Homework 7: Partial Solutions

This assignment is mostly computational. Here are some remarks on a few of the problems.

Chapter 9, Problem 5. The bootstrap distribution is:

x	X_1	X_2	$\cdots$	X_n
$\mathbb{P}(X^* = x)$	$1/n$	$1/n$	$\cdots$	$1/n$

So

$$\mathbb{E}(X_i^*|X_1,\ldots,X_n) = \sum_x x\,\mathbb{P}(X^* = x) = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}$$

and

$$\mathbb{V}(X_i^*|X_1,\ldots,X_n) = \sum_x (x-\overline{X})^2\,\mathbb{P}(X^* = x) = \frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2.$$

Now $\overline{X}_n^* = n^{-1}\sum_{i=1}^n X_i^*$ and hence,

$$\mathbb{E}(\overline{X}^*|X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i^*|X_1,\ldots,X_n) = \overline{X}$$

and

$$\mathbb{V}(\overline{X}^*|X_1,\ldots,X_n) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}(X_i^*|X_1,\ldots,X_n) = \frac{1}{n^2}\sum_{i=1}^n (X_i-\overline{X})^2.$$

Now we get the unconditional mean and variance:

$$\mathbb{E}(\overline{X}^*) = \mathbb{E}\Big(\mathbb{E}(\overline{X}^*|X_1,\ldots,X_n)\Big) = \mathbb{E}(\overline{X}) = \mu$$

and

$$\mathbb{V}(\overline{X}^*) \;=\; \mathbb{V}\Big(\mathbb{E}(\overline{X}^*|X_1,\ldots,X_n)\Big) + \mathbb{E}\Big(\mathbb{V}(\overline{X}^*|X_1,\ldots,X_n)\Big)$$

$$
\begin{aligned}
&= \mathbb{V}(\overline{X}) + \mathbb{E}\left(\frac{1}{n^2}\sum_{i=1}^n (X_i - \overline{X})^2\right) \\
&= \frac{\sigma^2}{n} + \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}(X_i - \overline{X})^2) \\
&= \frac{\sigma^2}{n} + \frac{n-1}{n^2}\mathbb{E}\left(\frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}\right) \\
&= \frac{\sigma^2}{n} + \frac{n-1}{n^2}\sigma^2 \\
&= \frac{\sigma^2}{n}\left(2 - \frac{1}{n}\right).
\end{aligned}
$$

Chapter 9, Problem 7. We have computed the exact distribution of $\widehat{\theta}$ before. Here it is again:

$$
\mathbb{P}(\widehat{\theta} \le c) = \mathbb{P}(\max\{X_1, \ldots, X_n\} \le c) = \prod_{i=1}^n \mathbb{P}(X_i \le c) F_\theta(c)^n = \left(\frac{c}{\theta}\right)^n.
$$

The density of $\widehat{\theta}$ is therefore

$$
g(\widehat{\theta}) = \left(\frac{n}{\theta}\right)\left(\frac{\widehat{\theta}}{\theta}\right)^{n-1}.
$$

The parametric bootstrap will mimick this fairly well. The nonparametric bootstrap will not.

Let $X_{(n)} = \max\{X_1, \ldots, X_n\}$. Then,

$$
\begin{aligned}
\mathbb{P}(\widehat{\theta}^* = \widehat{\theta}) &= \mathbb{P}(\max\{X_1^*, \ldots, X_n^*\} = \max\{X_1, \ldots, X_n\} \\
&= \mathbb{P}(X_{(n)} \text{ is in the bootstrap sample}) \\
&= 1 - \mathbb{P}(X_{(n)} \text{ is not in the bootstrap sample}) \\
&= 1 - \left(1 - \frac{1}{n}\right)^n \\
&\to 1 - e^{-1} = .632.
\end{aligned}
$$

Thus, $\mathbb{P}(\widehat{\theta} = \theta)$ while $\mathbb{P}(\widehat{\theta}^* = \widehat{\theta}) \approx .632$ which shows that the nonparametric bootstrap does not mimick the true distribution well.