ECON 709 - PS 4

Alex von Hafften*

10/4/2020

Most of the problems assume a random sample $\{X_1, ..., X_n\}$ from a common distribution F with density f such that $E(X) = \mu$ and $Var(X) = \sigma^2$ for generic random variable $X \sim F$. The sample mean and variances are denoted $\bar{X}_n$ and $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, with the bias corrected variance $s^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

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

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

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

*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Emily Case, Sarah Bass, and Danny Edgel.

(b) $s_{n+1}^2 = ((n-1)s_n^2 + (n/(n+1))(X_{n+1} - \bar{X}_n)^2)/n$

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

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.

Consider sample raw moments: $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$. Raw sample moments are unbiased:

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

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?

Consider sample central moments: $\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^k$. In general, I expect $\hat{m}_k$ to be biased. For example, as shown in lecture, $\hat{m}_2 = \hat{\sigma}_2$ is a biased estimator for variance σ_2.

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

$$\begin{aligned}
Var(\hat{\mu}_k) &= E(\hat{\mu}_k^2) - E(\hat{\mu}_k)^2 \\
&= E\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right)^2\right) - \mu_k^2 \\
&= \frac{1}{n^2}E\left(\left(\sum_{i=1}^{n} X_i^k\right)^2\right) - \mu_k^2 \\
&= \frac{1}{n^2}E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i^k X_j^k\right) - \mu_k^2 \\
&= \frac{1}{n^2}E\left(\sum_{i=1}^{n} X_i^{2k} + \sum_{i=1}^{n}\sum_{j=1;i\neq j}^{n} X_i^k X_j^k\right) - \mu_k^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n} E[X_i^{2k}] + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1;i\neq j}^{n} E[X_i^k]E[X_j^k] - \mu_k^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mu_{2k} + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1;i\neq j}^{n} \mu_k^2 - \mu_k^2 \\
&= \frac{1}{n^2}n\mu_{2k} + \frac{1}{n^2}(n^2-n)\mu_k^2 - \mu_k^2 \\
&= \frac{1}{n}\mu_{2k} + \mu_k^2 - \frac{\mu_k^2}{n} - \mu_k^2 \\
&= \frac{\mu_{2k} - \mu_k^2}{n}
\end{aligned}$$

5. Show that $E(s_n) \leq \sigma$. (Hint: Use Jensen's inequality, CB Theorem 4.7.7).

Because $g(x) = \sqrt{x}$ is a concave function, we can apply Jensen's inequality:

$$E(s_n) = E\left(\sqrt{(n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}\right) \leq \sqrt{E\left((n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right)} = \sqrt{\sigma^2} = \sigma$$

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

I show that $\hat{\sigma}^2 = n^{-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2\right)$ and $n^{-1}\sum_{i=1}^{n}(X_i - \mu)^2 - (\bar{X}_n - \mu)^2 = n^{-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2\right)$, so by transitivity $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \mu)^2 - (\bar{X}_n - \mu)^2$:

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

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

7. Find the covariance of $\hat{\sigma}^2$ and $\bar{X}_n$. Under what condition is this zero?

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

$$
\begin{aligned}
E\Big[\Big(\bar{X}_n-\mu\Big)\sum_{i=1}^{n}(X_i-\mu)^2\Big] &= E\Big[\Big(n^{-1}\sum_{i=1}^{n}X_i-\mu\Big)\sum_{i=1}^{n}(X_i-\mu)^2\Big] \\
&= n^{-1}E\Big[\sum_{i=1}^{n}(X_i-\mu)^3 + \sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n}(X_i-\mu)^2(X_j-\mu)\Big] \\
&= n^{-1}\left[\sum_{i=1}^{n}E[(X_i-\mu)^3] + \sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n}E[(X_i-\mu)^2][E(X_j)-\mu]\right] \\
&= n^{-1}\left[\sum_{i=1}^{n}E[(X_i-\mu)^3] + \sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n}E[(X_i-\mu)^2][\mu-\mu]\right] \\
&= n^{-1}\left[\sum_{i=1}^{n}E[(X_i-\mu)^3] + \sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n}E[(X_i-\mu)^2](0)\right] \\
&= n^{-1}\sum_{i=1}^{n}E[(X_i-\mu)^3] \\
&= E[(X_i-\mu)^3]
\end{aligned}
$$

