Part 1

1

Problem: Briefly, in your own words, why is it advantageous to know a statistic's sampling distribution?
Solution: I can think of one really good reason, comparison. Knowing how a statistic behaves allows us to look at different samples and compare them. If we're going to compare the sample mean weight of two breeds of dog, we can't confidently identify differences if we don't understand the properties of the sample mean like how much it varies. Hypothesis testing only works the way it does because of our knowledge of sampling distributions.

1.2

Problem: What is the distinction between a sampling distribution and the broader definition of a probability distribution?
Solution: The distinction here is that a sampling distribution is a type of probability distribution. Sampling distributions are specifically the probability distributions of statistics.

3

Problem: What is the distinction between the term standard error and the broader definition of standard deviation?
Solution: Surprisingly, the term **standard error** doesn't seem to be explicitly mentioned in Casella & Berger. Thankfully though it's defined in Jay L Devore Probability and Statistics for Engineering and the Sciences. According to that text, the standard error of an estimator is its standard deviation. That is, $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$ (Devore, 2012). So, in other words, it is the standard deviation of a statistic. While the standard deviation can measure the variability of a sample, the standard error measures the variation of statistic drawn from that sample. So the standard deviation of the sample mean for instance tells us how much the same mean fluctuates. It's used to help us understand how precise our estimator is. The standard error of the sample mean can help us understand how precise its estimation of the population mean is. Note: not how accurate is is, but precise.

Part 2

3

Complete the proof of Lemma 5.3.2

Lemma 5.3.2

We use the notation χ_p^2 to denote a chi squared random variable with p degrees of freedom.

- If Z is a $N(0,1)$ random variable, then, $Z^2 \sim \chi_1^2$. That is, the square of a standard normal random variable is a chi squared random variable.
- If $X_1, \cdots, X_n$ are independent, and $X_i \sim \chi_{p_i}^2$, then $X_1 + \cdots + X_n \sim \chi^2_{p_1+\ldots+p_n}$.

Solution

Thankfully we worked on this one in class!

A

Goal: If Z is a $N(0,1)$ random variable, then, $Z^2 \sim \chi_1^2$.

For these proofs moment generating functions are fantastic. Let's examine the mgf of a squared standard normal.

$$
\begin{aligned}
M_{Z^2}(t) &= E(e^{tz^2}) \\
&= \int e^{tz^2} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \\
&= \int \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{t}(1-2t)\right) dz \\
&= \frac{1}{\sqrt{1-2t}} \int \frac{\sqrt{1-2t}}{\sqrt{2\pi}} exp\left(-\frac{1-zt}{2}z^2\right) dz
\end{aligned}
$$

From here, the integrand is pdf of $N(0,(1-2t)^{-1})$. Our goal so far has been to rearrange that integral for that purpose. So, that integral evaluates to 1 as it's over all possible values for that pdf. Lastly, we are left with the value outside the integral.

$$M_{Z^2}(t) = (1-2t)^{-1/2}$$

which is the mgf of a χ^2 random variable.

B

Goal: If $X_1, \cdots, X_n$ are independent, and $X_i \sim \chi^2_{p_i}$, then $X_1 + \cdots + X_n \sim \chi^2_{p_1+\ldots+p_n}$

For the second proof we will proceed in a similar way, using moment generating functions.

The goal here will be to look at the behavior of the sum in the mgf and rearrange it so it's a form we can understand.

$$\begin{aligned}
M_{\sum x_i}(t) &= E[e^{t\sum x_i}] \\
&= E\left[\prod e^{tx_i}\right] && \text{(Properties of exponent)} \\
&= \prod E[e^{tx_i}] && \text{(X's are independent)} \\
&= \prod M_{X_i}(t) \\
&= \prod (1-2t)^{-p_i/2} \\
&= (1-2t)^{-\frac{1}{2}(p_1+\cdots+p_n)}
\end{aligned}$$

which is the mgf of a χ^2 with $p = p_1 + \cdots + p_n$.

