Question 1: Point Estimator Uniform

Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$ and let $\hat{\theta} = \max\{X_1, \ldots, X_n\}$. Find the bias, standard error (SE), and mean squared error (MSE) of this estimator.

1.1 Bias

$$
\begin{align}
\hat{F}(x) &= \mathbb{P}(\max\{X_1, \ldots, X_n\} \leq x) \tag{1}\\
&= \mathbb{P}(X_1 < x, X_2 < x, \ldots, X_n < x) \tag{2}\\
&= \mathbb{P}(X_1 < x)\mathbb{P}(X_2 < x) \ldots \mathbb{P}(X_n < x) \quad \text{because } X_i \text{ are i.i.d} \tag{3}\\
&= \prod_{i=1}^{n} \mathbb{P}(X_i < x) \tag{4}\\
&= \prod_{i=1}^{n} F_{X_i}(x) \tag{5}\\
&= \prod_{i=1}^{n} \frac{x}{\theta} \tag{6}\\
&= (\frac{x}{\theta})^n \tag{7}\\
& \tag{8}
\end{align}
$$

The density is:

$$
\begin{align}
\frac{d}{dx}\hat{F}(x) &= \hat{f}(x) \tag{9}\\
&= \frac{d}{dx}(\frac{x}{\theta})^n = \frac{n}{\theta}(\frac{x}{\theta})^{n-1} \tag{10}
\end{align}
$$

The expected value $\mathbb{E}(\hat{\theta})$ is:

$$\mathbb{E}(\hat{\theta}) = \int_0^\theta x\frac{n}{\theta}(\frac{x}{\theta})^{n-1}dx \tag{11}$$

$$= \int_0^\theta x^n(\frac{n}{\theta^n})dx \tag{12}$$

$$= \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{\theta^n}\cdot\frac{x^{n+1}}{n+1}\Big|_0^\theta = \frac{n\theta^{n+1}}{\theta^n(n+1)} = \frac{n\theta}{n+1} \tag{13}$$

$\text{Bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$

$$Bias(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta \tag{14}$$

$$= \frac{n\theta}{n+1} - \theta \tag{15}$$

$$= -\frac{1}{n+1} \tag{16}$$

1.2 SE

$$SE(\hat{\theta}) = \sqrt{\mathbb{V}(\hat{\theta})} \tag{17}$$

$$\mathbb{V}(\hat{\theta}) = \mathbb{E}(\hat{\theta}^2) - \mathbb{E}(\hat{\theta})^2 \tag{18}$$

$$\tag{19}$$

Examining $\mathbb{E}(\hat{\theta}^2)$:

$$\mathbb{E}(\hat{\theta}^2) = \int_0^\theta x^2\frac{n}{\theta}(\frac{x}{\theta})^{n-1}dx \tag{20}$$

$$= \int_0^\theta x^{n+1}(\frac{n}{\theta^n})dx \tag{21}$$

$$= (\frac{n}{\theta^n})\int_0^\theta x^{n+1}dx \tag{22}$$

$$= \frac{n}{\theta^n}\cdot\frac{x^{n+2}}{n+2}\Big|_0^\theta = \frac{n\theta^{n+2}}{\theta^n(n+2)} = \frac{n\theta^2}{n+2} \tag{23}$$

We know that:

$$\mathbb{E}(\hat{\theta})^2 = \frac{n\theta}{n+1}$$

Combining these two, we get:

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

Simplifying:

$$\frac{n\theta^2}{(n+1)^2(n+2)}$$

Thus the SE is:

$$SE = \sqrt{\frac{n\theta^2}{(n+1)^2(n+2)}} = \frac{\sqrt{n}\theta}{(n+1)\sqrt{n+2}}$$

1.3 MSE

$$MSE(\hat{\theta}) = \frac{n\theta^2}{(n+1)^2(n+2)} + \left(\frac{-1}{n+1}\right)^2 \tag{24}$$

$$= \frac{n\theta^2}{(n+1)^2(n+2)} + \frac{1}{(n+1)^2} \tag{25}$$

$$= \frac{n\theta^2 + n + 2}{(n+1)^2(n+2)} \tag{26}$$

Question 2: Chapter 7 # 1

Theorem 7.3

$$\begin{aligned}\text{At any fixed value of } x,\\ \mathbb{E}[\hat{F}_n(x)] = F(x),\\ \mathrm{Var}(F_n(x)) = \frac{F(x)(1-F(x))}{n},\\ \mathrm{mse} = \frac{F(x)(1-F(x))}{n} \to 0,\\ P[F_n(x) \to F(x)].\end{aligned}$$

