

# Estimators and Bootstrap Samples

## Contents

<b>Estimation</b>	<b>1</b>
Estimators for $\alpha$ and $\beta$ . . . . .	1
Bootstrap Estimates for the Standard Error . . . . .	2
Empirical Estimates for the estimators. . . . .	2
<b>Bootstrap Estimates for area of petal.</b>	<b>3</b>

In this project given that the population has a Gamma distribution, we will explore the estimation of the shape and scale parameters obtained using the method moments.

## Estimation

Suppose we are given the following sample of data values coming from a Gamma distribution.

```
values <- c(2.670, 0.612, 1.218, 0.899, 2.036, 0.643, 0.489, 2.189, 0.132, 0.177,
            0.363, 0.664, 0.252, 1.866, 0.418, 1.711, 2.708, 2.960, 1.234, 0.384,
            0.460, 0.805, 1.317, 0.407, 0.963, 0.578, 1.591, 1.818, 1.554, 1.240,
            0.694, 0.441, 0.278, 1.391, 0.399, 2.706, 0.973, 0.471, 2.149, 0.851)
```

### Estimators for $\alpha$ and $\beta$ .

Recall from class that we calculated the estimators for  $\alpha$  and  $\beta$  using the method of moments for a sample of data values of size  $n$  as:

$$\hat{\alpha} = \frac{\overline{X}^2}{(1/n) \sum X_i^2 - \overline{X}^2} \quad \hat{\beta} = \frac{(1/n) \sum X_i^2 - \overline{X}^2}{\overline{X}}$$

Write functions “alpha\_hat” and “beta\_hat” that take input a sample of data values “samp” and output the given estimators.

```
alpha_hat <- function(samp){
  n <- length(samp)
  mean_sq <- (mean(samp))^2
  sum_sq <- sum(samp^2)
  estimate <- mean_sq / (sum_sq / n - mean_sq)
  return(estimate)
}

beta_hat <- function(samp){
  estimate <- mean(samp) / alpha_hat(samp)
  return(estimate)
}
```

Use the Estimators to estimate the values for  $\alpha$  and  $\beta$  for the Gamma distribution from which we got the sample of data values “values”.

```
a_hat <- alpha_hat(values)
b_hat <- beta_hat(values)
a_hat
```

```
## [1] 2.004434
```

```
b_hat
```

```
## [1] 0.5576512
```

What is the estimate for the mean  $\mu$  for the population distribution, using the estimates for  $\alpha$  and  $\beta$ .

```
mu_hat <- a_hat * b_hat
mu_hat
```

```
## [1] 1.117775
```

Note that this is the same as the mean of the “values”.

```
mean(values)
```

```
## [1] 1.117775
```

## Bootstrap Estimates for the Standard Error

Using bootstrap samples (take  $B = 10000$ ) coming from “values”, derive the estimates for the standard error for each of the estimators  $\hat{\alpha}$  and  $\hat{\beta}$ .

```
B <- 10000
alpha_hats_boot <- replicate(B, alpha_hat(sample(values, size=length(values), replace=TRUE)))
sd(alpha_hats_boot)
```

```
## [1] 0.4007326
```

```
beta_hats_boot <- replicate(B, beta_hat(sample(values, size=length(values), replace=TRUE)))
sd(beta_hats_boot)
```

```
## [1] 0.08432576
```

## Empirical Estimates for the estimators.

Now fix  $\alpha = 2$  and  $\beta = 0.5$ . Apply the estimators  $\hat{\alpha}$  and  $\hat{\beta}$  to  $B = 10000$  random samples of size 40 coming from a Gamma Distribution with shape parameter  $\alpha = 2$  and scale parameter  $\beta = 0.5$ , and use this to estimate the standard error of the estimators.

```
alpha <- 2
beta <- 0.5
n <- 40

alpha_hats_true <- replicate(B, alpha_hat(rgamma(n, shape=alpha, scale=beta)))
sd(alpha_hats_true)
```

```
## [1] 0.5868389
```

```
beta_hats_true <- replicate(B, beta_hat(rgamma(n, shape=alpha, scale=beta)))
sd(beta_hats_true)
```

```
## [1] 0.1408142
```

Compare your standard errors to those that you got from the bootstrap sample in the previous part. What do you notice? ANSWER: Standard errors of the estimators obtained from samples of the Gamma Distribution are slightly higher than those obtained from the bootstrap samples.

## Bootstrap Estimates for area of petal.

Recall that the area of an ellipse with major axis  $a$  and minor axis  $b$  is given by

$$\pi ab.$$

Under the assumption that the shape of the petal of the iris flower is approximately an ellipse with the major and minor axes given by petal length and petal width (or vice versa), we can estimate the petal surface area (only one side) of iris of setosa species by the following estimator:

$$\hat{\theta} = \pi \cdot \bar{X} \cdot \bar{Y},$$

where  $\bar{X}$  and  $\bar{Y}$  are the sample means of petal length and petal width respectively. (Note it might also have been reasonable to calculate the mean of  $\pi \cdot x \cdot y$  for every pair  $(x, y)$  of petal length and petal width.)\

Estimate the surface area of the petal of a iris setosa flower using the iris dataset.

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

data(iris)
setosa <- filter(iris, Species == "setosa")
# divide by 4 because we need major axis and minor axis, not length and width
theta_hat <- pi * mean(setosa$Petal.Length) * mean(setosa$Petal.Width) / 4
theta_hat

## [1] 0.28247
```

Now use bootstrap sampling to estimate the standard error in your estimated petal surface area. Make sure you identify the correct output.

```
boot_pet_area_set <- replicate(B, {
  samp <- slice_sample(setosa, n=dim(setosa)[1], replace=TRUE)
  x <- mean(samp$Petal.Length)
  y <- mean(samp$Petal.Width)
  # divide by 4 because we need major axis and minor axis, not length and width
  pi*x*y/4
})
sd(boot_pet_area_set)

## [1] 0.01908957
```