ECON 6190

Problem Set 3

Gabe Sekeres

September 30, 2024

Problem 1. Let $\bar{X}_n$ and s_n^2 be the sample mean and variances. Suppose another observation X_{n+1} becomes available.

(a) We have that

$$\bar{X}_{n+1} = \frac{\sum_{i=1}^{n+1} X_i}{n+1} = \frac{X_{n+1} + \sum_{i=1}^{n} X_i}{n+1} = \frac{X_{n+1} + n\frac{\sum_{i=1}^{n} X_i}{n}}{n+1} = \frac{X_{n+1} + n\bar{X}_n}{n+1}$$

(b) We have that

$$\begin{aligned}
ns_{n-1}^2 &= n\left(\frac{1}{n}\sum_{i=1}^{n+1}(X_i - \bar{X}_{n+1})^2\right) \\
&= \sum_{i=1}^{n+1}(X_i - \bar{X}_n + \bar{X}_n - \bar{X}_{n+1})^2 \\
&= \sum_{i=1}^{n+1}(X_i - \bar{X}_n)^2 - (n+1)(\bar{X}_n - \bar{X}_{n+1})^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + (X_{n+1} - \bar{X}_n)^2 - (n+1)\left(\bar{X}_n - \frac{X_{n+1} + n\bar{X}_n}{n+1}\right)^2 \\
&= (n-1)s_n^2 + (X_{n+1} - \bar{X}_n)^2 - (n+1)\left(\frac{\bar{X}_n - X_{n+1}}{n+1}\right)^2 \\
&= (n-1)s_n^2 + \frac{n}{n+1}(X_{n+1} - \bar{X}_n)^2
\end{aligned}$$

Problem 2. Find the distributions of:

(a) $(\bar{X}_n - \bar{Y}_n)/\sqrt{2\sigma^2/n}$: We have that

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}} = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i - Y_i}{\sqrt{\frac{2\sigma^2}{n}}} = \frac{1}{\sqrt{n}}\frac{1}{\sigma\sqrt{2}}\sum_{i=1}^{n} X_i - Y_i$$

Thus, the mean is

$$\mathbb{E}\left[\frac{1}{\sqrt{n}}\frac{1}{\sigma\sqrt{2}}\sum_{i=1}^{n} X_i - Y_i\right] = \frac{1}{\sqrt{n}}\frac{1}{\sigma\sqrt{2}}\sum_{i=1}^{n}\mathbb{E}[X_i] - \mathbb{E}[Y_i] = \frac{1}{\sqrt{n}}\frac{1}{\sigma\sqrt{2}}\sum_{i=1}^{n}(\mu - \mu) = 0$$

and the variance is

$$var\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma\sqrt{2}}\sum_{i=1}^{n} X_i - Y_i\right) = \frac{1}{2n\sigma^2}\sum_{i=1}^{n} var(X_i) + var(Y_i) = \frac{1}{2n\sigma^2}\sum_{i=1}^{n} 2\sigma^2 = 1$$

Since X and Y are iid normal and mutually independent, we can say that

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}} \sim \mathcal{N}(0,1)$$

(b) $(\bar{X}_n - \bar{Y}_n)/\sqrt{2s_X^2/n}$: We have that from a theorem in class, since X is iid normal, that $\frac{ns_X^2}{\sigma^2} \sim \chi_n^2$. This means that our distribution is

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2s_X^2}{n}}} = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}} \cdot \frac{1}{\sqrt{\frac{ns_X^2}{\sigma^2}\frac{1}{\sqrt{n}}}} \sim \frac{\mathcal{N}(0,1)}{\frac{\sqrt{\chi_{n-1}^2}}{\sqrt{n-1}}} \sim t_{n-1}$$

A student's t distribution with n degrees of freedom.

