Stat 346/446: Theoretical Statistics II: Homework 1 Solutions

Textbook Exercises

5.21 (346 & 446: 2 pts.) What is the probability that the larger of two continuous i.i.d. random variables will exceed the population median? Generalize this result to samples of size n.

Let X and Y be two i.i.d. continuous random variables with a population median m. We have

$$P(X \leq m) = P(X > m) = 0.5 \quad and \quad P(Y \leq m) = P(Y > m) = 0.5$$

The probability that the larger of two continuous i.i.d. random variables will exceed the population median can be written as:

$$P(\max(X, Y) > m) = 1 - P(\max(X, Y) \leq m).$$

Since $\max(X, Y) \leq m$ is equivalent to $X \leq m$ and $Y \leq m$, and the random variables are i.i.d., we have:

$$P(\max(X, Y) \leq m) = P(X \leq m)P(Y \leq m)$$

Thus

$$P(\max(X, Y) > m) = 1 - P(\max(X, Y) \leq m) = 1 - P(X \leq m)P(Y \leq m) = 1 - (\frac{1}{2})^2 = \frac{3}{4}$$

Generalize to a sample of size n:
For n i.i.d. continuous random variables $X_1, X_2, \ldots, X_n$,

$$\begin{aligned} P(\max(X_1, X_2, \ldots, X_n) > m) &= 1 - P(\max(X_1, X_2, \ldots, X_n) \leq m) \\ &= 1 - P(X_1 \leq m) \cdot P(X_2 \leq m) \cdots P(X_n \leq m) \\ &= 1 - \left(\frac{1}{2}\right)^n. \end{aligned}$$

5.18 (346 : 3 pts, 446 : 5 pts.) Let X be a random variable with a Student's t distribution with p degrees of freedom.

(a) (346 & 446: 2 pts.) Derive the mean and variance of X.

Let $X \sim t_p$, and $X = Z/\sqrt{V/p}$, where: $Z \sim N(0, 1)$ is a standard normal distribution, $V \sim \chi^2_p$ is a chi-squared distribution with p degrees of freedom, and Z and V are independent. Since $E[Z] = 0$ and the denominator $\sqrt{V/p}$ is positive and finite for $p > 1$, the expectation simplifies to:

$$E[X] = E\left[\frac{Z}{\sqrt{V/p}}\right] = \frac{1}{\sqrt{p}} E[Z] \cdot E\left[\frac{1}{\sqrt{V}}\right].$$

As $E[Z] = 0$, the product becomes:

$$E[X] = 0 \quad (\text{when } p > 1).$$

The variance of X is:

$$\text{Var}(X) = E[X^2] - (E[X])^2 = E[X^2].$$

Using $X^2 = Z^2/(V/p)$:
Since $Z^2 \sim \chi_1^2$ and $V/p \sim \frac{\chi_p^2}{p}$, it follows that X^2 is the ratio of two independent chi-squared distributions, scaled by their degrees of freedom. Hence:

$$X^2 \sim F(1,p).$$

The mean of an $F(1,p)$ distribution is given by:

$$E[X^2] = \frac{p}{p-2}, \quad \text{for } p > 2.$$

Thus:

$$\mathrm{Var}(X) = \frac{p}{p-2}, \quad (\text{when } p > 2).$$

(b) (346 & 446: 1 pts.) Show that X^2 has an F distribution with 1 and p degrees of freedom.

As shown in part (a), $X^2 = Z^2/(V/p)$:
Since $Z^2 \sim \chi_1^2$ and $V/p \sim \frac{\chi_p^2}{p}$, it follows that X^2 is the ratio of two independent chi-squared distributions, scaled by their degrees of freedom. we derived that:

$$X^2 \sim F(1,p).$$

(c) (446: 2 pts.)Let $f(x|p)$ denote the pdf of X. Show that

$$\lim_{p\to\infty} f(x|p) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

at each value of x, $-\infty < x < \infty$. This correctly suggests that as $p \to \infty$, X converges in distribution to a $n(0,1)$ random variable. *(Hint: Use Stirling's Formula.)*

The PDF of t_p is:

$$f(x|p) = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\sqrt{p\pi}} \left(1 + \frac{x^2}{p}\right)^{-\frac{p+1}{2}}.$$

The constant term is:

$$C_p = \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\sqrt{p\pi}}.$$

Using Stirling's formula:

$$\Gamma(n) \sim \sqrt{2\pi}(n-1)^{n-\frac{1}{2}} e^{-n+1},$$

for large p:

$$\Gamma\left(\frac{p+1}{2}\right) \sim \sqrt{2\pi}\left(\frac{p-1}{2}\right)^{\frac{p}{2}} e^{-\frac{p-1}{2}},$$

$$\Gamma\left(\frac{p}{2}\right) \sim \sqrt{2\pi}\left(\frac{p-2}{2}\right)^{\frac{p}{2}-\frac{1}{2}} e^{-\frac{p-2}{2}}.$$

