SE"),
col=c("green","purple"),lty=c(1,1),lwd=c(2,2),bty="n")
}
```

The Bootstrap Distribution is $\approx$ Normal

In spite of the rather non-normal distribution of repair times

the bootstrap distribution looks very normal.

This is not surprising since the sample mean is the sum of many terms,

all with equal variance

$$\bar{X} = \sum_{i=1}^{n} (X_i/n) \quad \text{and} \quad \frac{\max\{\text{var}(X_1/n), \ldots, \text{var}(X_n/n)\}}{\text{var}(X_1/n) + \ldots + \text{var}(X_n/n)} = \frac{\sigma_X^2}{n\sigma_X^2} = \frac{1}{n} = \frac{1}{1664}$$

$\Longrightarrow$ CLT $\Longrightarrow$ normal sampling distribution for $\bar{X}$.

The CLT should work equally well for the bootstrap $\bar{X}^\star$ distribution

The histograms confirm this.

Theory Mean of $\bar{X}$ and $\bar{X}^\star$

Theory $\Longrightarrow$ for a random sample $X_1,\ldots,X_n$ from some cdf F with mean μ_F

the mean or expectation of the sample mean $\bar{X}$ is μ_F,

$$E_F(\bar{X}) = E_F\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n E_F(X_i)}{n} = \frac{n\cdot\mu_F}{n} = \mu_F = \mu_F(X) = E_F(X)$$

The mean of the $\bar{X}$ sampling distribution $= X$ population mean.

We say that $\bar{X}$ is an unbiased estimator of μ_F.

Same theory says: $\bar{X}^\star$ is an unbiased estimator of the mean of $\hat{F}_n$, i.e., of $\bar{X}$

for random samples $X_1^\star,\ldots,X_n^\star$ from $\hat{F}_n$. $\quad E_{\hat{F}_n}(\bar{X}^\star) = E_{\hat{F}_n}(X^\star) = \bar{X}$

The random variable $X^\star$ takes on the values $X_1,\ldots,X_n$ with probability $1/n$ each.

Theory Variance of $\bar{X}$ and $\bar{X}^\star$

$$\sigma_F^2(\bar{X}) = \text{var}_F(\bar{X}) = \text{var}_F\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^n \text{var}_F(X_i)$$

$$= \frac{1}{n^2} \cdot n \cdot \text{var}_F(X) = \frac{\sigma_F^2(X)}{n}$$

This holds for any distribution F for X with $E(X^2) < \infty$, thus also for $\hat{F}_n$ of $X^\star$, i.e.,

$$\text{var}_{\hat{F}_n}(\bar{X}^\star) = \frac{\text{var}_{\hat{F}_n}(X^\star)}{n} = \frac{\sigma^2_{\hat{F}_n}(X^\star)}{n}$$

where

$$\text{var}_{\hat{F}_n}(X^\star) = E_{\hat{F}_n}(X^\star - E_{\hat{F}_n}(X^\star))^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{with} \quad E_{\hat{F}_n}(X^\star) = \bar{X}$$

The random variable $X^\star$ takes on the values $X_1, \ldots, X_n$ with probability $1/n$ each.

Theory SE of $\bar{X}$ and $\bar{X}^\star$

Thus the standard error of $\bar{X}$ is $\quad \mathrm{SE}(\bar{X}) = \sigma_F(\bar{X}) = \sigma_F(X)/\sqrt{n}$.

$$\mathrm{SE}(\bar{X}^\star) = \sigma_{\hat{F}_n}(\bar{X}^\star) = \frac{\sigma_{\hat{F}_n}(X^\star)}{\sqrt{n}} = \frac{1}{\sqrt{n}} \times \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2/n} = \frac{\hat{\sigma}_F}{\sqrt{n}}$$

The SE of the $\bar{X}^\star$ bootstrap distribution $=$ estimated SE of $\bar{X}$ sampling distribution.

$\widehat{=} \qquad \bar{X}^\star$ bootstrap distribution $=$ estimated $\bar{X}$ sampling distribution

We can get $\mathrm{SE}(\bar{X}^\star)$ directly from the $\bar{X}^\star$ bootstrap distribution as

$$\mathrm{SE}(\bar{X}^\star) \approxeq \mathrm{SE}_{\mathrm{boot},\bar{X}} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}\left(\bar{X}_i^\star - \bar{\bar{X}}^\star\right)^2} \qquad \text{with} \qquad \bar{\bar{X}}^\star = \frac{1}{B}\sum_{j=1}^{B} \bar{X}_j^\star$$

without knowing the standard error formula for the mean, i.e., $\mathrm{SE}(\bar{X}) = \sigma_F(X)/\sqrt{n}$.

Here $\approxeq$ becomes $=$ as $B \longrightarrow \infty$. Law of large numbers. We can force B large!

What Do Bootstrap Distribution Histograms Show?

We can check the unbiasedness property of the $\bar{X}$ estimator by comparing the mean of the bootstrap distribution for $\bar{X}^{\star}$, indicated by a green vertical line, with the theoretical mean under $\hat{F}_n$, namely $\bar{X}$, indicated by a purple vertical line.

The mean of the bootstrap distribution for $\bar{X}^{\star}$ is just the average of all B bootstrap estimates $\bar{X}_1^{\star}, \ldots, \bar{X}_B^{\star}$.

This check can only be performed while sampling from $\hat{F}_n$, but $\hat{F}_n \approx F$, and thus unbiasedness can be expected to hold for sampling from F as well.

The reason for not getting an exact match of theory and bootstrap mean in the previous histogram is that we have $B = 1000$ and not $B = \infty$.
Good approximation for $B = 1000$, even better for $B = 10000$!

What Does the Bootstrap Do for Us?

$\hat{\sigma}/\sqrt{n}$ is also called the substitution estimate of $\sigma/\sqrt{n}$, the standard error of $\bar{X}$.

This requires that we know the formula for this standard error.

As pointed out previously, $\mathrm{SE}_{\mathrm{boot},\bar{X}} \approxeq \mathrm{SE}(\bar{X}^\star) = \hat{\sigma}/\sqrt{n}$ where $\mathrm{SE}_{\mathrm{boot},\bar{X}}$ can be calculated directly from the $\bar{X}_1^\star, \ldots, \bar{X}_B^\star$ without knowing the above SE formula $\sigma/\sqrt{n}$.

Below the bootstrap distribution histograms the value for $\mathrm{SE}_{\mathrm{boot},\bar{X}}$ is indicated as a green line segment while $\mathrm{SE}(\bar{X}^\star) = \frac{\hat{\sigma}}{\sqrt{n}}$ is indicated by the purple line segment.

Agreement is quite good for large B.

What is the Big Deal?

The unbiasedness property $E(\bar{X}) = \mu_X$ and the formula $\sigma/\sqrt{n}$ for $\mathrm{SE}(\bar{X})$ are known well enough and quite ingrained.

Why go through the massive resampling and recalculation of bootstrap estimates?

When using the natural plug-in estimate $\hat{\theta} = \theta(\hat{F}_n)$ for other distribution parameters $\theta(F)$, such formulas are not so easy to come by.

The next set of histograms illustrate this with the two estimators S^2 and S, where

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Mean and SE of S^2 and S

It is well established that S^2 is unbiased, i.e., $E(S^2) = \sigma_F^2 = \sigma^2$.

However, S is biased and an explicit formula for $E_F(S)$ is not available.

We only have the following approximate formula

(from a 2-term Taylor expansion of $S = \sqrt{S^2}$ around $\sigma = \sqrt{\sigma^2}$)

$$E_F(S) \approx \sigma - \frac{1}{8}\frac{1}{\sigma^3}\,\mathrm{var}_F(S^2)$$

With some significant effort one gets

$$\mathrm{SE}_F(S^2) = \sqrt{\mathrm{var}_F(S^2)} = \sqrt{E_F(S^2 - \sigma_F^2)^2} = \sqrt{\frac{\mu_4(F) - \sigma^4}{n} + \frac{2\sigma^4}{n(n-1)}}$$

with $\mu_4(F) = E_F(X - \mu)^4$ and again by a 1-term Taylor expansion

$$\mathrm{SE}_F(S) = \sqrt{\mathrm{var}_F(S)} \approx \sqrt{\mathrm{var}_F(S^2)\,\frac{1}{4\sigma^2}} = \frac{\mathrm{SE}_F(S^2)}{2\sigma}$$

1- and 2-Term Taylor Expansions

For a smooth function f we have

$$f(x) \approx f(\mu) + (x-\mu)f'(\mu) \quad \text{and} \quad f(x) \approx f(\mu) + (x-\mu)f'(\mu) + \frac{1}{2}(x-\mu)^2 f''(\mu)$$

For $f(x) = \sqrt{x}$ we have $f'(x) = \frac{1}{2\sqrt{x}}$ and $f''(x) = -\frac{1}{4x^{3/2}}$.

