

bootstrap

Bin He

October 28, 2019

Purpose

Demonstrate the use of bootstrap to estimate the properties of an estimator, such as mean, variance and so on.

In this case, we actually know the statistical properties of the estimator, in this case sample mean. If a random variable is normally distributed with mean of μ and variance of σ^2 , the sample mean is an unbiased estimator for the population mean μ . Additionally, \bar{x} itself is a normally distributed random variable with its mean and variance as follows:

$$\bar{x} = \sum_{i=1}^n x_i / n$$

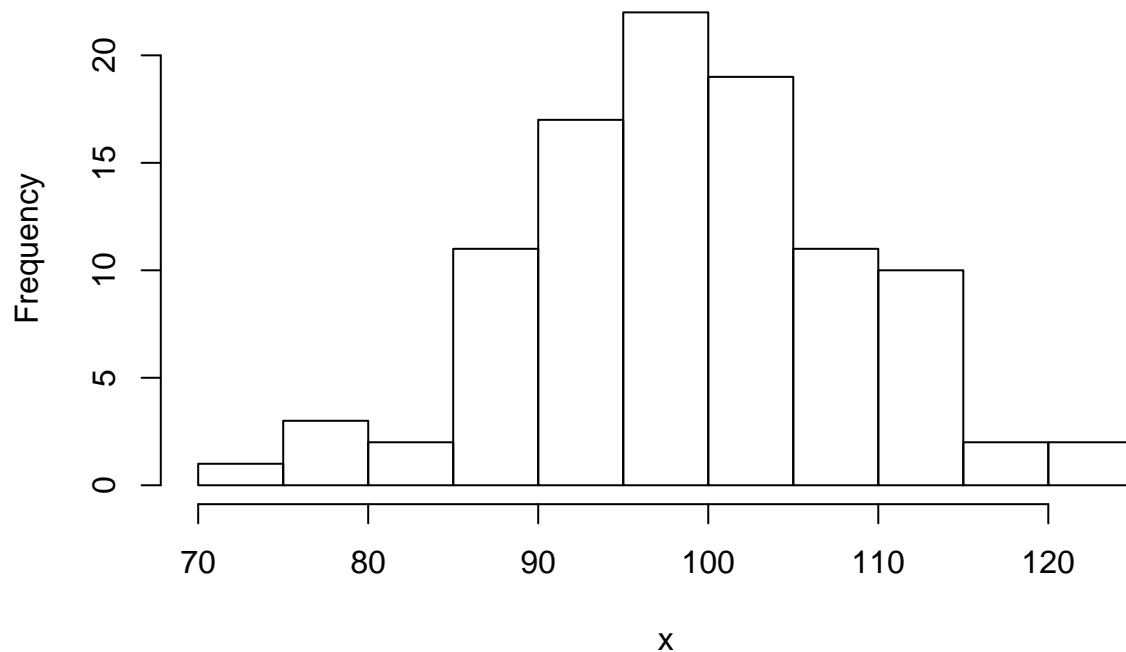
$$E(\bar{x}) = \mu$$

$$Var(\bar{x}) = \frac{\sigma^2}{n}$$

What we are going to do is to calculate the 95% confidence interval for the sample mean estimator for a normally distributed random variable, both using the known statistical property and using the bootstrap method.

```
# let our random variable be normally distributed with mean of mu and variance of sigma^2
mu <- 100
sigma2 <- 100
n <- 100
s <- rnorm(n, mu, sqrt(sigma2)) # take a sample of size n
hist(s, xlab = "x")
```