(c) $(\bar{X}_n - \bar{Y}_n)/\sqrt{2s_Y^2/n}$: Similarly as (b), since Y is iid normal, we have that $\frac{ns_Y^2}{\sigma^2} \sim \chi_n^2$. This means that our distribution is

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2s_Y^2}{n}}} = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}} \cdot \frac{1}{\sqrt{\frac{ns_Y^2}{\sigma^2}\frac{1}{\sqrt{n}}}} \sim \frac{\mathcal{N}(0,1)}{\frac{\sqrt{\chi_{n-1}^2}}{\sqrt{n-1}}} \sim t_{n-1}$$

A student's t distribution with n degrees of freedom.

(d) $(\bar{X}_n - \bar{Y}_n)/\sqrt{(s_X^2 + s_Y^2)/n}$: We have that since X and Y are each normal, we have that $\frac{ns_X^2}{\sigma^2}, \frac{ns_Y^2}{\sigma^2} \sim \chi_n^2 \Rightarrow \frac{ns_X^2}{\sigma^2} + \frac{ns_Y^2}{\sigma^2} \sim 2\chi_n^2$. This means that the distribution is

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{s_X^2+s_Y^2}{n}}} = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}} \cdot \frac{1}{\sqrt{\frac{n(s_X^2+s_Y^2)}{\sigma^2}\frac{1}{\sqrt{n}}}} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{2\chi_{n-1}^2}{n}}} \sim t_{2n-2}$$

A student's t distribution with $2n-2$ degrees of freedom.

(e) $(\bar{X}_n - \bar{Y}_n)/\sqrt{s_n^2/n}$: Since $Z \sim \mathcal{N}(0, 2\sigma^2)$, since X and Y are mutually independent, we have that $\frac{ns_n^2}{\sigma^2} \sim \chi_n^2$. This means that the distribution is

$$\frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{s_n^2}{n}}} = \frac{\bar{X}_n - \bar{Y}_n}{\sqrt{\frac{2\sigma^2}{n}}} \cdot \frac{1}{\sqrt{\frac{ns_n^2}{\sigma^2}\frac{1}{\sqrt{2n}}}} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} \sim t_{n-1}$$

A student's t distribution with $n-1$ degrees of freedom.

Problem 3. Use X_1, X_2, X_3 to construct a statistic with the following distributions:

(a) Chi-square distribution with 3 degrees of freedom. Note that since $X_i \sim \mathcal{N}(i, i^2)$, we have that $\frac{X_i - i}{i} \sim \mathcal{N}(0,1)$. Then if $Z_i \sim \mathcal{N}(0,1)$, we have that

$$\sum_{i=1}^{3} Z_i = (X_1 - 1) + \frac{X_2 - 2}{2} + \frac{X_3 - 3}{3} = X_1 + \frac{X_2}{2} + \frac{X_3}{3} - 3 \sim \chi_3^2$$

From the definition of chi-square distributions

(b) t distribution with 2 degrees of freedom. From above, we have that $\frac{X_i - i}{i} \sim \mathcal{N}(0,1)$. Then from the definition of the t distribution, and defining $Z_i \sim \mathcal{N}(0,1)$, we have that

$$\frac{\frac{1}{3}\left((X_1 - 1) + \frac{X_2-2}{2} + \frac{X_3-3}{3}\right)}{\frac{1}{\sqrt{3}}\sqrt{\frac{1}{2}\left[\left(-\frac{X_2-2}{2} - \frac{X_3-3}{3}\right)^2 + \left(-(X_1-1) - \frac{X_3-3}{3}\right)^2 + \left(-(X_1-1) - \frac{X_2-2}{2}\right)^2\right]}} = \frac{\bar{Z} - 0}{\frac{s_Z}{\sqrt{3}}} \sim t_2$$

Problem 4. Show that $Y = \min\{X_1, \dots, X_n\}$ is a sufficient statistic for θ, where $f(x \mid \theta) = e^{-(x-\theta)}\mathbb{1}\{x \geq \theta\}$.