$$S = \sqrt{S^2} = f(S^2) \quad \approx \quad f(\sigma^2) + (S^2-\sigma^2)f'(\sigma^2) + \frac{1}{2}(S^2-\sigma^2)^2 f''(\sigma^2)$$

$$E(S) = E\left(\sqrt{S^2}\right) = Ef(S^2) \quad \approx \quad f(\sigma^2) + 0 + \frac{1}{2}f''(\sigma^2)E(S^2-\sigma^2)^2 = \sigma - \frac{\text{var}(S^2)}{8\sigma^3}$$

$$\text{var}(S) = \text{var}\left(\sqrt{S^2}\right) = \text{var}(f(S^2)) \quad \approx \quad \left(f'(\sigma^2)\right)^2 \text{var}(S^2) = \frac{\text{var}(S^2)}{4\sigma^2}$$

Covariance Rules

$$\text{cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E(XY) - E(X)E(Y) \quad \Longrightarrow \text{cov}(X,X) = \text{var}(X)$$

For X and Y independent, i.e., $f(x,y) = f_X(x)f_Y(y)$, we have

$$\text{cov}(X,Y) = \int\int (x-\mu_X)(y-\mu_Y)f_X(x)f_Y(y)dxdy$$

$$= \int (x-\mu_X)f_X(x)dx \int (y-\mu_Y)f_Y(y)dy = 0\cdot 0 = 0$$

$$\begin{aligned}
\text{cov}\left(\sum_i X_i, \sum_j Y_j\right) &= E\left(\left(\sum_i X_i - E(\sum_i X_i)\right)\left(\sum_j Y_j - E(\sum_j Y_j)\right)\right) \\
&= E\left(\sum_i [X_i - E(X_i)] \sum_j [Y_j - E(Y_j)]\right) \\
&= \sum_i \sum_j E\left([X_i - E(X_i)][Y_j - E(Y_j)]\right) = \sum_i \sum_j \text{cov}(X_i, Y_j)
\end{aligned}$$

Alternate Sample Variance Formula

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{2n(n-1)}\sum_{i \neq j}(X_i - X_j)^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2$$

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - \bar{X} - (X_j - \bar{X}))^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\left[(X_i - \bar{X})^2 + (X_j - \bar{X})^2 - 2(X_i - \bar{X})(X_j - \bar{X})\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}\sum_{j=1}^{n}(X_j - \bar{X})^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - \bar{X})(X_j - \bar{X}) \\
&= n \cdot \sum_{i=1}^{n}(X_i - \bar{X})^2 + n \cdot \sum_{j=1}^{n}(X_j - \bar{X})^2 - 2\sum_{i=1}^{n}(X_i - \bar{X})\sum_{j=1}^{n}(X_j - \bar{X}) \\
&= 2n \cdot \sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad \text{q.e.d.}
\end{aligned}$$

$$\Longrightarrow E(S^2) = \frac{E\left(\sum_{i \neq j}(X_i - X_j)^2\right)}{2n(n-1)} = \frac{\sum_{i \neq j}E(X_i - X_j)^2}{2n(n-1)} = \frac{2n(n-1)\sigma^2}{2n(n-1)} = \sigma^2$$

Some Significant Effort

$$\text{var}(S^2) = \frac{1}{4n^2(n-1)^2}\text{var}\left(\sum_{i\neq j}(X_i - X_j)^2\right)$$

w.l.o.g. assume $E(X_i) = 0$

$$\begin{aligned}
