Problem Set #4

Danny Edgel
Econ 709: Economic Statistics and Econometrics I
Fall 2020

October 3, 2020

Collaborated with Sarah Bass, Emily Case, Michael Nattinger, and Alex Von Hafften

Question 1

Suppose that another observation X_{n+1} becomes available. Show that:

(a) $\mathbf{\overline{X}_{n+1} = (n\overline{X}_n + X_{n+1})/(n+1)}$

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

(b) $\mathbf{s^2_{n+1} = \frac{1}{n}((n-1)s^2_n + (n/(n+1))(X_{n+1} - \overline{X}_n)^2)}$

Using the relation from (a), we can derive:

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

Question 2

For some integer k, set $\mu_k = E(X^k)$. Construct an unbiased estimator $\hat{\mu}_k$ for μ_k, and show its unbiasedness.

Define $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$. If the bias of this estimator is equal to zero, then it is unbiased:

$$E(\hat{\mu}_k) - \mu_k = 0$$

$$E(\frac{1}{n}\sum_{i=1}^{n} X_i^k) - E(X^k) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = X^k$$

Since $\{X_i\}_{i=1}^{n}$ is assumed to be a random sample and X is assumed to be i.i.d., $E(X_i^k) = E(X^k)$,[1] so this equality holds. Thus, $\hat{\mu}_k$ is an unbiased estimator.

Question 3

Consider the central moment $m_k = E((X-\mu)^k)$. Construct an estimator $\hat{m}_k$ for m_k without assuming a known μ. In general, do you expect $\hat{m}_k$ to be biased or unbiased?

Let $\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^m$, where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. In general, I expect this esimator to be be biased. To see why, take $\hat{m}_2$. From the lecture, we know that $\hat{m}_2 = \sigma_n^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2 - (\overline{X}_n - \mu)^2$ with the known exact bias $\frac{1}{n}\sigma_X^2$. We could correct for this downward bias, but the higher-order central moment will differ non-proportionally. We cannot derive a general, unbiased estimator for $m_k = E((X-\mu)^k)$.

Question 4

Calculate the variance of $\hat{\mu}_k$ that you proposed above, and call it $Var(\hat{\mu}_k)$.

The variance of any analog estimator, $\hat{a}_i$ is calculated as $\frac{1}{n^2}\sum_{i=1}^{n} Var(\hat{a}_i)$. Thus, we can derive:

$$Var(\hat{\mu}_k) = \frac{1}{n^2}\sum_{i=1}^{n} Var(\hat{\mu}_k) = \frac{1}{n} Var(x_i^k) = \frac{1}{n}\left(E(X_i^2 k) - E(X_i^k)\right) = \frac{1}{n}(\mu_{2k} - \mu_k)$$

[1] This is because X_i and X_j are independent $\forall i \neq j$, so $E(X_i X_j) = E(X_i)E(X_j)$.

Question 5

Show that $E(s_n) \leq \sigma$ using Jensen's inequality (CB Theorem 4.7.7).

According to Jensen's inequality, if g is a convex function, then $E[g(x)] \geq g(E[x])$. Since S_n^2 is an unbiased estimator of σ^2, $E(S_n^2) = \sigma^2$. Further, $\sqrt{\sigma^2} = \sigma$. Note that the $f(x) = \sqrt{x}$ is a concave function, so $g(x) = -f(x)$ is a convex function. Then,

$$\begin{aligned} E\left[-\sqrt{s_n^2}\right] &\geq -\sqrt{E(s_n^2)} \\ -E\left[s_n\right] &\geq -\sqrt{\sigma^2} \\ E\left[s_n\right] &\leq \sigma \end{aligned}$$

Question 6

Show algebraically that $\hat{\sigma}^2 = n^{-1}\sum_{i-1}^{n}(X_i - \mu)^2 - (\overline{X}_n - \mu)^2$.