Proof. We have that

$$f_X(x \mid \theta) = \begin{cases} \prod_{i=1}^n e^{-(x-\theta)} & X_i \geq \theta \; \forall \; i = 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

Thus, if $\exists \; i$ s.t. $X_i < \theta$, $\frac{f_X(x|\theta)}{f_Y(y)} = 0$ which is not dependent on θ. Thus, we consider the case where $X_i \geq \theta \; \forall \; i = 1, \dots, n$. We have that

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y}\mathbb{P}\{Y \leq y\} \\ \mathbb{P}\{Y \leq y\} &= 1 - \prod_{i=1}^n \mathbb{P}\{X_i > y\} \\ &= 1 - \mathbb{1}_{Y \geq \theta} \prod_{i=1}^n \int_y^\infty e^{-(x-\theta)} dx \\ \text{and since we assume that } \min_i X_i \geq \theta, \quad &= 1 - \prod_{i=1}^n e^{-(y-\theta)} \\ &= 1 - e^{-n(y-\theta)} \end{aligned}$$

Thus, we have that

$$f_Y(y) = \frac{\partial}{\partial y}\left[1 - e^{-n(y-\theta)}\right] = ne^{-n(y-\theta)}$$

And thus,

$$\frac{f_X(x \mid \theta)}{f_Y(y)} = \frac{\prod_{i=1}^n e^{-(x-\theta)}}{ne^{-n(y-\theta)}} = \frac{e^{-n(x-\theta)}}{ne^{-n(y-\theta)}} = \frac{e^{-nx}}{ne^{-ny}}$$

Since this does not depend on θ, $Y = \min_i X_i$ is a sufficient statistic for θ. □

Problem 5. Show that $\min_i \frac{X_i}{i}$ is a sufficient statistic.

Proof. We will use the Factorization Theorem. Note that

$$f_X(x \mid \theta) = \begin{cases} \prod_{i=1}^n e^{i\theta - x} & \min_i\{X_i\} \geq i\theta \\ 0 & \text{otherwise} \end{cases}$$

Defining an indicator function, we get that

$$f_X(x \mid \theta) = \mathbb{1}_{\min_i \frac{X_i}{i} \geq \theta} \prod_{i=1}^n e^{i\theta - x} = \mathbb{1}_{\min_i \frac{X_i}{i} \geq \theta} e^{\sum_{i=1}^n i\theta - nx} = \mathbb{1}_{\min_i \frac{X_i}{i} \geq \theta} e^{\sum_{i=1}^n i\theta} e^{-nx}$$

Then we can define

$$h(x) = e^{-nx}$$

and

$$g\left(\min_i \frac{X_i}{i} \middle| \theta\right) = \mathbb{1}_{\min_i \frac{X_i}{i} \geq \theta} e^{\sum_{i=1}^n i\theta}$$

And since $f_X(x \mid \theta) = h(x) g\left(\min_i \frac{X_i}{i} \middle| \theta\right)$, $\min_i \frac{X_i}{i}$ is a sufficient statistic. □

Problem 6. Show that any one-to-one function of a sufficient statistic is also a sufficient statistic.

Proof. We have that a statistic $T(x)$ is sufficient, meaning that there exist functions $h(x)$ and $g(T(X) \mid \theta)$ such that

$$f_X(x \mid \theta) = h(x)g(T(X) \mid \theta)$$

If there exists $T'(x)$ such that $T'(x) = f(T(X))$ for some bijective f, then we can say that since f is bijective and invertible, $T(X) = f^{-1}(T'(X))$

$$f_X(x \mid \theta) = h(x)g(f^{-1}(T'(X) \mid \theta)$$

so defining $g^\star := g \circ f^{-1}$, we get that

$$f_X(x \mid \theta) = h(x)g^\star(T'(X) \mid \theta)$$

and since the conditions of the Factorization Theorem hold, T' is also a sufficient statistic. $\square$

Problem 7. The distribution of $\mathcal{N}(0, \sigma^2)$ is

$$\phi_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

Thus, since every observation x is squared, meaning that $x^2 = |X|^2$, we can simply set $x = |X|$, and have

$$g(|X| \mid \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{|X|^2}{2\sigma^2}}$$

and

$$h(x) = -e^{i\pi} = 1$$

so $\phi_X(x) = g(|X| \mid \sigma^2)h(x)$ and $|X|$ is sufficient by the Factorization Theorem.