\text{var}\left(\sum_{i\neq j}(X_i - X_j)^2\right) &= \text{cov}\left(\sum_{i\neq j}(X_i - X_j)^2, \sum_{k\neq \ell}(X_k - X_\ell)^2\right) \\
&= \sum_{i\neq j}\sum_{k\neq \ell}\text{cov}\left((X_i - X_j)^2, (X_k - X_\ell)^2\right) \\
&= n(n-1)(n-2)(n-3)\text{cov}\left((X_1 - X_2)^2, (X_3 - X_4)^2\right) \\
&\quad + 4n(n-1)(n-2)\text{cov}(\left((X_1 - X_2)^2, (X_1 - X_3)^2\right) \\
&\quad + 2n(n-1)\text{cov}(\left((X_1 - X_2)^2, (X_1 - X_2)^2\right)
\end{aligned}$$

Note that $n(n-1)(n-2)(n-3) + 4n(n-1)(n-2) + 2n(n-1) = n^2(n-1)^2$

Special Terms 1

$$\operatorname{cov}\left((X_1-X_2)^2,(X_3-X_4)^2\right)=0$$ by independence

$$\operatorname{cov}\left((X_1-X_2)^2,(X_1-X_3)^2\right)=E\left((X_1-X_2)^2(X_1-X_3)^2\right)-E(X_1-X_2)^2E(X_1-X_3)^2$$

$$E(X_1-X_2)^2=E(X_1^2+X_2^2-2X_1X_2)=E(X_1^2)+E(X_2^2)-2E(X_1X_2)=\sigma^2+\sigma^2\cdot 0\cdot 0=2\sigma^2$$

$$E\left((X_1-X_2)^2(X_1-X_3)^2\right)=E\left((X_1^2+X_2^2-2X_1X_2)(X_1^2+X_3^2-2X_1X_3)\right)$$

$$=E\left(X_1^4+X_1^2X_3^2-2X_1^3X_3+X_2^2X_1^2+X_2^2X_3^2-2X_2^2X_1X_3-2X_1^3X_2-2X_1X_2X_3^2+4X_1^2X_2X_3\right)$$

$$=\mu_4+\sigma^4+0+\sigma^4+\sigma^4-0-0-0+0=\mu_4+3\sigma^4$$

$$\Longrightarrow\quad \operatorname{cov}\left((X_1-X_2)^2,(X_1-X_3)^2\right)=\mu_4+3\sigma^4-(2\sigma^2)^2=\mu_4-\sigma^4$$

Special Terms 2

$$\mathrm{cov}\left(\left((X_1-X_2)^2,(X_1-X_2)^2\right)=E\left((X_1-X_2)^4\right)-\left(E(X_1-X_2)^2\right)^2\right.$$

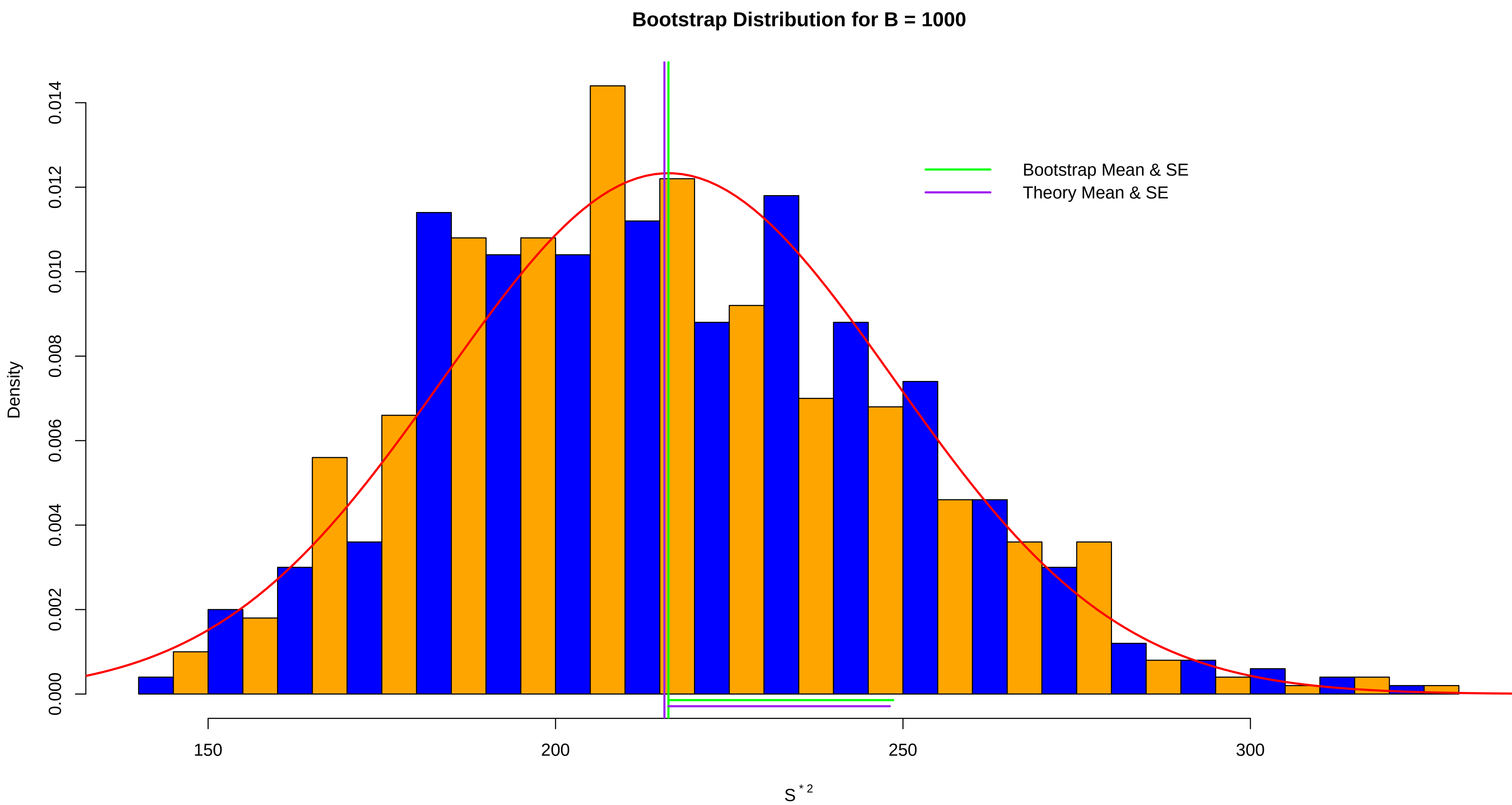
$$= E\left(X_1^4-4X_1^3X_2+6X_1^2X_2^2-4X_1X_2^3+X_2^4\right)-(2\sigma^2)^2$$

$$= \mu_4-0+6\sigma^2\sigma^2-0+\mu_4-4\sigma^4=2\mu_4+2\sigma^4$$

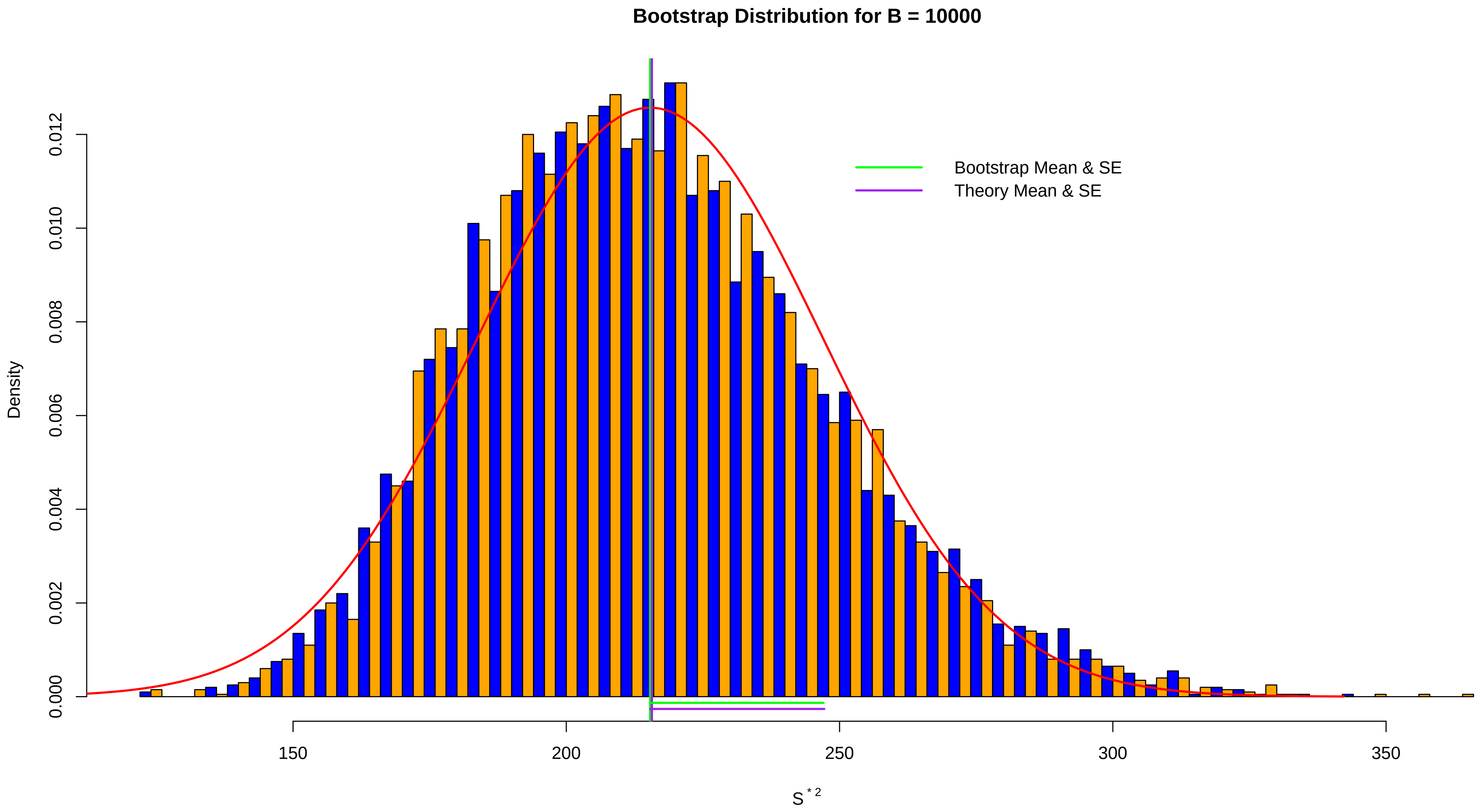
$$\begin{aligned}
\mathrm{var}\left(\sum_{i\neq j}(X_i-X_j)^2\right) &= 4n(n-1)(n-2)[\mu_4-\sigma^4]+2n(n-1)[2\mu_4+2\sigma^4] \\
&= 4n(n-1)[(n-2)(\mu_4-\sigma^4)+\mu_4+\sigma^4] \\
&= 4n(n-1)[(n-1)(\mu_4-\sigma^4)-(\mu_4-\sigma^4)+\mu_4+\sigma^4] \\
&= 4n^2(n-1)^2\left[\frac{\mu_4-\sigma^4}{n}+\frac{2\sigma^4}{n(n-1)}\right]
\end{aligned}$$

$$\Longrightarrow \qquad \mathrm{var}(S^2)=\frac{\mu_4-\sigma^4}{n}+\frac{2\sigma^4}{n(n-1)}$$

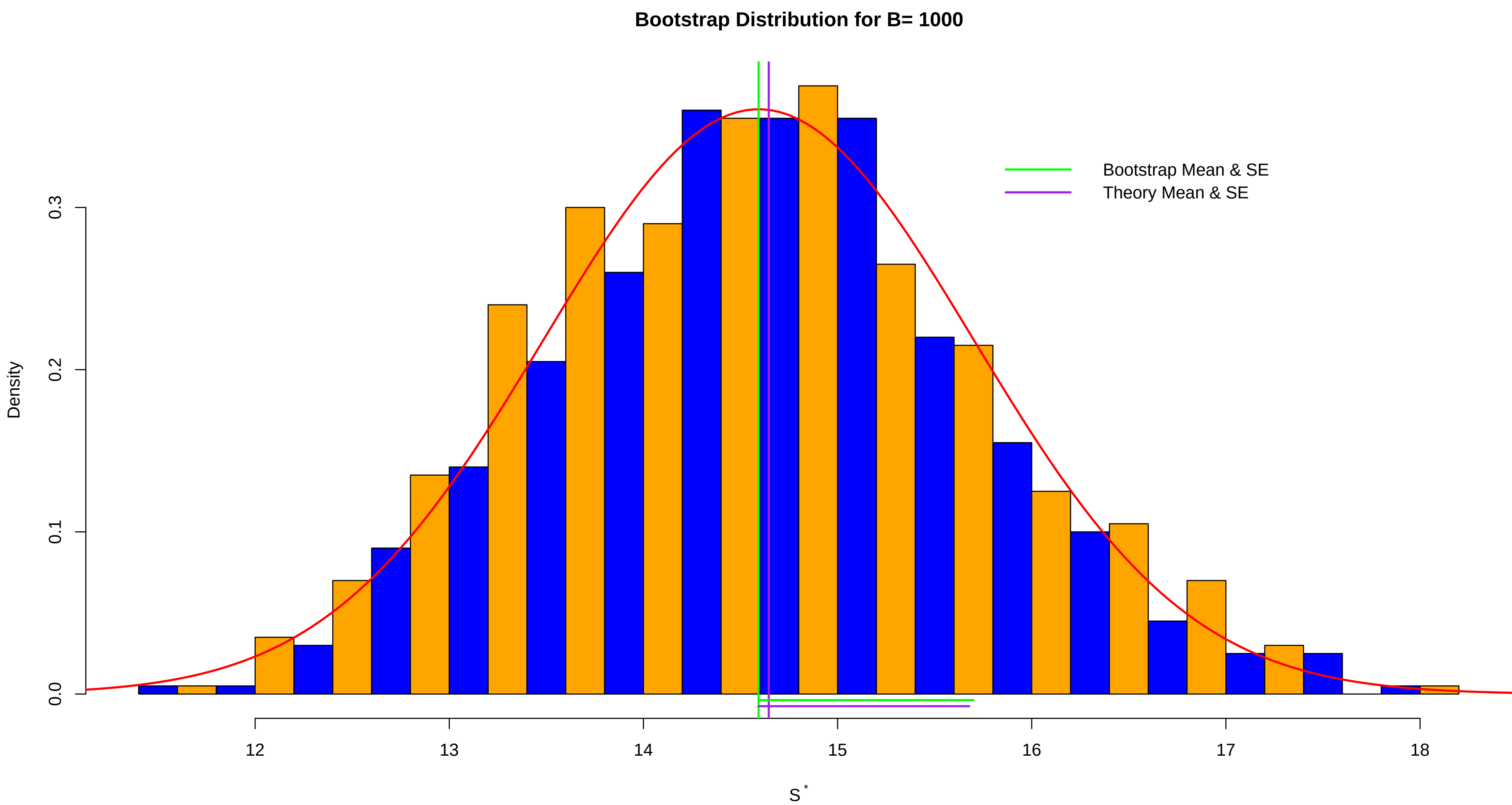
Bootstrap Distribution of $S^{\star 2}$ ($\approx$ Normal!)



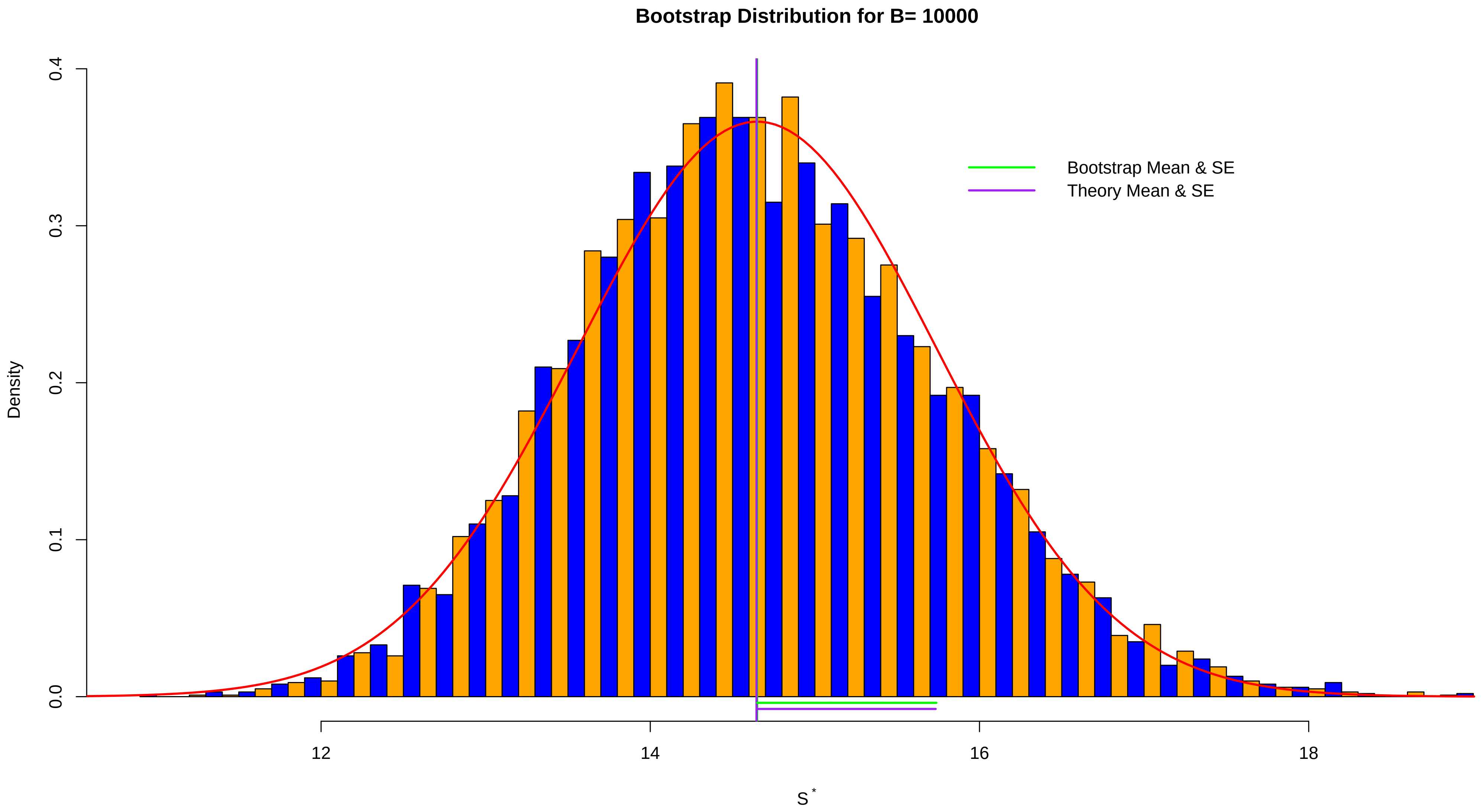
Bootstrap Distribution of $S^{\star 2}$



Bootstrap Distribution of $S^{\star}$ ($\approx$ Normal!)



