HW6

Kwan Ho Lee

6/4/2021

Problem 2

(a) Find the distribution of θ (in terms of θ and n).

(b) Derive the analytic expression for the variance of θ.

We want to find the variance of the MLE $\hat{\theta}$.

$$\mathrm{var}(\hat{\theta}) = \mathrm{E}(\hat{\theta}^2) - \mathrm{E}(\hat{\theta})^2$$

$$\mathrm{E}_\theta(\hat{\theta}) = \int_0^\theta x f_{X_{(n)}}(x)dx = \int_0^\theta xn\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}dx = \frac{n}{n+1}\theta$$

$$\mathrm{E}_\theta(\hat{\theta}^2) = \int_0^\theta x^2 f_{X_{(n)}}(x)dx = \int_0^\theta x^2 n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}dx = \frac{n}{n+2}\theta^2$$

$$\mathrm{var}(\hat{\theta}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = n\theta^2\left(\frac{1}{n+2} - \frac{n}{(n+1)^2}\right)$$

```
[1] 0.003327123
```

(c) Generate a data set of size n = 50 and θ = 3. Then generate B = 5000 bootstrap samples using parametric bootstrap. Use the bootstrap samples to approximate Var(θ). Compare your answer to (b).

```
[1] 0.003332027
```

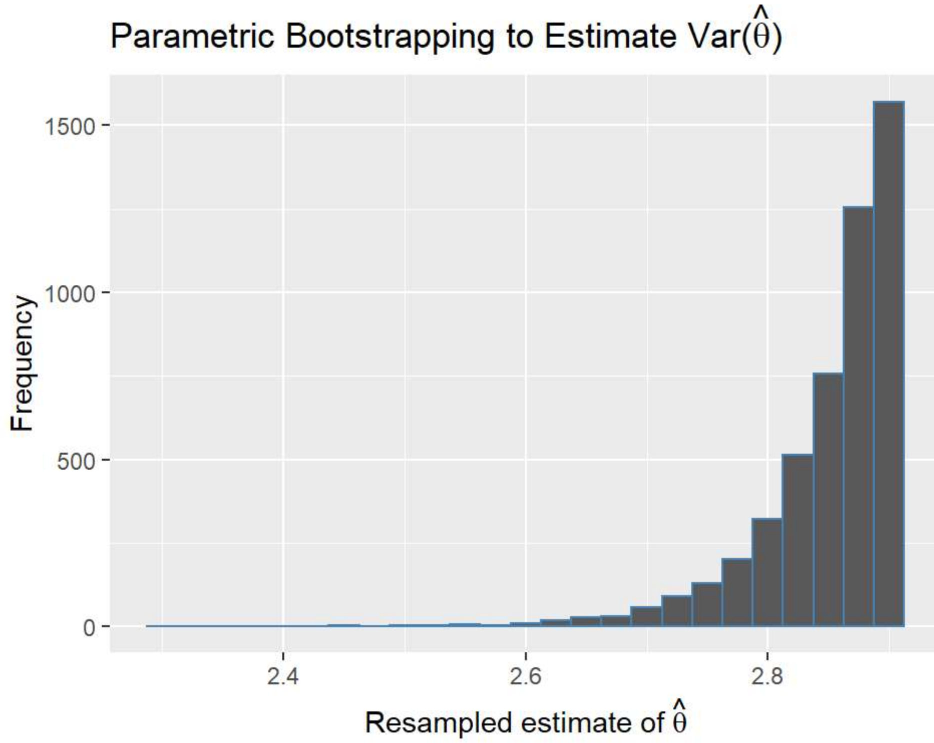
The parametric bootstrapped variance of $\hat{\theta}$ is 0.003332 and the analytical variance is 0.0033271. The log-ratio of these two values is 0.0014728, indicating that the parametric bootstrapped estimate is very close to its analytical variance value.