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - 2\overline{X}_n\frac{1}{n}\sum_{i=1}^{n}X_i + \frac{1}{n}\sum_{i=1}^{n}\overline{X}_n^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - 2\overline{X}_n^2 + \overline{X}_n^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}_n^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}X_i^2 - 2\mu\overline{X}_n + \mu^2 - (\overline{X}_n^2 - 2\mu\overline{X}_n + \mu^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2\mu X_i + \mu^2) - (\overline{X}_n - \mu)^2 \\ \hat{\sigma}^2 &= n^{-1}\sum_{i-1}^{n}(X_i - \mu)^2 - (\overline{X}_n - \mu)^2 \end{aligned}$$

Question 7

Find the covariance of $\hat{\sigma}^2$ and $\overline{X}_n$. Under what condition is this zero? (See lecture question for hint)

From the covariance definition, we can solve:

$$\begin{aligned}
Cov(\hat{\sigma^2}, \overline{X}_n) &= E\left[(\hat{\sigma^2} - E(\hat{\sigma^2}))(\overline{X}_n - E(\overline{X}_n))\right] \\
&= E\left[\hat{\sigma^2}(\overline{X}_n - \mu)\right] - E(\hat{\sigma^2})E\left[\overline{X}_n - \mu\right] \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\overline{X}_n-\mu)^2\right)(\overline{X}_n - \mu\right] - \hat{\sigma^2}(\mu-\mu) \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2(\overline{X}_n - \mu - (\overline{X}_n-\mu)^3\right)\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2(\overline{X}_n - \mu\right] - E\left[(\overline{X}_n-\mu)^3\right]
\end{aligned}$$

Where, since $\{X_i\}_{i=1}^n$ are independent.:

$$\begin{aligned}
E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2(\overline{X}_n - \mu)\right] &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)\right)\right] \\
&= \frac{1}{n^2}E\left[\sum_{i=1}^{n}(X_i-\mu)^3\right] + 2\frac{1}{n^2}E\left[\sum_{i\neq j}^{n}(X_i-\mu)(X_j-\mu)\right] \\
&= \frac{1}{n}E\left[(X_i-\mu)^3\right]
\end{aligned}$$

And:

$$\begin{aligned}
E\left[(\overline{X}_n-\mu)^3\right] &= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right)^3\right] \\
&= \frac{1}{n^3}E\left[\left(\left(\sum_{i=1}^{n}(X_i-\mu)\right)^2 + 2\sum_{i\neq j}^{n}(X_i-\mu)(X_j-\mu)\right)\left(\sum_{i=1}^{n}(X_i-\mu)\right)\right] \\
&= \frac{1}{n^3}E\left[\left(\sum_{i=1}^{n}(X_i-\mu)\right)^3\right] + E\left[\sum_{i\neq j}^{n}(X_i-\mu)(X_j-\mu)\right] \\
&+ E\left[2\sum_{i\neq j}^{n}(X_i-\mu)(X_j-\mu) + 3\sum_{i\neq j\neq k}^{n}(X_i-\mu)(X_j-\mu)(X_k-\mu)\right] \\
&= \frac{1}{n^3}E\left[\left(\sum_{i=1}^{n}(X_i-\mu)\right)^3\right] \\
&= \frac{1}{n^2}E\left[(X_i-\mu)^3\right]
\end{aligned}$$

Taken together,

$$Cov(\hat{\sigma^2}, \overline{X}_n) = \left(\frac{1}{n} - \frac{1}{n^2}\right) E[(X_i - \mu)^3]$$

Thus, this covariance is zero if $E[(X_i - \mu)^3] = 0$, which is if the distribution of X has no skewness.

Question 8

Suppose that X_i are independent but not necessarily identically distributed (i.n.i.d.) with $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$.