Bootstrap Distribution of $S^\star$



Bootstrap Distributions $\approx$ Normal

Again we note the remarkable normality of these bootstrap distributions.

We can think of S and S^2 being influenced by all the X_i in diminishing capacity as n gets large. Note the X_i/n and X_i^2/n terms in

$$S^2 = \frac{n}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2/n = \frac{n}{n-1}\frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n} = \frac{n}{n-1}\left(\sum_{i=1}^{n} X_i^2/n - \left(\sum_{i=1}^{n} X_i/n\right)^2\right)$$

This suggests linearization, i.e., approximate S^2 and S by linear functions of the X_i and X_i^2 and invoke the CLT.

W.l.o.g. $\mu = E(X_i) = 0 \Rightarrow V = \sum_{i=1}^{n} X_i^2/n \approx \mathcal{N}(\sigma^2, (\mu_4 - \sigma^4)/n)$ & $\bar{X} \approx \mathcal{N}(0, \sigma^2/n)$.
$\bar{X}^2$ is negligible against $V \implies S^2 \approx V \approx \mathcal{N}(\sigma^2, (\mu_4 - \sigma^4)/n)$.

By a 1-term Taylor expansion of $f(S^2) = \sqrt{S^2} = S$ around σ^2

$$\implies \quad S - \sigma = \sqrt{S^2} - \sqrt{\sigma^2} \approx \frac{1}{2\sqrt{\sigma^2}}(S^2 - \sigma^2) \approx \mathcal{N}(0, (\mu_4 - \sigma^4)/n)/(4\sigma^2))$$

Not All Bootstrap Distributions are Normal

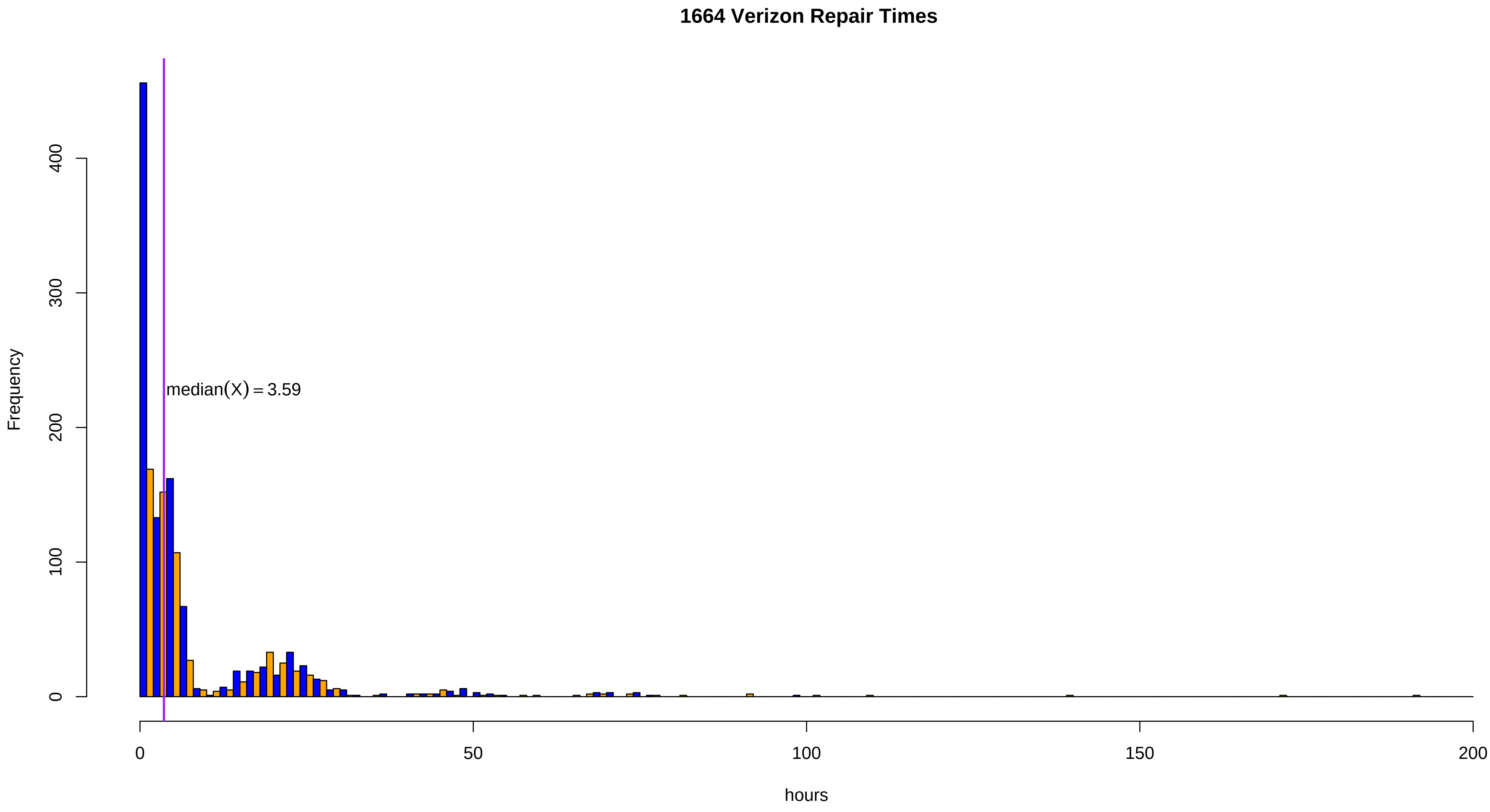
The above linearization argument is reasonable in many situations, because reasonable estimates $\hat{\theta}$ tend to be consistent, i.e., close to θ as n gets large.

If $\hat{\theta} \approx \mathcal{N}(\theta, \tau^2/n)$ then $f(\hat{\theta}) \approx \mathcal{N}(f(\theta), (f'(\theta))^2 \tau^2/n)$ for smooth functions f.

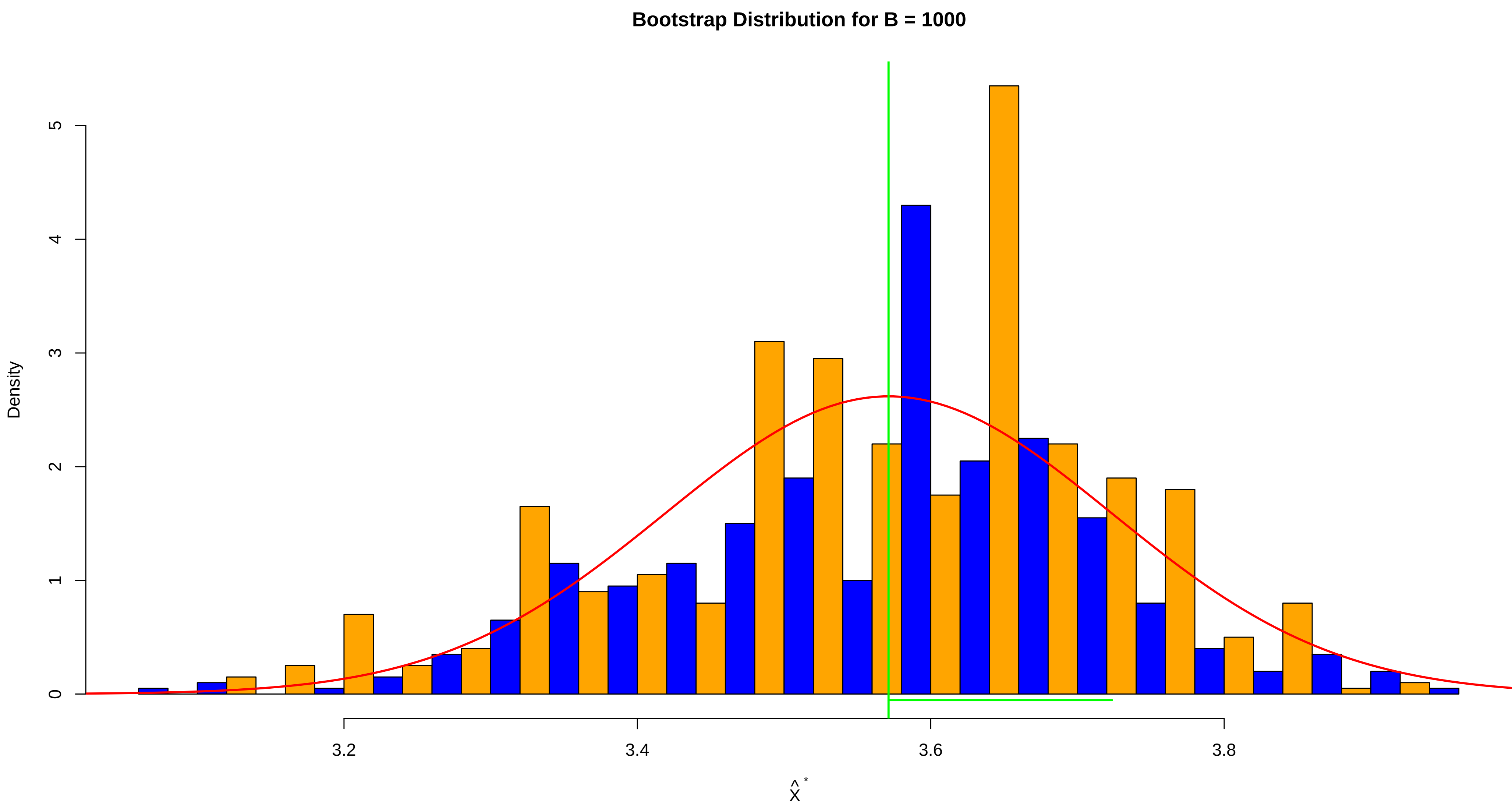
However, we do not always get approximate normality for the bootstrap distribution of an estimator $\hat{\theta}$.