5.3

Let $X_1, \cdots, X_n$ be iid rvs with continuous cdf F_X, and suppose $E[X_i] = \mu$. Define the rvs $Y_1, \cdots, Y_n$ by:

$$Y_i = \begin{cases} 1 & if\ X_i > \mu \\ 0 & if\ X_i \leq \mu \end{cases}$$

Find the distribution of $\sum Y_i$

Solution:

We can think of each Y_i as a bernoulli random variable where the probability of success is based on a value of X_i exceeding the average of X.

More formally:

$$Y_i \sim Bern(p = P(X_i > \mu))$$

We're trying to find the distribution of $\sum Y_i$ and thankfully that's not too bad. The sum of bernoulloi rvs is just a binomial. The probability parameter stays the same, but we can change X_i to X_1 because our rvs are all iid.

$$\sum Y_i \sim Bin(n, p = P(X_1 > \mu))$$

5.6

If X has pdf $f_X(x)$, and Y, independent of X, has pdf $f_Y(y)$, establish formulas for the random variable Z in each of the following situations.

A

Problem: $Z = X - Y$
Solution:

For all sections of this problem I will be using the general formula for bivariate transformations.

$$f_{u,v}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|J|$$

Where $|J|$ is the absolute value of the jacobian. This strategy just seems like the most straightforward way to accomplish this.

So let's rearrange the transformation a bit.

$$Z = X - Y \qquad Y = X - Z \qquad X = U$$
$$Y = U - Z$$

From here we can set up our Jacobian.

$$\begin{aligned} J &= \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dz} \\ \frac{dy}{du} & \frac{dy}{dz} \end{vmatrix} \\ &= \begin{vmatrix} 1 & 0 \\ 1 & -1 \end{vmatrix} \\ &= 1 \cdot -1 - 1 \cdot 0 \\ &= -1 \end{aligned}$$

From here we plug in our pieces.

$$\begin{aligned} f_{u,v}(u, v) &= f_{X,Y}(h_1(u, v), h_2(u, v))|J| \\ &= f_{X,Y}(f_X(u), f_Y(u - z))| - 1| \\ &= f_X(u) \cdot f_Y(u - z) && \text{(X and Y are independent)} \end{aligned}$$

To clairfy, we used the fact that X and Y are independent to split up the joint $f_{X,Y(x,y)}$ into $f_X(x) \cdot f_Y(y)$.

Now, we need the marginal distribution to finish.

$$f_Z(z) = \int f_X(u) f_Y(u-z) du$$

B

Problem: $Z = XY$
Solution:

Alright I'm gonna skip the details on this one.

$$Z = XY \qquad Y = \frac{Z}{X} \qquad X = U$$
$$Y = \frac{Z}{U}$$

$$\begin{aligned} J &= \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dz} \\ \frac{dy}{du} & \frac{dy}{dz} \end{vmatrix} \\ &= \begin{vmatrix} 1 & 0 \\ \frac{-z}{u^2} & \frac{1}{u} \end{vmatrix} \\ &= 1 \cdot \frac{1}{u} - \left(\frac{-z}{u^2} \cdot 0 \right) \\ &= \frac{1}{u} \end{aligned}$$

$$\begin{aligned} f_{U,Z}(u,z) &= f_X(u) f_Y\left(\frac{z}{u}\right) \cdot \left| \frac{1}{u} \right| \\ f_Z(z) &= \int f_X(u) f_Y\left(\frac{z}{u}\right) \cdot \frac{1}{u} du \end{aligned}$$

C

Problem: $Z = \frac{X}{Y}$
Solution:

$$Z = \frac{X}{Y} \qquad Y = \frac{X}{Z} \qquad X = U$$
$$Y = \frac{U}{Z}$$

$$\begin{aligned}
J &= \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dz} \\ \frac{dy}{du} & \frac{dy}{dz} \end{vmatrix} \\
&= \begin{vmatrix} 1 & 0 \\ \frac{1}{z} & \frac{-z}{u^2} \end{vmatrix} \\
&= 1 \cdot \frac{-u}{z^2} - \left(\frac{1}{z} \cdot 0 \right) \\
&= \frac{-u}{z^2}
\end{aligned}$$

$$\begin{aligned}
f_{U,Z}(u,z) &= f_X(u) f_Y\left(\frac{u}{z}\right) \cdot \left| \frac{u}{z^2} \right| \\
f_Z(z) &= \int f_X(u) f_Y\left(\frac{u}{z}\right) \cdot \frac{u}{z^2} du
\end{aligned}$$

5.11

Problem: Suppose $\bar{X}$ and S^2 are calculated from a random sample $X_1, \cdots, X_n$ drawn from a population with finite variance σ^2. We know that $E[S^2] = \sigma^2$. Prove that $E[S] \leq \sigma$, and, if $\sigma^2 > 0$, then $E[S] < \sigma$.