Histogram of s



```
sbar <- mean(s)

# 1. using the known statistical property of the estimator to calculate CI
p <- 0.95 # calculate 95% confidence interval
z <- qnorm(1-(1-p)/2)
ci.95 <- c(sbar-z*sqrt(sigma2/n), sbar+z*sqrt(sigma2/n))
print("Using the known statistical property of the sample mean estimator to calculate 95% confidence interval")

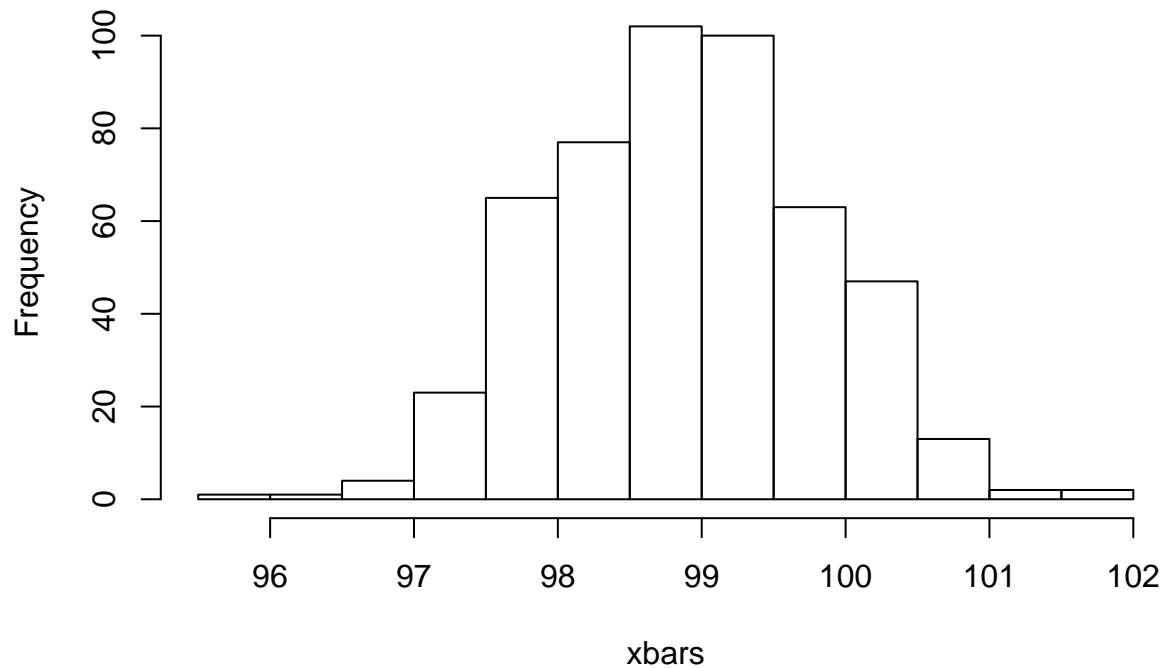
## [1] "Using the known statistical property of the sample mean estimator to calculate 95% confidence interval"

print(ci.95)

## [1] 96.90732 100.82725

# 2. using the known statistical property of the estimator to calculate CI
# first, generate bootstrap samples
N = 500
xbars <- c(); xvars <- c() # store the mean and variance from each bootstrap sample
for (i in c(1:N)){
  bs <- sample(s, n, replace=T)
  xbars <- c(xbars, mean(bs))
  xvars <- c(xvars, var(bs))
}
hist(xbars)
```

Histogram of xbars



```
print("Using bootstrap sample to calculate 95% confidence interval")
```

```
## [1] "Using bootstrap sample to calculate 95% confidence interval"
```

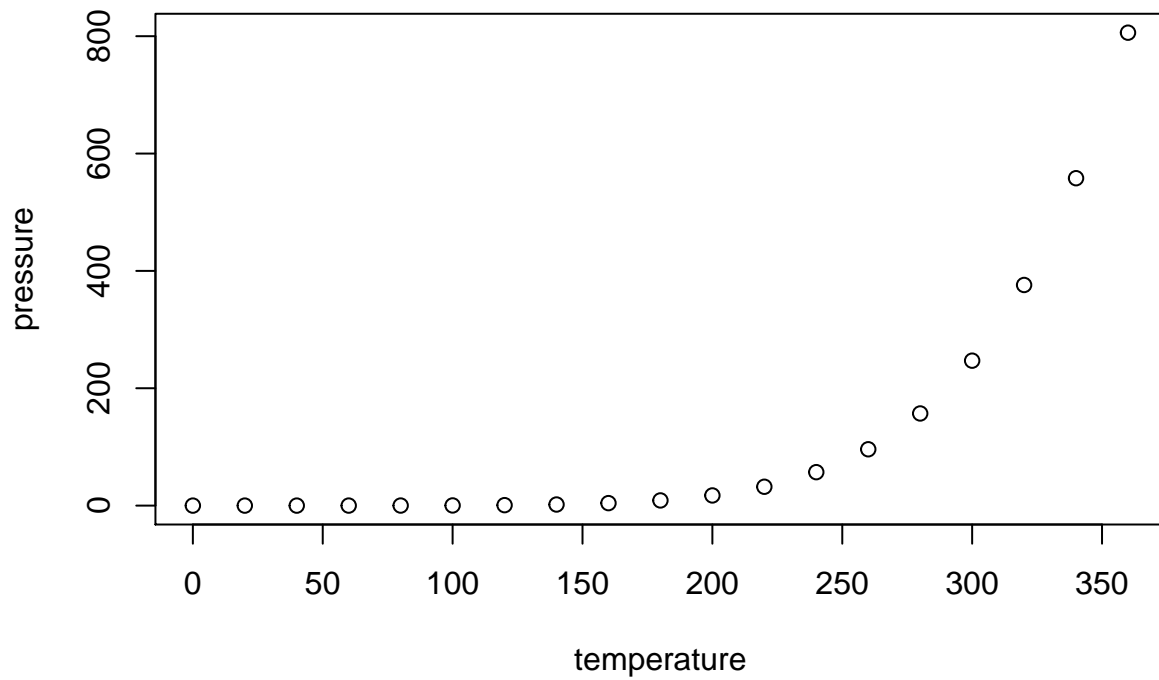
```
print(quantile(xbars, probs = c(0.025, 0.975)))
```

```
##      2.5%      97.5%
```

```
## 97.1622 100.5641
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.