ECMA 31350, Homework 1

Andi Liu

1. **Problem 1**
   Let $X_1, \dots, X_n$ be a random sample from an unknown distribution P with $\mu = \mathbb{E}_p[X]$ and $\sigma^2 = \mathbb{V}ar_p(X)$. Show that

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

   is a consistent estimator of σ^2.

   Solution:

$$\begin{aligned}
\mathbb{E}[\widehat{\sigma}_n^2] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \bar{X}_n)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu + \mu - \bar{X}_n)^2] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)^2] + \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)(\mu - \bar{X}_n)] + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\bar{X}_n - \mu)^2]
\end{aligned}$$

   Using $\mathbb{E}[(X_i - \mu)^2] = \sigma^2$ and $\mathbb{E}[\bar{X}_n] = \mu$, this simplifies to:

$$\mathbb{E}[\widehat{\sigma}_n^2] = \sigma^2 + \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)(\mu - \bar{X}_n)] + \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(\bar{X}_n - \mu)^2]$$

   From which we can see that as n approaches infinity, the first term approaches σ^2, amd the second and third terms approach 0, thus

$$\mathbb{E}[\widehat{\sigma}_n^2] \xrightarrow{p} \sigma^2$$

2. **Problem 2**
   Let $X_1, \dots, X_n$ be a random sample from a uniform distribution $U[0, \theta]$. Show that $\hat{\theta_n} = \max\{X_1, \dots, X_n\}$ is a consistent estimator of θ. That is, for any $\epsilon > 0$,

$$\lim_{n\to\infty}\mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) = 0$$

   Solution: Thus, the cumulative distribution function for U is given as

$$F(x) = \begin{cases} \frac{x}{\theta} & \text{if } 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

   Then, the probability that $\hat{\theta}_n$ is less than or equal to some value x is given as

$$\begin{aligned}
\mathbb{P}(\hat{\theta}_n \le x) &= \mathbb{P}(max\{X_1, \dots, X_n\} \le x) \\
&= P(X_1 \le x, \dots, X_n \le x)
\end{aligned}$$

As we know that the random variables are independent, we can write this as the product of individual probabilities:

$$\begin{aligned}\mathbb{P}(\hat{\theta}_n \leq x) &= \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) \\ &= \left(\frac{x}{\theta}\right)^n\end{aligned}$$

Now we want to determine the probability that $|\hat{\theta}_n - \theta| > \epsilon$.

We must consider two cases, where $\hat{\theta}_n \geq \theta + \epsilon$ and $\hat{\theta}_n \leq \theta - \epsilon$. First note that if $\epsilon \geq \theta$, trivially, our probability is 0: it is impossible for $\hat{\theta}_n$ to be greater than $\theta + \epsilon$ if $\epsilon \geq \theta$, or for $\hat{\theta}_n$ to be less than $\theta - \epsilon$ if $\epsilon \geq \theta$. Thus, we can assume from now that $\epsilon < \theta$.

(a) The case where $\hat{\theta}_n > \theta + \epsilon$ is trivial: by definition, $\hat{\theta}_n$ can never be larger than θ, so this probability is 0, thus it is also true that

$$\lim_{n\to\infty} \mathbb{P}(\hat{\theta}_n > \theta + \epsilon) = 0$$

(b) Now consider the case where $\hat{\theta}_n < \theta - \epsilon$. We have

$$\mathbb{P}(\hat{\theta}_n < \theta - \epsilon) = \left(\frac{\theta - \epsilon}{\theta}\right)^n$$

Since we have assumed that $\epsilon < \theta$, it must be that $0 < \frac{\theta - \epsilon}{\theta} < 1$, and

$$\lim_{n\to\infty} \mathbb{P}(\hat{\theta}_n < \theta - \epsilon) = 0$$

Now we have shown that in all cases, $\lim_{n\to\infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) = 0$, completing our proof.
■

3. **Problem 3**
Airlines often overbook flights to maximize profit. Last year, I flew about 50 times and about 5 of those flights were overbooked. Let $X_i = 1$ if the flight it overbooked, and $X_i = 0$ otherwise. Given $X_1, \ldots, X_{50}$, how would you estimate the probability of a flight being overbooked, $p = P(X = 1)$? Denoting your estimator by $\hat{p}_n$, what is the asymptotic distribution of $\sqrt{n}(\hat{p} - p)$? Based on this asymptotic approximation, how would you construct a 95% confidence interval for this parameter?

Solution: Notice that X_i is a Bernoulli random variable with parameter p. Thus, we can write that $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Furthermore, by definition of a Bernoulli random variable, we have $\mathbb{E}[\hat{p}_n] = p$ and $\mathbb{V}ar[\hat{p}_n] = \frac{p(1-p)}{n}$. Thus, the asymptotic distribution of $\sqrt{n}(\hat{p}_n - p)$ is given as

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{d} N(0, p(1-p))$$

To construct a 95% confidence interval, we can use the fact that the 95% confidence interval for a normal distribution is given as

$$\left[\hat{p}_n - 1.96\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + 1.96\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}\right]$$

4. **Problem 4**

   Write a Monte Carlo simulation to check coverage of the confidence interval for the mean with known variance and with estimated variance. For each simulation:

   (1) Draw $X_1, \ldots, X_n$ from a distribution of your choice, for which you know the mean and the variance.

   (2) Construct two confidence intervals: with known standard deviation and with the estimated one.

   (3) Check if each of them covers the true value of the mean and save the result.

   Perform 1000 simulations and report how often each of the intervals covers the true value. Repeat this exercise for $n = 30, 100, and 500$. Discuss the results.

   Solution: I chose to draw $X_1, \ldots, X_n$ from a normal distribution with mean 0 and variance 1. I then constructed the two confidence intervals as described in the previous problem. I performed 1000 simulations for each value of n, and the results are shown in the table below.

n	Known Variance	Estimated Variance
30	0.94.3	0.93.5
100	0.94.8	0.95.1
500	0.95.1	0.95.1

   We can see that the coverage is very close to the expected 95% for all values of n. This is expected, as the confidence interval is constructed using the normal distribution, which is symmetric. Thus, the coverage should be close to 95% for all values of n.

   The Python code used to generate these results is included as a separate file.

My submission for lecture 2's problem set is attached as a separate, juptyer notebook file.