2.1 $\mathbb{E}[\hat{F}_n(x)] = F(x)$

****Proof****

Using the definition of expected value:

$$\mathbb{E}(\hat{F}_n(x)) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(x - x_i)\right)$$

Each x_i is independent so,

$$\mathbb{E}(\hat{F}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}(x - x_i)\right)$$

Lets evaluate the $\mathbb{E}\left(\mathbb{1}(x - x_i)\right)$.

$$\mathbb{E}\left(\mathbb{1}(x - x_i)\right) = \int_{-\infty}^{\infty} f_X(x) \cdot \mathbb{1}(x - x_i)dx$$

$f_X(x)$ is the density function for the underlying random variable. As, $\mathbb{1}(x - x_i) = 1$ for all $x_i \geq x$, and $\mathbb{1}(x - x_i) = 0$ for all $x_i < x$ we an simplify this as:

$$\mathbb{E}\left(\mathbb{1}(x - x_i)\right) = \int_{x_i}^{\infty} f_X(x)dx = \mathbb{P}(x \geq x_i)$$

This is exactly the definition of the CDF. So, the expected value of the indicator function is the CDF(i.e.) $\mathbb{E}\left(\mathbb{1}(x - x_i)\right) = F(x)$.

From here, we can plug back into the original equation for the expected value of the empirical CDF.

$$\mathbb{E}(\hat{F}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left(\mathbb{1}(x - x_i)\right) = \frac{1}{n}\sum_{i=1}^{n} F(x)$$

$F_X(x)$ does not depend on the index of summation, so we get:

$$\mathbb{E}(\hat{F}_n(x)) = F(x)$$

$\square$

2.2 Var$(\hat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$

****Proof****

Lets start with the definition of the variance:

$$Var(\hat{F}_n(x)) = Var\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(x - x_i)\right)$$

Because each x_i is i.i.d. and we have that:

$$Var(\hat{F}_n(x)) = \frac{1}{n^2}\sum_{i=1}^{n} Var\left(\mathbb{1}(x - x_i)\right)$$

Lets calculate the variance of $\mathbb{1}(x - x_i)$:

$$Var(\mathbb{1}(x - x_i)) = \int_{-\infty}^{\infty} f_X(x)\mathbb{1}(x - x_i)^2 dx - F(x)^2 \tag{27}$$

$$= \int_{x_i}^{\infty} f_X(x)1^2 dx - F(x)^2 \tag{28}$$

$$= \int_{x_i}^{\infty} f_X(x) dx - F(x)^2 \tag{29}$$

$$= \mathbb{P}(x \geq x_i) - F(x)^2 \tag{30}$$

$$= F(x) - F(x)^2 \tag{31}$$

$$= F(x)(1 - F(x)) \tag{32}$$

Thus, we have it that:

$$Var(\hat{F}_n(x)) = \frac{1}{n^2}\sum_{i=1}^{n} F(x)(1-F(x))$$

Simplified this is:

$$Var(\hat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$$

$\square$

2.3 mse $= \frac{F(x)(1-F(x))}{n} \to 0$

****Proof****

The mean squared error(MSE) is given as

$$MSE(\hat{F}_n) = Var(\hat{F}_n) + Bias(\hat{F}_n)$$

We computed the variance in 2.2, so now we just need to compute the bias. The bias is:

$$Bias(\hat{F}_n) = \mathbb{E}(\hat{F}_n) - F(x) \tag{33}$$
$$= F(x) - F(x) = 0 \tag{34}$$

