

Empirical Bootstrap Distribution

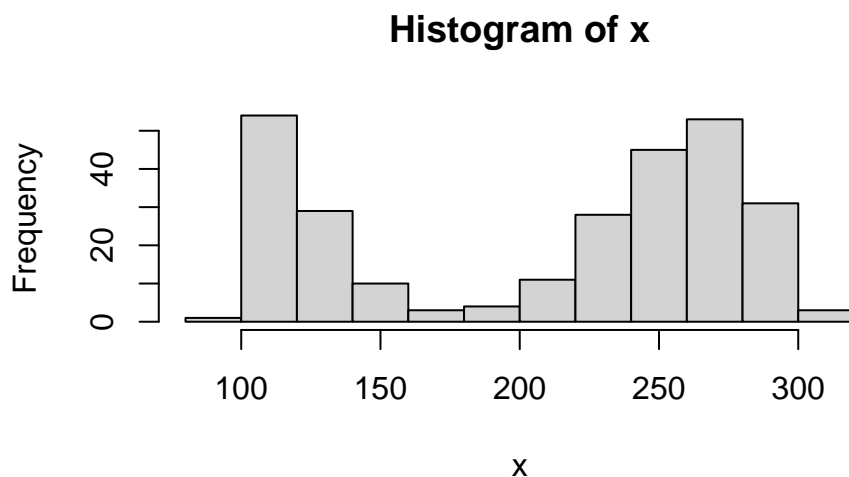
Becky Tang

2024-04-09

Example 1: bootstrapping for a mean

Consider the following sample of eruption times (in seconds) of the Old Faithful geyser:

```
# obtain data
data("faithful")
x <- faithful$eruptions * 60
hist(x)
```

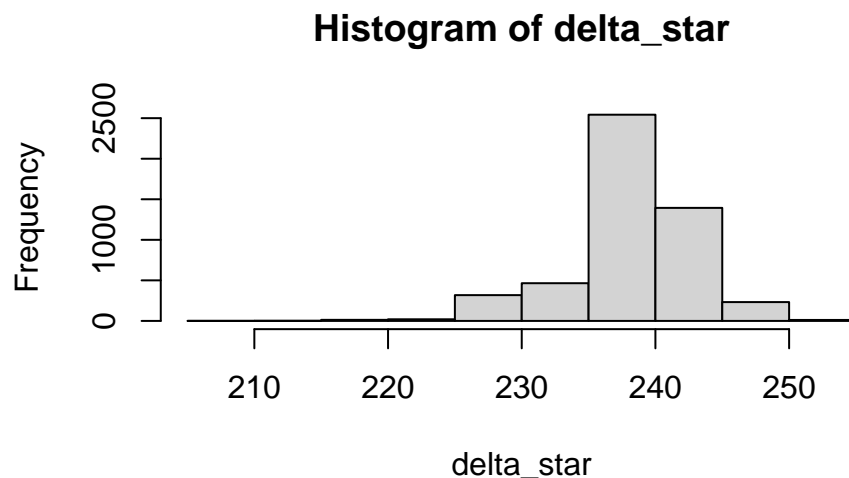


This data is certainly not Normal!

Perhaps we'd like to obtain a bootstrap distribution for the median eruption time.

```
n <- length(x)
B <- 5000
delta_star <- rep(NA, B)
for(i in 1:B){
  xstar <- sample(x, size = n, replace = T)
  delta_star[i] <- median(xstar)
}

# bootstrap distribution of medians
hist(delta_star)
```



```
# bootstrap estimate of bias
mean(delta_star) - median(x)
```

```
## [1] -1.104252
```

Example 2: bootstrapping for MSE

The coefficient of variation of a distribution is the quantity $\frac{\sigma}{\mu}$, where σ and μ are the standard deviation and mean of the distribution, respectively. Let us generate some Poisson data and obtain estimates of the mean squared error (MSE) of the estimator $\frac{s}{\bar{x}}$ where s is the sample standard deviation and \bar{x} is the sample mean.