A good example is the sample median $\hat{X}$.

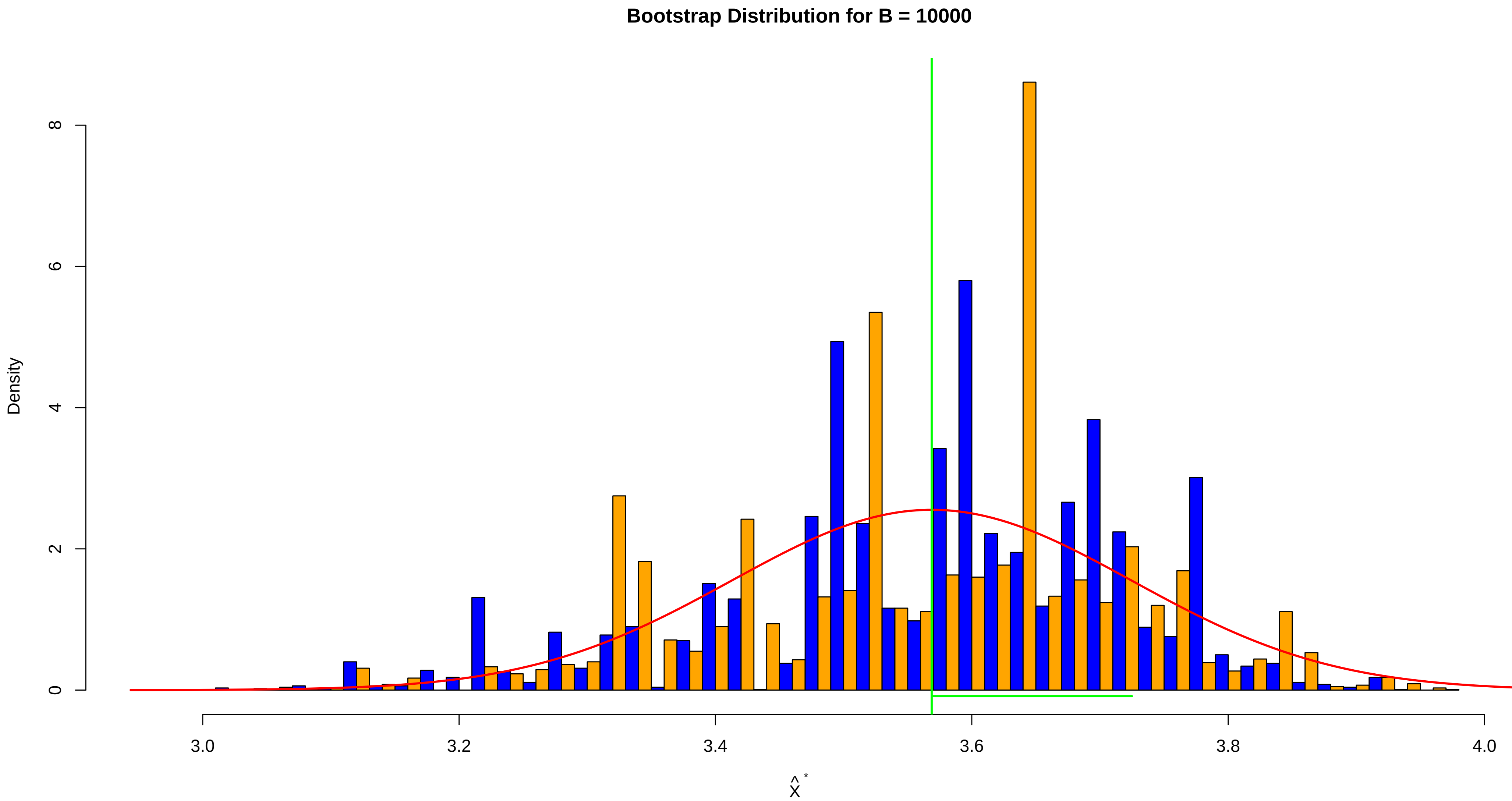
Verizon Repair Data with Median



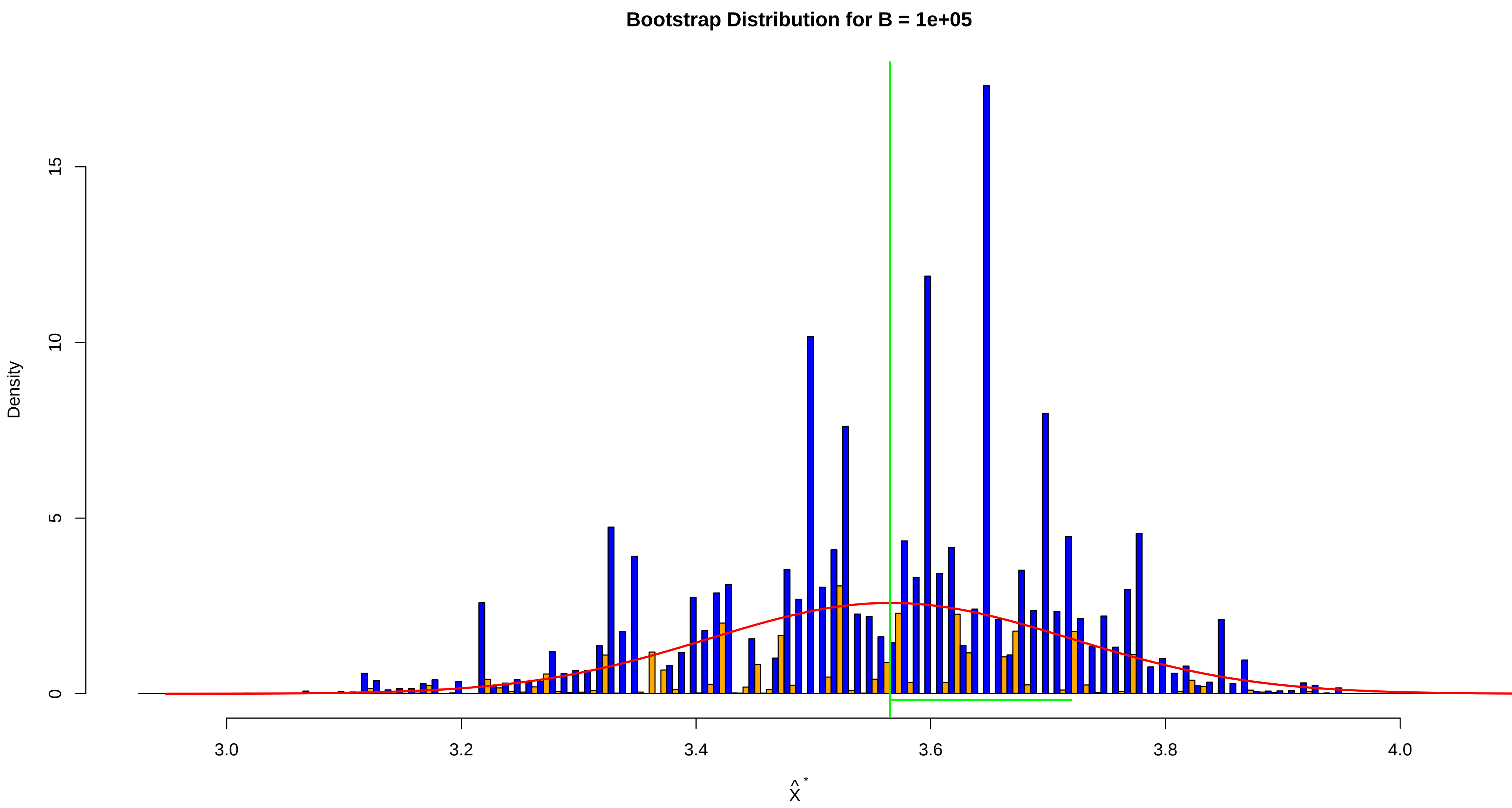
Bootstrap Distribution of Medians



Bootstrap Distribution of Medians



Bootstrap Distribution of Medians



What Happened to Normality?

The sample median is the average of the two middle observations when n is even.

In a bootstrap sample $X_1^\star, \ldots, X_n^\star$ theses two middle observations mostly come from few observations in the middle of the original sample $X_1, \ldots, X_n$.

This is a small and very discrete set of values $\Longrightarrow$ ragged bootstrap distribution.

Theorem: The sample median has an approximate normal distribution provided the cdf F from which the sample is drawn has $F'(m) > 0$ near the median m of F.

Proof idea: $\hat{X} \leq x \iff B_n(x) = \#\{X_i \leq x\} \geq (n+1)/2, \quad B_n \sim$ binomial $\approx$ normal.

However, our bootstrap sample is drawn from $\hat{F}_n$, a step function, not smooth!

Approximate Normality of Sample Median $\hat{X}$

Let $F(m) = 1/2$, i.e., $m =$ population median and assume that $F'(m) > 0$ exists.