Thus the Bias of the empirical CDF is 0, and we have that the mean squared error is:

$$MSE(\hat{F}_n) = \frac{F(x)(1-F(x))}{n} + 0$$

Proving that this value tends toward 0 in the limit as $n \to \infty$ is simple as, $F(x)(1-F(x))$ does not depend on n and can be treated as some finite constant, C. Thus we have,

$$\lim_{n\to\infty} \frac{F(x)(1-F(x))}{n} = \lim_{n\to\infty} \frac{C}{n} = 0$$

$\square$

2.4 $P[F_n(x) \to F(x)]$

****Proof****

This proof uses Chebyshev's inequality

Let $\epsilon > 0$

$$\mathbb{P}\left(\left|\hat{F}(x) - \mathbb{E}(\hat{F}(x))\right| > \epsilon\right) \leq \frac{Var(\hat{F}(x))}{\epsilon^2}$$

Now, plugging in the values for the expected value and the variance that we calculated previously,

$$\mathbb{P}\left(\left|\hat{F}(x) - F(x))\right| > \epsilon\right) \leq \frac{F(x)(1-F(x))}{n\epsilon^2}$$

Taking the limit of both sides gives us our result that:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\hat{F}(x) - F(x))\right| > \epsilon\right) \leq \lim_{n\to\infty} \frac{F(x)(1-F(x))}{n\epsilon^2} = 0$$

So,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\hat{F}(x) - F(x))\right| > \epsilon\right) \leq 0$$

Trivially, because it is a probability distribution

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\hat{F}(x) - F(x))\right| > \epsilon\right) \geq 0$$

Therefore,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\hat{F}(x) - F(x))\right| > \epsilon\right) = 0$$

And therefore, we say that $\hat{F}(x) \xrightarrow{p} F(x)$ $\square$

Question 3: Chapter 7 #4

Let $X_1, \ldots, X_n \sim F$ and let $\hat{F}_n(x)$ be the empirical distribution function. For a fixed x, use the central limit theorem to find the limiting distribution of $\hat{F}_n(x)$.

Proof

The emperical CDF is defined as follows:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(x - x_i)$$

where,

$$\mathbb{1}(y) = \begin{cases} 0 & y < 0 \\ 1 & y \geq 0 \end{cases}$$

We can treat the result of the indicator function as a random variable, as x_i is random:

$$Y_i = \mathbb{1}(x - x_i)$$

Substituting this in:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

To establish that the CLT applies, we have to show that Y_i has bounded expected value and variance

$$E(Y_i) = E(\mathbb{1}(x - x_i)) = \int_{-\infty}^{\infty} f_X(x) \cdot \mathbb{1}(x - x_i) dx$$

$f_X(x)$ is the density function for the underlying random variable. As, $\mathbb{1}(x - x_i) = 1$ for all $x_i > x$, we an simplify this as:

$$E(Y_i) = \int_{x_i}^{\infty} f_X(x) dx = F(x)$$

The varience of Y_i:

$$Var(Y_i) = Var(\mathbb{1}(x - x_i)) = E(\mathbb{1}(x - x_i)^2) - E(\mathbb{1}(x - x_i))^2$$

Lets examine $E(\mathbb{1}(x-x_i)^2)$:

$$E(\mathbb{1}(x-x_i)^2) = \int_{-\infty}^{\infty} f_X(x) \cdot \mathbb{1}(x-x_i)^2 dx$$

We can do something similar:

$$\int_{x_i}^{\infty} f_X(x) \cdot 1^2 dx = F(x)$$

Varience is

$$Var(Y_i) = F(x) - F(x)^2 = F(x)(1-F(x))$$

The expected value and variance are bounded, so we can apply the CLT.

$$Z_n = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i - E(Y_i)}{\sqrt{\frac{\sigma_Y^2}{n}}} \xrightarrow{d} N(0,1)$$

Plugging in:

$$Z_n = \frac{\hat{F}_n(x) - F(x)}{\frac{F(x)(1-F(x))}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

This implies that

$$\hat{F}_n(x) \xrightarrow{d} N\left(F(x), \frac{F(x)(1-F(x))}{n}\right)$$

Now, lets take the limit as n goes to infinity to evaluate what happens to the limiting distribution as we take more and more samples of the underlying random variable.

$$\lim_{n\to\infty} N\left(F(x), \frac{F(x)(1-F(x))}{n}\right) \to N\left(F(x), 0\right)$$

So the limiting distribution is a normal distribution with no variance centered around $F(x)$.