$$
\begin{aligned}
&E\Big[(\bar{X}_n-\mu)^3\Big] \\
&= E\Big[(n^{-1}\sum_{i=1}^{n}X_i-\mu)^3\Big] \\
&= n^{-3}\sum_{i=1}^{n}E\Big[(X_i-\mu)^3\Big] + n^{-3}\sum_{i=1}^{n}\sum_{j=1;i\neq j}^{n}E\Big[(X_i-\mu)^2(X_j-\mu)\Big] + n^{-3}\sum_{i=1}^{n}\sum_{j=1;j\neq i}^{n}\sum_{k=1;k\neq i;k\neq j}^{n}E\Big[(X_i-\mu)(X_j-\mu)(X_k-\mu)\Big] \\
&= n^{-3}nE\Big[(X_i-\mu)^3\Big] + n^{-3}n(n-1)E[(X_i-\mu)^2][E(X_j)-\mu] + n^{-3}n(n-1)(n-1)[E(X_i)-\mu][E(X_j)-\mu)][E(X_k)-\mu] \\
&= n^{-2}E\Big[(X_i-\mu)^3\Big] + n^{-3}n(n-1)E[(X_i-\mu)^2](0) + n^{-3}n(n-1)(n-1)(0)(0)(0) \\
&= n^{-2}E\Big[(X_i-\mu)^3\Big]
\end{aligned}
$$

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

The covariance of $\hat{\sigma}^2$ and $\bar{X}_n$ is zero if the kurtosis ($E[(X_i-\mu)^3]$) is zero.

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

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

$$E(\bar{X}_n) = E\left(n^{-1}\sum_{i=1}^{n} X_i\right) = n^{-1}\sum_{i=1}^{n} E(X_i) = n^{-1}\sum_{i=1}^{n} \mu_i$$

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

$$Var(\bar{X}_n) = Var\left(n^{-1}\sum_{i=1}^{n} X_i\right) = n^{-2}\sum_{i=1}^{n} Var(X_i) = n^{-2}\sum_{i=1}^{n} \sigma_i^2$$

9. Show that if $Q \sim \chi^2_r$, then $E(Q) = r$ and $Var(Q) = 2r$.

Note that $Q = \sum_{i=1}^n X_i^2$ with $X_i \sim N(0,1)$), then $M_X(t) = \exp\left(\frac{1}{2}t^2\right)$.

$$\begin{aligned}
M_X^{(1)}(t) &= \exp\left(\frac{1}{2}t^2\right)t \\
M_X^{(2)}(t) &= \exp\left(\frac{1}{2}t^2\right) + \exp\left(\frac{1}{2}t^2\right)t^2 \\
M_X^{(3)}(t) &= \exp\left(\frac{1}{2}t^2\right)t + \exp\left(\frac{1}{2}t^2\right)t^3 + 2\exp\left(\frac{1}{2}t^2\right)t \\
&= 3\exp\left(\frac{1}{2}t^2\right)t + \exp\left(\frac{1}{2}t^2\right)t^3 \\
M_X^{(4)}(t) &= 3\exp\left(\frac{1}{2}t^2\right) + 3\exp\left(\frac{1}{2}t^2\right)t^2 + \exp\left(\frac{1}{2}t^2\right)t^4 + 3\exp\left(\frac{1}{2}t^2\right)t^2 \\
&= \exp\left(\frac{1}{2}t^2\right)t^4 + 6\exp\left(\frac{1}{2}t^2\right)t^2 + 3\exp\left(\frac{1}{2}t^2\right)
\end{aligned}$$

$$\begin{aligned}
E[X] &= M_X^{(1)}(0) = 0 \\
E[X^2] &= M_X^{(2)}(0) = 1 \\
E[X^3] &= M_X^{(3)}(0) = 0 \\
E[X^4] &= M_X^{(4)}(0) = 3
\end{aligned}$$

$$E(Q) = E\left(\sum_{i=1}^r X_i^2\right) = \sum_{i=1}^r E(X_i^2) = \sum_{i=1}^r (1) = r$$