$$P(\sqrt{n}(\hat{X} - m) \le x) = P\left(\hat{X} \le m + \frac{x}{\sqrt{n}}\right) = P\left(B_n\left(m + \frac{x}{\sqrt{n}}\right) \ge \frac{n+1}{2}\right)$$

$$= P\left(B_n\left(m + \frac{x}{\sqrt{n}}\right) - n \cdot F\left(m + \frac{x}{\sqrt{n}}\right) \ge n \cdot \left(\frac{1}{2} - F\left(m + \frac{x}{\sqrt{n}}\right)\right) + \frac{1}{2}\right)$$

Write $B_n = B_n(m + x/\sqrt{n})$ and $p_n = F(m + x/\sqrt{n})$ and note that the CLT $\Longrightarrow$ $(B_n - np_n)/\sqrt{np_n(1-p_n)} \approx Z \sim \mathcal{N}(0,1)$. Further $p_n \to 1/2$ and

$$\frac{1}{2} - p_n = \frac{1}{2} - F\left(m + \frac{x}{\sqrt{n}}\right) = F(m) - F\left(m + \frac{x}{\sqrt{n}}\right) \approx -F'(m) \cdot x/\sqrt{n}$$

$$\frac{n(.5 - p_n) + .5}{\sqrt{np_n(1-p_n)}} \approx -2F'(m)x \quad \text{as } n \to \infty \quad \text{and}$$

$$P(\sqrt{n}(\hat{X} - m) \le x) \approx P\left(Z \ge -2F'(m)x\right) = P\left(Z \le 2F'(m)x\right)$$

$$\Longrightarrow \sqrt{n}(\hat{X} - m) \approx \mathcal{N}(0, 1/(2F'(m))^2) \quad \text{or} \quad \hat{X} \approx \mathcal{N}(m, 1/(2\sqrt{n}F'(m))^2)$$

Sample Median for Weibull Samples

Here we generalize the bootstrap concept to the parametric bootstrap.

We have a sample of size n from a Weibull(α, β) distribution $\mathcal{W}(\alpha, \beta)$ with cdf

$$F_{\alpha,\beta}(x) = 1 - \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) \qquad \text{for} \quad x > 0\,,\ \ \alpha > 0\,,\ \ \beta > 0\,.$$

We use the following two quantile estimates

$$\hat{X} = \text{median}(X_1, \ldots, X_n) = \hat{x}_{.5} \qquad \text{and} \qquad \hat{x}_{p_0} \quad \text{with} \quad p_0 = 1 - \exp(-1) = .6321\,.$$

Note that the target quantiles are $m = \text{median}(X) = \alpha\,(-\log(.5))^{1/\beta}$ and $x_{p_0} = \alpha$

from which derives the following expression $\beta = \log(-\log(.5))/(\log(m) - \log(\alpha))$

From these quantile estimates we have as estimates for α and β

$$\hat{\alpha} = \hat{x}_{p_0} \qquad \text{and} \qquad \hat{\beta} = \frac{\log(-\log(.5))}{\log(\hat{X}) - \log(\hat{x}_{p_0})}$$

Parametric Bootstrap Weibull Samples

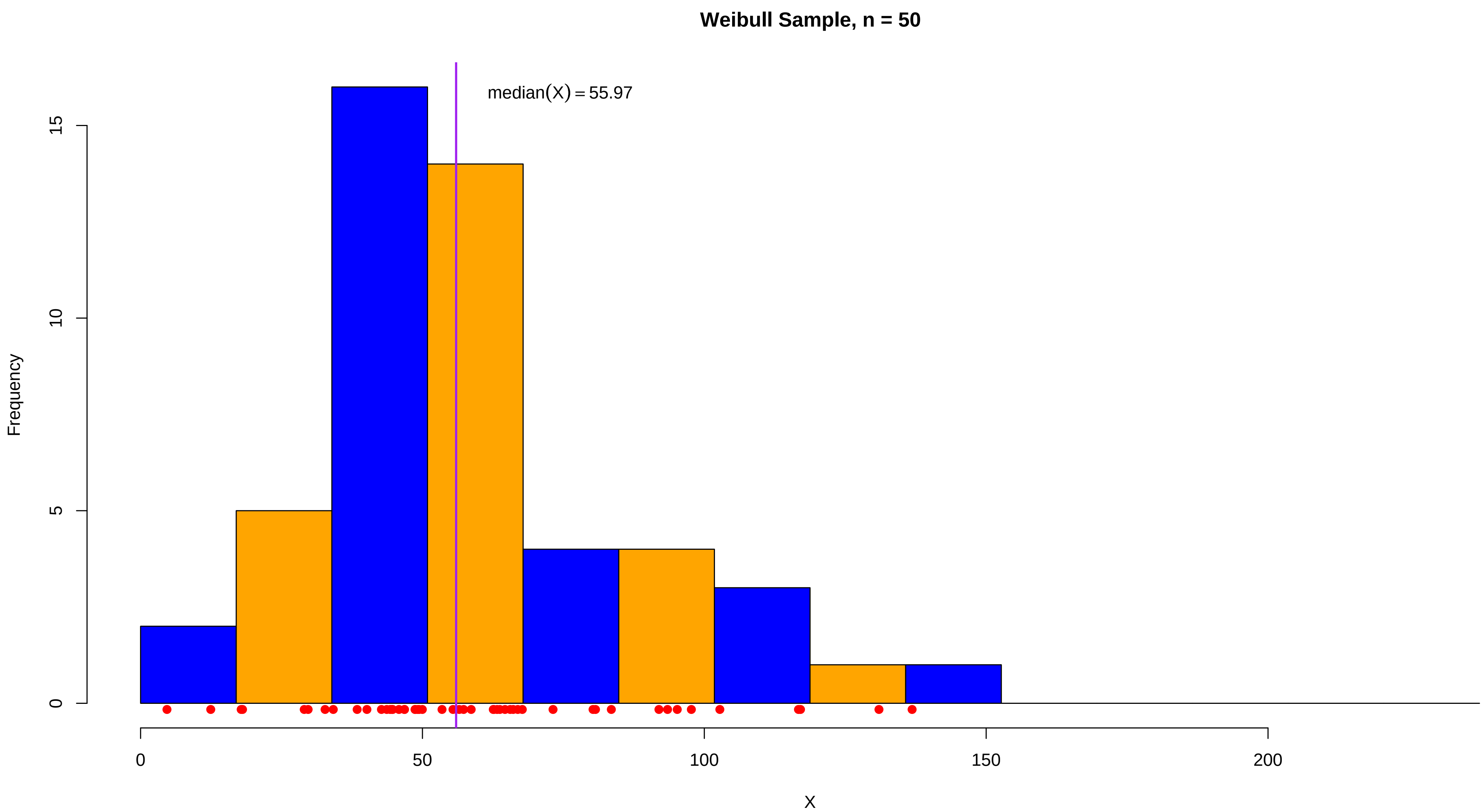
The above estimates $\hat{\alpha}$ and $\hat{\beta}$ define an estimated Weibull distribution $\hat{F} = \mathcal{W}(\hat{\alpha}, \hat{\beta})$ from which we can obtain bootstrap random samples of size n, i.e., $X_1^\star, \ldots, X_n^\star$.

Think of $\hat{F}$ as having the same role as our previous $\hat{F}_n$, which is known as the nonparametric maximum likelihood estimator of F, hence nonparametric bootstrap.

For each such bootstrap sample calculate $\hat{X}^\star$.

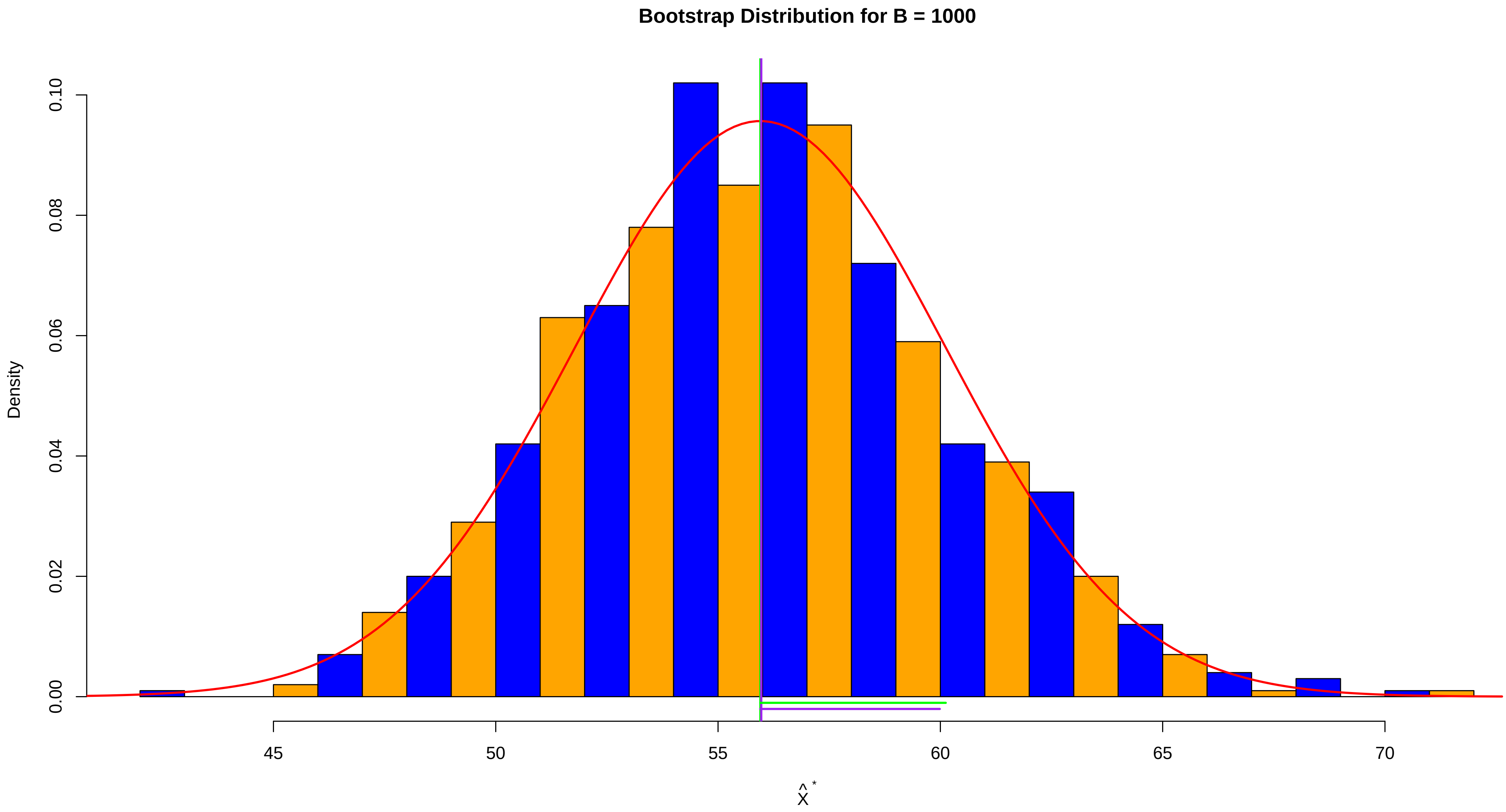
Repeating this $B = 1000$ or more times we get a bootstrap distribution for $\hat{X}^\star$.