Therefore the limiting distribution is $F(x)$. □

Question 4: Chapter 7 #9

100 people are given a standard antibiotic to treat an infection and another 100 are given a new antibiotic. In the first group, 90 people recover, in the second group, 85 people recover. Let p1 be the probability of recovery under the standard treatment and let p_2 be the probability of recovery under the new treatment. We are interested in estimating $\theta = p_1 - p_2$. Provide an estimate, standard error, an 80 percent confidence interval, and a 95 percent confidence interval for θ.

4.1 Estimate

Let $X_1, X_2, \ldots, X_{100}$ be random variables 0 or 1, indicating whether a patient recovers from the first group. Similarly, let $Y_1, Y_2, \ldots, Y_{100}$ be random variables 0 or 1, indicating weather a patient recovers from the second group.

We have it that $X_i \sim \text{Bernoulli}(p1)$ and $Y_i \sim \text{Bernoulli}(p2)$

Lets calculate $\hat{p_1}$. p_1 is the expected value for each bernoulli random variable X_i, so it's estimator can can be expressed as:

$$\hat{p_1} = \mathbb{E}(\hat{F}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

As we take more samples, this sum converges to the expected value of the emerical CDF for $\hat{p_1}$ which gives us a good estimator for p_1 as the expected value of its CDF is equal to p_1.

Similarly, we can define $\hat{p_2}$:

$$\hat{p_2} = \mathbb{E}(\hat{F}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Thus, we can use the estimator that $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i$.

In this question $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{90}{100} - \frac{85}{100} = 0.05$

4.2 Standard Error

Now we can begin calculating the standard error for $\hat{\theta}$

SE

$$
\begin{aligned}
SE(\hat{\theta}) &= \sqrt{Var(\hat{\theta})} && (35)\\
&= \sqrt{Var(\hat{p1} - \hat{p2})} && (36)\\
&= \sqrt{Var(\hat{p1}) + Var(\hat{p2})} && (37)\\
&= \sqrt{Var(\frac{1}{n}\sum_{i=1}^{n} X_i) + Var(\frac{1}{n}\sum_{i=1}^{n} Y_i)} && (38)\\
&= \sqrt{\frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) + \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i)} \qquad \text{because each } X_i, Y_i \text{ are iid} && (39)\\
&= \sqrt{\frac{1}{n^2}\sum_{i=1}^{n} \hat{p_1}(1-\hat{p_1}) + \frac{1}{n^2}\sum_{i=1}^{n} \hat{p_2}(1-\hat{p_2})} \quad \text{because } X_i, Y_i \sim \text{Bernoulli} && (40)\\
&= \sqrt{\frac{\hat{p1}(1-\hat{p1}) + \hat{p2}(1-\hat{p2})}{n}} && (41)
\end{aligned}
$$

The standard error for the estimator is $\sqrt{\frac{\hat{p1}(1-\hat{p1})+\hat{p2}(1-\hat{p2})}{n}}$.
Plugging in the numbers we got from the experiment,

$$\sqrt{\frac{.90(1-.90)+.85(1-.85)}{100}} \approx \sqrt{0.002175} = 0.0466$$

4.3 Confidence intervals

From the central limit theorem, we get that, each $\hat{p}_1$ and $\hat{p}_2$ converge to a normal distribution as they are both the sums of bernoulli random variables that have finite expectation and varience.

The difference of two random variables that converge to a normal distribution also converge, so we know that $\hat{\theta}$ converges.

$$\sqrt{n}(\hat{\theta_n} - \theta) \xrightarrow{d} N(0, SE(\hat{\theta})^2)$$

Thus we can normalize the distribution by dividing by the SE,

$$\frac{\sqrt{n}(\hat{\theta_n} - \theta)}{\sqrt{SE(\hat{\theta})^2}} \xrightarrow{d} N(0,1)$$

We can use this standard normal distribution to compute a confidence interval as follows, if $1 - \alpha$ is the percent confidence we want and C is the confidence interval:

$$C = (\hat{\theta} - z_{\frac{\alpha}{2}} \cdot Se(\hat{\theta}), \hat{\theta} + z_{\frac{\alpha}{2}} \cdot SE(\hat{\theta}))$$

where z_x is the z-value for a given x

4.3.1 80% Confidence

We want an 80% confidence interval so, $\alpha = .2$

$$C = (\hat{\theta} - z_{\frac{\alpha}{2}} \cdot Se(\hat{\theta}), \hat{\theta} + z_{\frac{\alpha}{2}} \cdot SE(\hat{\theta}))$$