Substituting into C_p:

$$C_p \sim \frac{\sqrt{2\pi}\left(\frac{p-1}{2}\right)^{\frac{p}{2}} e^{-\frac{p-1}{2}}}{\sqrt{2\pi}\left(\frac{p-2}{2}\right)^{\frac{p}{2}-\frac{1}{2}} e^{-\frac{p-2}{2}}\sqrt{p\pi}}.$$

$$\begin{aligned}
\frac{\sqrt{2\pi}\left(\frac{p-1}{2}\right)^{\frac{p}{2}} e^{-\frac{p-1}{2}}}{\sqrt{2\pi}\left(\frac{p-2}{2}\right)^{\frac{p}{2}-\frac{1}{2}} e^{-\frac{p-2}{2}}\sqrt{p\pi}} &= \frac{\left(\frac{p-1}{2}\right)^{\frac{p}{2}} e^{-\frac{p-1}{2}}}{\left(\frac{p-2}{2}\right)^{\frac{p}{2}-\frac{1}{2}} e^{-\frac{p-2}{2}}\sqrt{p\pi}} \\
&= \frac{\left(\frac{p-2}{2}\cdot\frac{p-1}{p-2}\right)^{\frac{p}{2}}}{\left(\frac{p-2}{2}\right)^{\frac{p}{2}-\frac{1}{2}}\sqrt{p\pi}} e^{-\frac{1}{2}} \\
&= \frac{\left(\frac{p-2}{2}\right)^{\frac{p}{2}}\left(\frac{p-1}{p-2}\right)^{\frac{p}{2}}}{\left(\frac{p-2}{2}\right)^{\frac{p}{2}-\frac{1}{2}}\sqrt{p\pi}} e^{-\frac{1}{2}} \\
&= \frac{\left(\frac{p-2}{2}\right)^{\frac{1}{2}}\cdot\left(\frac{p-1}{p-2}\right)^{\frac{p}{2}}}{\sqrt{p\pi}} e^{-\frac{1}{2}}. \\
&= \frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}}\cdot\left(\frac{p-2}{p}\right)^{\frac{1}{2}}\cdot\left(\frac{p-1}{p-2}\right)^{\frac{p}{2}}
\end{aligned}$$

As $p \to \infty$, we have:

$$\left(\frac{(p-2)}{p}\right)^{\frac{1}{2}} \to 1,$$

and

$$\lim_{p\to\infty}\left(\frac{p-1}{p-2}\right)^{\frac{p}{2}} = \lim_{p\to\infty}\left(1+\frac{1}{p-2}\right)^{\frac{p-2}{2}+1} = e^{\frac{1}{2}}.$$

Therefore:

$$\lim_{p\to\infty}\frac{e^{-\frac{1}{2}}}{\sqrt{2\pi}}\cdot\left(\frac{p-2}{p}\right)^{\frac{1}{2}}\cdot\left(\frac{p-1}{p-2}\right)^{\frac{p}{2}} = \frac{e^{\frac{1}{2}-\frac{1}{2}}}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}}.$$

Simplify to get:

$$\lim_{p\to\infty} C_p = \frac{1}{\sqrt{2\pi}}.$$

For the term $\left(1+\frac{x^2}{p}\right)^{-\frac{p+1}{2}}$, using $\ln(1+u) \sim u$ for small u:

$$\left(1+\frac{x^2}{p}\right)^{-\frac{p+1}{2}} \sim e^{-\frac{p+1}{2}\cdot\frac{x^2}{p}} = e^{-x^2/2}.$$

Therefore:

$$\lim_{p\to\infty} f(x|p) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

This is the PDF of $N(0,1)$, so t_p converges to $N(0,1)$ as $p \to \infty$.

Extra Problems

1. (346 : 5 pts, 446 : 3 pts.) Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu_1, \sigma^2)$ and $Y_1, Y_2, \ldots, Y_m$ be a random sample from $N(\mu_2, \sigma^2)$. Also assume that the random vectors $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_m)$ are mutually independent. Notice that the two populations have the same variance, but different means.