What does this say about the sample standard deviation as an estimator of the true standard deviation?

Setup:

Thoughts Before Proof:

I did not reach this solution organically. I initially assumed that this would work much like the proof for $E[S^2] = \sigma^2$. However I quickly found myself a bit stuck due to the square root. I suppose the big hint here is that the goal of the proof involves an inequality. Anyway, this proof will be leveraging Jensen's Inequality. Using this inequality was not my idea, but I will at least attempt to put in the work to convince myself of it. For this, we require some setup.

Theorems and Definitions:

Definition 4.7.6: A function $g(x)$ is **convex** if $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$ for all x and y, and $0 < \lambda < 1$. A function $g(x)$ is **concave** if $-g(x)$ is convex.

Alternative Convex Definition (From Probability by Ross): A twice-differentiable real-valued function $f(x)$ is said to be **convex** if $f''(x) \geq 0 \; \forall \; x$. Similarly, it is said to be **concave** if $f''(x) \leq 0 \; \forall \; x$.

Theorem 4.7.7 (Jensens Inequality)

For any random variable X, if $g(x)$ is a **convex** function then $E(g(X)) \geq g(E(X))$.

Equality holds if and only if, for every line $a + bx$ that is tangent to $g(x)$ at $x = E(X)$, P(g(X) = a+bX) = 1

Also, Jensen's Inequality applies to concave functions as well. If g is **concave** then $E[g(X)] \leq g(E[X])$. (From page 156 of Casella & Berger).

Note that I include both definitions of convex/concave because the first is from our book but I don't want to bother interpreting it. The other version from Sheldon Ross should be more than sufficient.

Solution Part 1

With all that out of the way, we can begin our solution to the first part of the proof.

For this problem we are examining $E[S] = E[\sqrt{S^2}]$.

Let $Y = S^2$ and $g(y) = \sqrt{y}$.

Using the alternative definition of convex we examine the second derivative of $f(y)$,

$$g''(y) = -\frac{1}{4}y^{-3/2}$$

From this, we note that

$$g''(y) \leq 0 \ \forall y$$

Therefore, $g(y)$ is concave.

From Theorem 4.7.7 we can use the alternative version of the inequality for concave functions.

$$\begin{aligned}
E[g(y)] &\leq g(E[y]) \\
E[\sqrt{S^2}] &\leq \sqrt{E[S^2]} && \text{(Substitue in values)} \\
&\leq \sqrt{\sigma^2} && \text{(Known expectation)} \\
&\leq \sigma && \text{(Value can't be negative due to the square)}
\end{aligned}$$

Completing the first portion of the proof.

Solution Part 2

Prove that if $\sigma^2 > 0$, then $E[S] < \sigma$.

For this part, our goal is to show that if σ^2 is non-zero, that $E[S]$ will never be equal to it. That is, the equality does not hold if σ^2 is non-zero. To show this we return back to the fundamentals.

For our random variable S,

$$\begin{aligned}
Var(S) &= E[S^2] - (E[S])^2 \\
Var(S) &= \sigma^2 - (E[S])^2 && \text{(Substitute known values)} \\
(E[S])^2 &= \sigma^2 - Var(S) \\
E[S] &= \sqrt{\sigma^2 - Var(S)}
\end{aligned}$$

From this equality, $E[S] = \sigma$ only if $Var(S) = 0$. Since we know that $\sigma^2 > 0$ so too must $Var(S) > 0$ as the population variance will now cause the sample to vary.

Therefore, returning back to our inequality, because $Var(S) > 0$,

$$E[\sqrt{S^2}] < \sigma$$

completing the proof.

Interpretation:

What this means is that the sample standard deviation will always underestimate the population standard deviation. More informally, it will always be an optimistic estimation of the population standard deviation.