Plugging in what we know:

$$C = (0.05 - z_{0.1} \cdot 0.0466, 0.05 + z_{0.1} \cdot 0.0466)$$

This is equivalent to:

$$C = (-0.010, 0.110)$$

4.3.2 95% Confidence

Similarly, we want an 95% confidence interval so, $\alpha = .05$:

$$C = (0.05 - z_{0.025} \cdot 0.0466, 0.05 + z_{0.025} \cdot 0.0466)$$

This comes out to:

$$C = (-0.041, 0.141)$$

□

Question 5: Supplemental Question

Let $x_1, x_2, \ldots, x_n$ be i.i.d. samples from a random variable X with expected value $\mu = \mathbb{E}(X)$ and variance $\sigma^2 = \mathrm{Var}(X)$. For each of the following three estimators for μ, compute the bias, variance, and MSE. Discuss the benefits and problems associated with each. Which one would you prefer to use?

(a) $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n x_i$.

(b) $\hat{\theta}_n = x_1$.

(c) $\hat{\theta}_n = 0$.

(a) $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n x_i$

Bias

$$\mathbb{E}(\hat{\theta}_n) - \theta = \mathbb{E}(\frac{1}{n}\sum_{i=1}^n x_i) - \mathbb{E}(\theta) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(x_i) - \mathbb{E}(\theta) = \mu - \mu = 0$$

Variance

We can split the varience over the sum because each x_i are i.i.d.

$$Var(\hat{\theta_n}) = Var(\frac{1}{n}\sum_{i=1}^n x_i) = \frac{1}{n^2}\sum_{i=1}^n Var(x_i) = \frac{\sigma^2}{n}$$

MSE

$$MSE(\hat{\theta_n}) = Var(\hat{\theta_n}) + Bias(\hat{\theta_n})^2$$
$$MSE(\hat{\theta_n}) = \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n}$$

(b) $\hat{\theta}_n = x_1$

Bias

$$\mathbb{E}(\hat{\theta}_n) - \theta = \mathbb{E}(x_i) - \mathbb{E}(X) = \mu - \mu = 0$$

Variance

$$Var(\hat{\theta_n}) = Var(x_i) = \sigma^2$$

MSE

$$MSE(\hat{\theta_n}) = \sigma^2 + 0 = \sigma^2$$

(c) $\hat{\theta}_n = 0$

Bias

$$\mathbb{E}(\hat{\theta}_n) - \theta = 0 - \mu = -\mu$$

Variance

$$Var(\hat{\theta_n}) = Var(0) = 0$$

MSE

$$MSE(\hat{\theta_n}) = 0 + (-\mu)^2 = \mu^2$$

Conclusion

(a) is the best of these methods as its MSE scales inversely with the amount of samples (n) collected from the random variable. Interestingly, if you the underlying random variable has an expected value of 0, then using $\hat{\theta}_n = 0$ provides the lowest MSE estimator. The problem with that is that you never know what the expected value of an underlying random phenomena is, so having something like part a is almost always better as you know that the MSE will converge to 0 the more samples collected.

$$\lim_{n\to\infty} \frac{\sigma^2}{n} = 0 \qquad \text{but} \lim_{n\to\infty} \mu^2 = \mu^2$$

So, if the expected value of the underlying random variable is not 0 (a) will be better than (b) in the limit, especially because (c) scales with the square of the expected value.

Examining (c), it is interesting and intuitive that (a) and (b) have the same MSE when we only have one sample; if you only have one sample to average over, it makes sense that the error of taking the average of the samples and just picking a single sample to be the estimator is the same. However, for any number of samples greater than 1, (a) is strictly better than (b), regardless of the properties of the underlying random variable. This makes sense, as in some way, taking the average of multiple samples will always remove some of the underlying randomness that comes with any individual sample. This is what the law of large numbers tells us more precisely and formally.

Overall, (a) is the best estimator as it is strictly better than (b) when given more than 1 sample, and is better than (c) in the limit, provided that the expected value of the underlying variable is not 0.