Since we are sampling from a smooth cdf (Weibull) we can expect $\approx$ normality from the previously stated theorem, see next few slides.

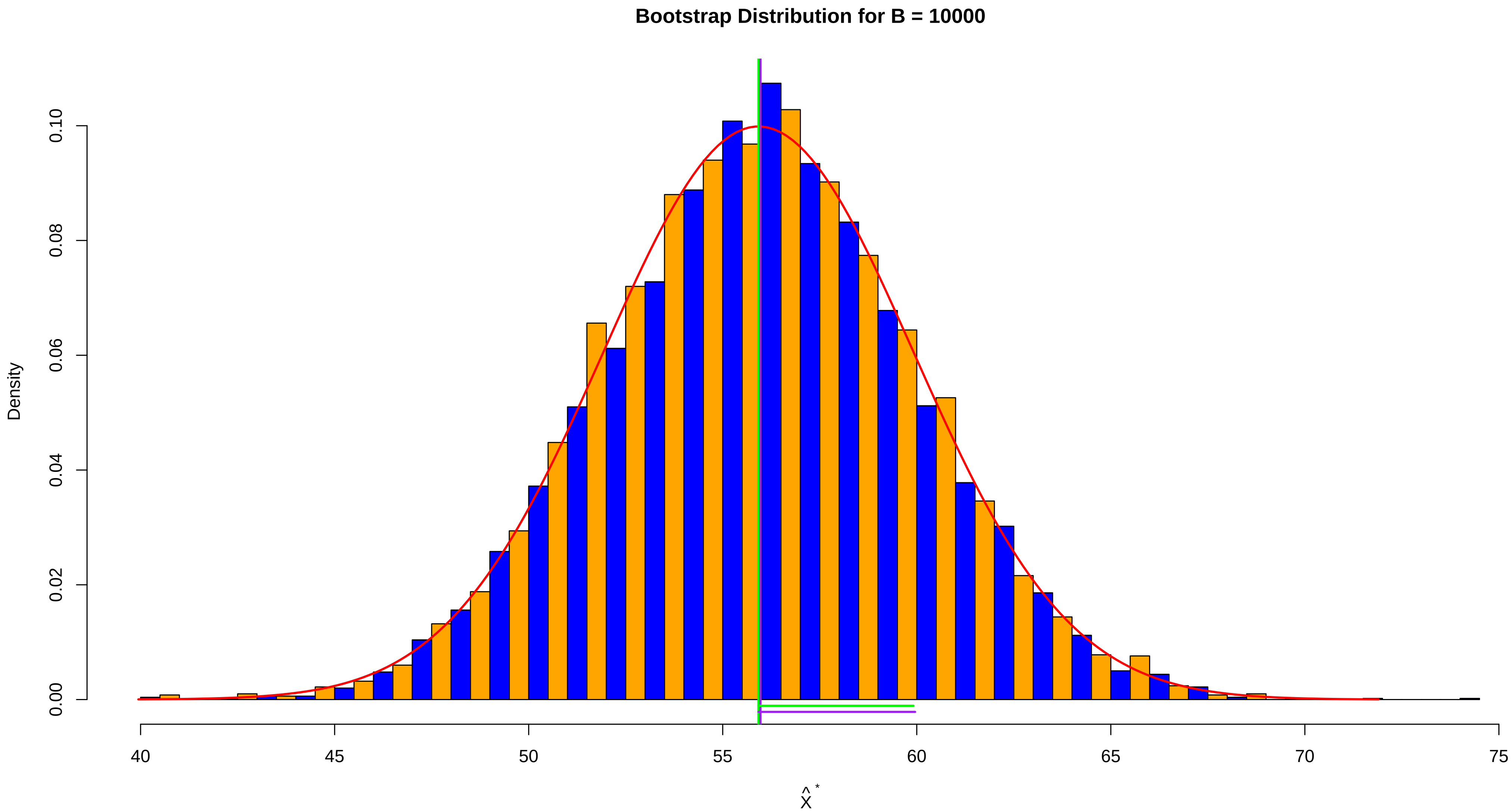


Weibull Sample, n = 50
median(X) = 55.97
Frequency
0
5
10
15
0
50
100
150
200
X

Parametric Bootstrap Distribution of Medians (Weibull)



Parametric Bootstrap Distribution of Medians (Weibull)



Estimation Uncertainty

The bootstrap distribution of $\bar{X}^\star$ is well approximated by a normal distribution, although the sampled population was far from normal.   Due to CLT!

Similarly, the CLT $\Longrightarrow$ the sampling distribution of $\bar{X} \approx \mathcal{N}(\mu, \sigma^2/n)$.

$$\Longrightarrow \quad P\left(|\bar{X}-\mu| \le z_{1-\alpha/2}\,\sigma/\sqrt{n}\right) \approx 1-\alpha = \gamma$$

$$\begin{aligned}\Longrightarrow \quad \gamma = 1-\alpha \quad &\approx \quad P\left(\bar{X} - z_{1-\alpha/2}\,\sigma/\sqrt{n} \le \mu \le \bar{X} + z_{1-\alpha/2}\,\sigma/\sqrt{n}\right) \\ &\approx \quad P\left(\bar{X} - z_{1-\alpha/2}\,s/\sqrt{n} \le \mu \le \bar{X} + z_{1-\alpha/2}\,s/\sqrt{n}\right) \\ &\approx \quad P\left(\bar{X} - t_{n-1,1-\alpha/2}\,s/\sqrt{n} \le \mu \le \bar{X} + t_{n-1,1-\alpha/2}\,s/\sqrt{n}\right)\end{aligned}$$

The first $\approx$ invokes the CLT, the second $\approx$ is due to replacing σ by s, and the third $\approx$ replaces $z_{1-\alpha/2}$ by $t_{n-1,1-\alpha/2}$ to adjust for the previous $\approx$ by analogy with Student-t confidence intervals, to adjust for not so large n.

The Bootstrap Step

Note that in the approximate confidence interval

$$\left[\bar{X} - t_{n-1,1-\alpha/2}\, s/\sqrt{n}\;,\;\; \bar{X} + t_{n-1,1-\alpha/2}\, s/\sqrt{n}\right]$$

we still make use of the theoretical formula $\mathrm{SE}(\bar{X}) = \sigma/\sqrt{n}$.

The bootstrap step consists in using

$$\mathrm{SE}(\bar{X}^{\star}) \approxeq \mathrm{SE}_{\mathrm{boot},\bar{X}} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}\left(\bar{X}_i^{\star} - \bar{\bar{X}}^{\star}\right)^2} \qquad \text{in place of } s/\sqrt{n},$$

i.e., use

$$\left[\bar{X} - t_{n-1,1-\alpha/2}\, \mathrm{SE}_{\mathrm{boot},\bar{X}}\;,\;\; \bar{X} + t_{n-1,1-\alpha/2}\, \mathrm{SE}_{\mathrm{boot},\bar{X}}\right]$$

In using $\mathrm{SE}_{\mathrm{boot},\bar{X}}$ we do not need the theoretical standard error formula of $\bar{X}$.

The Bootstrap Step in General

Suppose we have a sample $X_1,\ldots,X_n$ from some distribution $F \in \mathcal{F}$, where $\mathcal{F}$ is a family of possibilities for the unknown F.

When estimating a parameter $\theta(F)$ using some estimate $\hat{F}$ of F, i.e., using $\hat{\theta} = \theta(\hat{F})$ as estimate of $\theta(F)$, we can generate a bootstrap distribution of $\hat{\theta}^\star_1,\ldots,\hat{\theta}^\star_B$, calculated from bootstrap samples $X^\star_{b1},\ldots,X^\star_{bn}$, $b = 1,\ldots,B$.

If this bootstrap distribution is reasonably normal and centered on the original estimate $\hat{\theta}$ (unbiased), then the previous construction of a $100(1-\alpha)\%$ level confidence interval carries over, i.e.,

$$\left[\hat{\theta} - t_{n-1,1-\alpha/2}\,\mathrm{SE}_{\mathrm{boot},\hat{\theta}}\,,\;\hat{\theta} + t_{n-1,1-\alpha/2}\,\mathrm{SE}_{\mathrm{boot},\hat{\theta}}\right]$$

where $\mathrm{SE}_{\mathrm{boot},\hat{\theta}} = \sqrt{\frac{1}{B}\sum_{i=1}^{B}\left(\hat{\theta}^\star_i - \bar{\hat{\theta}}^\star\right)^2}$ with $\bar{\hat{\theta}}^\star = \frac{1}{B}\sum_{i=1}^{B}\hat{\theta}^\star_i$.