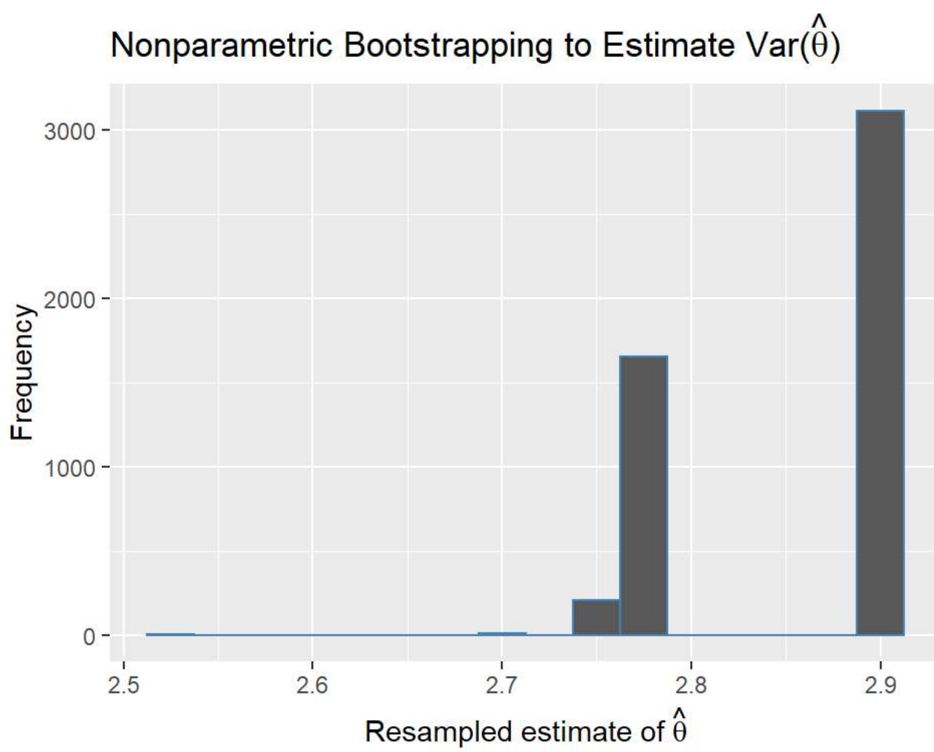
(d) With the same data set: repeat (c) with nonparametric bootstrap.

```
[1] 0.00426073
```

The nonparametric bootstrapped variance of $\hat{\theta}$ is 0.0042607 and the analytical variance is 0.0033271. The log-ratio of these two values is 0.2473327, indicating that the bootstrapped estimate is somewhat close to analytical value.

(e) With the same data set: plot the histograms of θ obtained from the parametric and nonparametric bootstraps.





Appendix

The following code is used in Problem 2 part (b)

```
n <- 50
t <- 3

## Calculate the analytical variance

var_anal <- n*t^2*(1/(n+2) - n/(n+1)^2)

var_anal
```

The following code is used in Problem 2 part (c)

```
n <- 50
t <- 3
B <- 5000

# Create the dataset

xs <- runif(n, 0, t)

# generate samples

tmax <- max(xs)
bs_par_raw <- runif(n*B, 0, tmax)
bs_par <- matrix(bs_par_raw, nrow=B, ncol=n)

# calculate max of each sample

maxs_par <- apply(bs_par, 1, max)

# calculate the parametric bootstrapped variance

var_boot_par <- var(maxs_par)

var_boot_par
```

The following code is used in Problem 2 part (d)

```
## Generate non-parametric bootstrap samples

# generate bootstrap

bs_nonpar_raw <- sample(xs, size=n*B, replace=TRUE)
bs_nonpar <- matrix(data=bs_nonpar_raw, nrow=B, ncol=n)

# calculate max of each bootstrap sample

maxs_nonpar<- apply(bs_nonpar, 1, max)

# calculate non-parametric bootstrapped variance of the mle

var_boot_nonpar = var(maxs_nonpar)

var_boot_nonpar
```

The following code is used in Problem 2 part (e)

```
# Plot parametric bootstrap samples

plt_var_par = ggplot(data=data.frame(x=maxs_par), aes(x=x)) +
    geom_histogram(color="steelblue", binwidth=0.025) +

    labs(
        x=TeX("Resampled estimate of $\\hat{\\theta}$"),
        y="Frequency",
        title=
        TeX("Parametric Bootstrapping to Estimate Var($\\hat{\\theta}$)")
    )

plt_var_par



# Plot non-parametric bootstrap samples

plt_var_nonpar = ggplot(data=data.frame(x=maxs_nonpar), aes(x=x)) +
    geom_histogram(color="steelblue", binwidth=0.025) +

    labs(
        x=TeX("Resampled estimate of $\\hat{\\theta}$"),
        y="Frequency",
        title=
        TeX("Nonparametric Bootstrapping to Estimate Var($\\hat{\\theta}$)")
    )

plt_var_nonpar
```