(a) (346 : 3 pts, 446 : 2 pts.) Let

$$S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{i=1}^m (Y_i - \overline{Y})^2}{n+m-2}.$$

Show that

$$\frac{(n+m-2)S^2}{\sigma^2} \sim \chi^2_{n+m-2} \quad \text{and} \quad E(S^2) = \sigma^2.$$

Let $X_i \overset{iid}{\sim} N(\mu_1, \sigma^2)$, $i = 1, \ldots, n$ and $Y_i \overset{iid}{\sim} N(\mu_2, \sigma^2)$, $i = 1, \ldots, m$. Assume that X_i and Y_i are independent.
According to Theorem 5.3.1:

$$\frac{(n-1)S_X^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{and} \quad \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2_{m-1}.$$

Since X_i and Y_i are independent, the sample variances S_X^2 and S_Y^2 are independent. Hence:

$$\frac{(n-1)S_X^2}{\sigma^2} \text{ and } \frac{(m-1)S_Y^2}{\sigma^2} \text{ are independent.}$$

Using Lemma 5.3.2:

$$\frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2_{(n-1)+(m-1)}.$$

$$\frac{(n-1)S_X^2}{\sigma^2} + \frac{(m-1)S_Y^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

By the definition of S^2:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{i=1}^m (Y_i - \overline{Y})^2}{n+m-2}.$$

Substituting into the equation:

$$\frac{(n+m-2)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{i=1}^m (Y_i - \overline{Y})^2}{\sigma^2}.$$

Using the earlier result:

$$\frac{(n+m-2)S^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

We calculate:

$$\begin{aligned}
E\left[(n+m-2)S^2\right] &= E\left[\sum_{i=1}^n (X_i - \overline{X})^2 + \sum_{i=1}^m (Y_i - \overline{Y})^2\right] \\
&= E\left[\sum_{i=1}^n X_i^2 - 2\overline{X}\sum_{i=1}^n X_i + n\overline{X}^2\right] + E\left[\sum_{i=1}^m Y_i^2 - 2\overline{Y}\sum_{i=1}^m Y_i + m\overline{Y}^2\right] \\
&= \sum_{i=1}^n E[X_i^2] - nE[\overline{X}^2] + \sum_{i=1}^m E[Y_i^2] - mE[\overline{Y}^2] \\
&= \sum_{i=1}^n \left[\sigma^2 + \mu_1^2\right] - nE[\overline{X}^2] + \sum_{i=1}^m \left[\sigma^2 + \mu_2^2\right] - mE[\overline{Y}^2] \\
&= n\left(\sigma^2 + \mu_1^2\right) - n\left(\frac{\sigma^2}{n} + \mu_1^2\right) + m\left(\sigma^2 + \mu_2^2\right) - m\left(\frac{\sigma^2}{m} + \mu_2^2\right) \\
&= n\sigma^2 + n\mu_1^2 - \sigma^2 - n\mu_1^2 + m\sigma^2 + m\mu_2^2 - \sigma^2 - m\mu_2^2 \\
&= (n+m-2)\sigma^2.
\end{aligned}$$

$$\Rightarrow E[S^2] = \frac{1}{n+m-2} E\left[(n+m-2)S^2\right] = \sigma^2.$$

$$\text{Thus, } E[S^2] = \sigma^2.$$

(b) (346 : 2 pts, 446 : 1 pts.) Let

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

where $S = \sqrt{S^2}$ from part (a). Show that $T \sim t_{n+m-2}$.

We already showed in part (a) that:

$$\frac{(n+m-2)S^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

We also have:

$$\overline{X} \sim N\left(\mu_1, \frac{\sigma^2}{n}\right), \quad \overline{Y} \sim N\left(\mu_2, \frac{\sigma^2}{m}\right).$$

Since $\overline{X}$ and $\overline{Y}$ are independent as functions of X_i's and Y_i's (which are independent), we conclude:

$$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right).$$

Thus:

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim N(0,1).$$

According to the definition of t distribution:

- $Z = \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}} \sim N(0,1)$,
- $V = \frac{(n+m-2)S^2}{\sigma^2} \sim \chi^2_{n+m-2}$.
- $\overline{X}$ and S^2_X are independent, $\quad \overline{Y}$ and S^2_Y are independent

Using Theorem 5.3.1, $\overline{X}$ and $\sum_{i=1}^n (X_i - \overline{X})^2$ are independent. Similarly:

$$\overline{Y} \text{ and } \sum_{i=1}^m (Y_i - \overline{Y})^2 \text{ are independent.}$$

Since X_i's and Y_i's are independent, $\overline{X}$ and $\overline{Y}$ are independent, and:

$$\overline{X} \text{ and } S^2 \text{ are independent.}$$

Thus, Z and V are independent. Using the definition of the t-distribution:

$$\frac{Z}{\sqrt{\frac{V}{n+m-2}}} \sim t_{n+m-2}.$$

Substitute Z and V:

$$\frac{\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{n}+\frac{\sigma^2}{m}}}}{\sqrt{\frac{\frac{(n+m-2)S^2}{\sigma^2}}{n+m-2}}} \sim t_{n+m-2}.$$

Simplify:

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}.$$

We have shown that:

$$T = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}.$$