Efron's Percentile Method

To extend this bootstrap idea to situations where the bootstrap distribution does not look normal, Efron suggested the following percentile method to construct a $100(1-\alpha)\%$ level confidence interval for θ:

Determine the $\alpha/2$- and $(1-\alpha/2)$-quantiles $\hat{\theta}^{\star}_{\alpha/2}$ and $\hat{\theta}^{\star}_{1-\alpha/2}$ of the bootstrap distribution and treat $[\hat{\theta}^{\star}_{\alpha/2}\,,\ \hat{\theta}^{\star}_{1-\alpha/2}]$ as $100(1-\alpha)\%$ level confidence interval.

This is close to previous method when the bootstrap distribution $\approx$ normal.

This method is transformation invariant: If $[\hat{\theta}_L, \hat{\theta}_U]$ is a confidence interval for θ then, $[\psi(\hat{\theta}_L), \psi(\hat{\theta}_U)]$ is a confidence of same level for $\psi(\theta)$ for any monotone increasing function ψ of θ.

This is especially appealing when the sampling distribution of $\psi(\hat{\theta})$ is approximately normal for some $\psi \nearrow$. No need to know ψ. S and $S^2 \Rightarrow$ corresponding intervals.

Hall's Percentile Method

If we knew the distribution of $\hat{\theta} - \theta$, say its cdf is $G(x) = P(\hat{\theta} - \theta \leq x)$, then we could use its quantiles $g_{\alpha/2}$ and $g_{1-\alpha/2}$ to get

$$\begin{aligned} 1 - \alpha &= P(g_{\alpha} \leq \hat{\theta} - \theta \leq g_{1-\alpha/2}) \\ &= P(\hat{\theta} - g_{1-\alpha/2} \leq \theta \leq \hat{\theta} - g_{\alpha}) \end{aligned}$$

and thus get the following $100(1-\alpha)\%$ level confidence interval for θ

$$[\hat{\theta} - g_{1-\alpha/2}\,,\ \hat{\theta} - g_{\alpha/2}]$$

Not knowing G we estimate it by the bootstrap distribution of $\hat{\theta}^{\star} - \hat{\theta}$, i.e.,

take its corresponding quantiles $g^{\star}_{\alpha/2}$ and $g^{\star}_{1-\alpha/2}$ in place of $g_{\alpha/2}$ and $g_{1-\alpha/2}$

$$\begin{aligned} [\hat{\theta} - g^{\star}_{1-\alpha/2}\,,\ \hat{\theta} - g^{\star}_{\alpha/2}] &= [\hat{\theta} - (\hat{\theta}^{\star}_{1-\alpha/2} - \hat{\theta})\,,\ \hat{\theta} - (\hat{\theta}^{\star}_{\alpha/2} - \hat{\theta})] \\ &= [2\hat{\theta} - \hat{\theta}^{\star}_{1-\alpha/2}\,,\ 2\hat{\theta} - \hat{\theta}^{\star}_{\alpha/2}] \end{aligned}$$

Not Transformation Invariant

Hall's percentile method is not transformation invariant.

If the sampling distribution of $\hat{\theta}$ is skewed to the right, we tend to get $\hat{\theta}$ values further away from θ on the right of θ and closer in on the left of θ.

$$1-\alpha = P(\theta + g_{\alpha} \le \hat{\theta} \le \theta + g_{1-\alpha/2})$$

Then $(\theta + g_{1-\alpha/2}) - \theta > \theta - (\theta + g_{\alpha/2})$ or $g_{1-\alpha/2} > -g_{\alpha/2}$ (> 0 typically).

In order for the interval $[\hat{\theta} - g_{1-\alpha/2}\,,\ \hat{\theta} - g_{\alpha/2}]$ not to miss its target θ when $\hat{\theta}$ is on the high side, it makes sense to reach further back by using the quantile $-g_{1-\alpha/2}$ at the lower end point.

Similarly, when $\hat{\theta}$ is on the low side, it is OK to reach less far up by using the quantile $-g_{\alpha/2}$ at the upper end point.

Which is Better?

Neither percentile method is uniformly best.

There are many other variants, that I won't go into.

There are also double bootstrap methods that try to calibrate and integrate the uncertainty in the first bootstrap step when stating the overall uncertainty with confidence intervals.

The literature is huge, with many good textbooks on the bootstrap method.

Efron & Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman & Hall

Davison & Hinkley (1997), *Bootstrap Methods and their Applications*, Cambridge University Press

The Bootstrap Has Given Wings to Statistics

We can handle statistical problems without having to assume convenient probability models for our data.

$\Longrightarrow$ Nonparametric Bootstrap.

We can handle inference in plausible probability models that before were mathematically intractable.

$\Longrightarrow$ Parametric Bootstrap.

The bootstrap distribution makes the sampling distribution more understandable to consumers of statistics.

The ideas go beyond simple random samples.

The Abstract Problem

We have some data set $\mathbf{X}$.

$\mathbf{X}$ is uncertain for various reasons (sampling variability, measurement error, etc.)

$\mathbf{X}$ was generated by a probability mechanism/model which we denote by P.

Statistical inference: Use $\mathbf{X}$ to make inference concerning the particulars of P.

A very simple and common data structure:

$\mathbf{X} = (X_1, \ldots, X_n)$ and the X_i are independent and identically distributed (i.i.d.).

Other structures involve known covariates, which can be thought of as being a known part of the specified probability model.

Keeping the data set as generic as possible we emphasize the wide applicability of the bootstrap method.

The Probability Model P & Estimates $\hat{P}$

The probability model P that generated $\mathbf{X}$ is unknown.

This is expressed as: P is one of many possible probability models, i.e., $P \in \mathcal{P}$.

Assume: we can generate data sets from any given probability model $P \in \mathcal{P}$.

We need a method that estimates P based on $\mathbf{X}$ via $\hat{P} = \hat{P}(\mathbf{X})$.

Thus we can generate bootstrap data sets $\mathbf{X}^\star$ from $\hat{P}$.

We are interested in $\theta = \theta(P)$ and estimate it by $\hat{\theta} = \theta(\hat{P}) \implies \hat{\theta}^\star = \theta(\hat{P}(\mathbf{X}^\star))$.

The uncertainty in $\hat{\theta}$ is assessed via the bootstrap distribution of $\hat{\theta}^\star_1, \ldots, \hat{\theta}^\star_B$.

$\implies$ many types of bootstrap confidence intervals for $\theta(P)$.

Batch Data Revisited

We assume the following batch data model

$$X_{ij} = \mu + b_i + e_{ij}, \quad j = 1, \ldots, n_i, \quad \text{and} \quad i = 1, \ldots, k \,,$$

$$\begin{array}{llll} \text{where} & b_i \sim \mathcal{N}(0, \sigma_b^2) & & \text{(between batch variation effect)} \\ \text{and} & e_{ij} \sim \mathcal{N}(0, \sigma_e^2) & & \text{(within batch variation effects)} \,. \end{array}$$

b_i and $\{e_{ij}\}$ are assumed to be mutually independent $\Longrightarrow X_{ij} \sim \mathcal{N}(\mu, \sigma_b^2 + \sigma_e^2)$

Quantity of interest is the p-quantile of the X_{ij} distribution $\mathcal{N}(\mu, \sigma_b^2 + \sigma_e^2)$, i.e.,

$$x_p = \mu + z_p \sqrt{\sigma_b^2 + \sigma_e^2} \quad \text{where} \quad z_p = \Phi^{-1}(p) \quad \text{standard normal quantile.}$$

Denote the data set of the above structure by

$$\mathbf{X} = \left\{X_{ij} : j = 1, \ldots, n_i, \quad \text{and} \quad i = 1, \ldots, k\right\}$$

Estimating Batch Data Parameters

$$SS_b = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{X}_{i\cdot} - \bar{X})^2 = \sum_{i=1}^{k} n_i (\bar{X}_{i\cdot} - \bar{X})^2 \quad \text{and} \quad SS_e = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_{i\cdot})^2 .$$

Take $\hat{\sigma}_e^2 = SS_e/(N-k)$ as unbiased estimate of σ_e^2 and $\hat{\tau}^2 = SS_b/(k-1)$ as unbiased estimate of

$$\tau^2 = \sigma_e^2 + \sigma_b^2 \frac{N}{k-1} \left(1 - \sum_{i=1}^{k} \left(\frac{n_i}{N} \right)^2 \right) = \sigma_e^2 + \sigma_b^2 \frac{N}{k-1} \frac{f}{f+1} .$$

$\Rightarrow \hat{\sigma}_b^2 = \left(\hat{\tau}^2 - \hat{\sigma}_e^2 \right) (k-1)(f+1)/(N f)$ as unbiased estimate for σ_b^2.

Redefine $\hat{\sigma}_b^2 = \max(0, \hat{\sigma}_b^2)$, it will no longer be unbiased.

The p-quantile estimate is

$$\hat{x}_p = \bar{X} + z_p \sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_b^2}$$

Batch Data Generation

In HW6 we constructed a function `batch.data.make` that created batch data of the type described above. This can be done for any set of batch sample sizes, $n_1,\ldots,n_k$, and for any number k of batches.

Besides `nvec` $=(n_1,\ldots,n_k)$, further inputs to `batch.data.make` are `sig.e` $=\sigma_e$, `sig.b` $=\sigma_b$, and `mu` $=\mu$.

By replacing μ, σ_e, and σ_b by estimates $\hat{\mu}=\bar{X}$, $\hat{\sigma}_e$, and $\hat{\sigma}_b$ in the call to `batch.data.make` we get a parametric bootstrap batch data set with same `nvec` $=(n_1,\ldots,n_k)$.

We can repeat this many times, say $B=10000$ times.

For all these parametric bootstrap batch data sets we compute

$$\bar{X}_\ell, \quad \hat{\sigma}^\star_{e,\ell}, \quad \hat{\sigma}^\star_{b,\ell} \qquad \text{and} \qquad \hat{x}_{p,\ell}=\bar{X}^\star_\ell+z_p\sqrt{\hat{\sigma}^{\star 2}_{e,\ell}+\hat{\sigma}^{\star 2}_{b,\ell}}\,, \qquad \ell=1,2,\ldots,B\,.$$

Parametric Bootstrap Distribution of Quantile Estimates for Batch Data