Notice that in the following, I am setting a seed using the `set.seed()` function. The input doesn't really matter. What this function does is ensure that the random generator in R generates the same sequence of random values. That way when we knit or run the same code that involves random sampling on different laptops, we will get the same result.

```
# generate some Poisson data
set.seed(309)
lambda_true <- 1
n <- 20
x <- rpois(n, lambda_true)
```

An estimate for coefficient variation from the data is:

```
cv_hat <- sd(x)/mean(x)
cv_hat
```

```
## [1] 1.00006
```

Bootstrap estimate

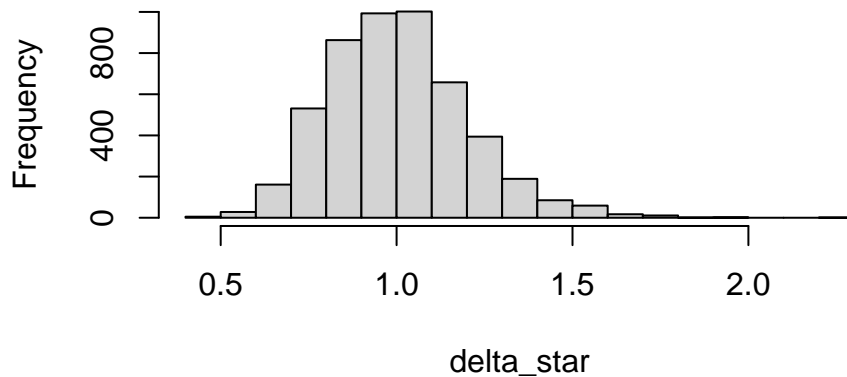
```
B <- 5000
delta_star <- rep(NA, B)
for(i in 1:B){
```

```

# if size isn't specified in sample(), defaults to length of x!
xstar <- sample(x, replace = T)
delta_star[i] <- sd(xstar)/mean(xstar)
}
hist(delta_star)

```

Histogram of delta_star



```

# bootstrap estimate of MSE
mean((delta_star - cv_hat)^2)

```

```
## [1] 0.03948413
```

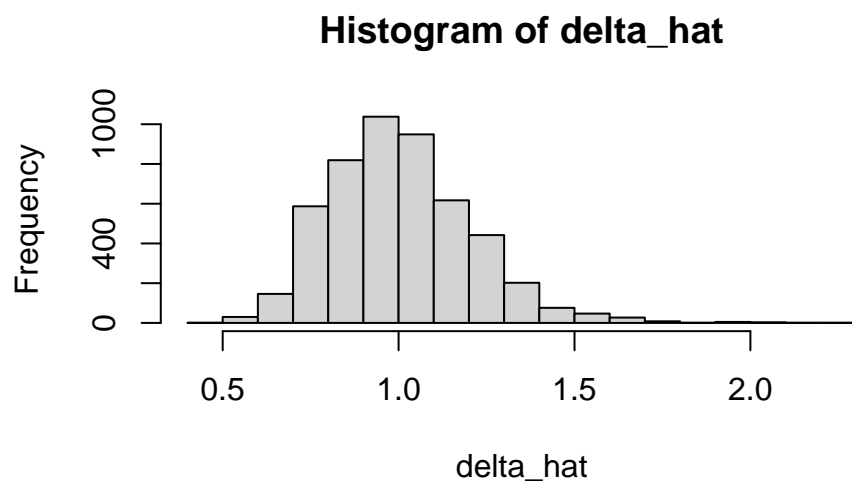
“Monte Carlo” estimate

Since we know the true distribution, we don’t need to bootstrap to estimate the MSE of $\frac{s}{\bar{x}}$ for $\frac{\sigma}{\mu}$. That is, rather than resampling from the original observation \mathbf{x} , we can actually simulate new data sets from the Poisson distribution. Then we can compare the estimates to the true coefficient of variation, which for a $\text{Poisson}(\lambda)$ distribution is $\frac{\sqrt{\lambda}}{\lambda}$.

```

delta_hat <- rep(NA, B)
for(i in 1:B){
  x_new <- rpois(n, lambda_true)
  delta_hat[i] <- sd(x_new)/mean(x_new)
}
hist(delta_hat)

```



```
cv_true <- sqrt(lambda_true)/lambda_true  
mean((delta_hat - cv_true)^2)
```

```
## [1] 0.04043435
```