$$\begin{aligned}
Var(Q) &= E(Q^2) - E(Q)^2 \\
&= E\left(\left(\sum_{i=1}^r X_i^2\right)^2\right) - r^2 \\
&= E\left(\sum_{i=1}^r \sum_{j=1}^r X_i^2 X_j^2\right) - r^2 \\
&= E\left(\sum_{i=1}^r X_i^4 + \sum_{i=1}^r \sum_{j=1;j\neq i}^r X_i^2 X_j^2\right) - r^2 \\
&= \sum_{i=1}^r E(X_i^4) + \sum_{i=1}^r \sum_{j=1;j\neq i}^r E(X_i^2)E(X_j^2) - r^2 \\
&= \sum_{i=1}^r (3) + \sum_{i=1}^r \sum_{j=1;j\neq i}^r (1)(1) - r^2 \\
&= 3r + r(r-1) - r^2 \\
&= 2r
\end{aligned}$$

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

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

$$
\begin{aligned}
E(\bar{X}_n - \bar{Y}_n) &= E\left(n_1^{-1} \sum_{i=1}^{n_1} X_i - n_2^{-1} \sum_{i=1}^{n_2} Y_i\right) \\
&= n_1^{-1} \sum_{i=1}^{n_1} E(X_i) - n_2^{-1} \sum_{i=1}^{n_2} E(Y_i) \\
&= n_1^{-1} \sum_{i=1}^{n_1} \mu_X - n_2^{-1} \sum_{i=1}^{n_2} \mu_Y \\
&= \mu_X - \mu_Y
\end{aligned}
$$

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

$$
\begin{aligned}
Var(\bar{X}_n - \bar{Y}_n) &= Var\left(n_1^{-1} \sum_{i=1}^{n_1} X_i + n_2^{-1} \sum_{i=1}^{n_2} Y_i\right) \\
&= n_1^{-2} \sum_{i=1}^{n_1} Var(X_i) + n_2^{-2} \sum_{i=1}^{n_2} Var(Y_i) \\
&= n_1^{-2} \sum_{i=1}^{n_1} \sigma_X^2 + n_2^{-2} \sum_{i=1}^{n_2} \sigma_Y^2 \\
&= \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}
\end{aligned}
$$

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

Consider any independent $W \sim N(\mu_W, \sigma_W)$ and $Z \sim N(\mu_Z, \sigma_Z)$. Therefore:

$$\begin{aligned}
M_W(t) &= \exp\left(t\mu_W + \frac{t^2}{2}\sigma_W^2\right) \\
M_Z(t) &= \exp\left(t\mu_Z + \frac{t^2}{2}\sigma_Z^2\right) \\
M_{W+Z}(t) &= E[\exp(t(W+Z))] \\
&= E[\exp(tW)\exp(tZ)] \\
&= E[\exp(tW)]E[\exp(tZ)] \\
&= M_W(t)M_Z(t) \\
&= \exp\left(t\mu_W + \frac{t^2}{2}\sigma_W^2\right)\exp\left(t\mu_Z + \frac{t^2}{2}\sigma_Z^2\right) \\
&= \exp\left(t\mu_W + \frac{t^2}{2}\sigma_W^2 + t\mu_Z + \frac{t^2}{2}\sigma_Z^2\right) \\
&= \exp\left(t(\mu_W + \mu_Z) + \frac{t^2}{2}(\sigma_W^2 + \sigma_Z^2)\right)
\end{aligned}$$

So $W + Z \sim N(\mu_W + \mu_Z, \sigma_W^2 + \sigma_Z^2)$.

By induction, $\sum_{i=1}^{n_1} X_i \sim (n_1\mu_X, n_1\sigma_X^2)$ and $\sum_{i=1}^{n_2} Y_i \sim (n_2\mu_Y, n_2\sigma_Y^2)$. So $\bar{X}_n \sim N(\mu_X, n_1^{-1}\sigma_X^2)$ and $\bar{Y}_n \sim N(\mu_Y, n_2^{-1}\sigma_Y^2)$. And $\bar{X}_n - \bar{Y}_n \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}\right)$.