(a) **Find $E(\overline{X}_n)$.**

$$E[\overline{X}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu_i$$

(b) **Find $Var(\overline{X}_n)$.**

$$\begin{aligned}
Var(\overline{X}_n) &= E\left[\overline{X}_n^2\right] - \left(E[\overline{X}_n]\right)^2 \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] - \left(\frac{1}{n}\sum_{i=1}^{n}\mu_i\right)^2 \\
&= \frac{1}{n^2}E\left[\sum_{i=1}^{n} X_i^2 + 2\sum_{i\neq j}^{n} X_i X_j\right] - \frac{1}{n^2}\left(\sum_{i=1}^{n}\mu_i^2 - 2\sum_{i\neq j}^{n}\mu_i\mu_j\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n}(E[X_i^2] - \mu_i^2)\right) + \frac{2}{n^2}\sum_{i\neq j}^{n}(E[X_i]E[X_j] - \mu_i\mu_j) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n} Var(X_i)\right) + \frac{2}{n^2}\sum_{i\neq j}^{n}(\mu_i\mu_j - \mu_i\mu_j) \\
Var(\overline{X}_n) &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2
\end{aligned}$$

Question 9

Show that if $Q \sim \chi^2_r$, then $E(Q) = r$ and $Var(Q) = 2r$ (hint: use the representation $Q = \sum_{i=1}^n X_i^2$ with X_i being i.i.d $\mathcal{N}(0,1)$).

$$E[Q] = E\left[\sum_{i=1}^r X_i^2\right] = \sum_{i=1}^r E[X_i^2] = \sum_{i=1}^r (\sigma_x^2 + \mu_x^2) = \sum_{i=1}^r (1) = r$$

$$\begin{aligned} Var(Q) = E[Q^2] - (E[Q])^2 &= E\left[\left(\sum_{i=1}^r X_i^2\right)^2\right] - r^2 \\ &= E\left[\sum_{i=1}^r X_i^4 + 2\sum_{i \neq j}^r X_i^2 X_j^2\right] - r^2 \\ &= \sum_{i=1}^r E\left[X_i^4\right] + 2\sum_{i \neq j}^r E[X_i^2]E[X_j^2] - r^2 \end{aligned}$$

Notice that $E\left[X_i^4\right]$ is the fourth moment of X_i, which is normally distributed with mean zero and variance one, and that $\sum_{i \neq j}^r E[X_i^2]E[X_j^2]$ is the number of combinations between two groups of r items, without replacement. Thus,

$$Var(Q) = \sum_{i=1}^r (3) + 2\left(\frac{r!}{2!(r-2)!}\right) - r^2 = 3r - r(r-1) - r^2 = 3r + r^2 - r - r^2 = 2r$$

Question 10

Suppose that $X_i \sim \mathcal{N}(\mu_X, \sigma_X^2) : i = 1, ..., n_1$ and $Y_i \sim \mathcal{N}(\mu_Y, \sigma_Y^2), i = 1, ..., n_2$ are mutually independent. Set $\overline{X}_n = n_1^{-1} \sum_{i=1}^{n_1} X_i$ and $\overline{X}_n = n_2^{-1} \sum_{i=1}^{n_2} Y_i$.

First, I will show that the sum of any set of independent, normally-distributed random variables is itself a normally-distributed random variable. Suppose that $X_1, X_2, ..., X_n$ are independent, normal random variables, where $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ for all $i \in \{1, ..., n\}$. Then the moment-generating function of their sum is:

$$M_{\sum X_i}(t) = E\left[e^{t(\sum X_i)}\right] = E\left[\prod_{i=1}^n e^{tX_i}\right] = \prod_{i=1}^n M_{X_i}(t)$$

Since the linear transformation of any normal random variable is also a normal random variable, $\overline{X}_n$ and $\overline{X}_y$ are normal random variables. Thus, $\overline{X}_n - \overline{Y}_n$ is also a normal random variable with the MGF:

(a) **Find $E(\overline{X}_n - \overline{Y}_n)$.**

(b) **Find** $Var(\overline{X}_n - \overline{Y}_n)$**.**

(c) **Find the distribution of** $\overline{X}_n - \overline{Y}_n$**.**