Lecture 3: Some Moments for Estimators of the Mean and Variance

MATH 667-01
Statistical Inference
University of Louisville

August 29, 2017

Introduction

- We start with the definition of a statistic given in Section 5.2 of Casella and Berger (2001)[1].
- We review several important definitions and theorems concerning expected values, moments, and independence from Sections 2.2, 2.3, 4.2, 4.5, and 4.6.
- Then we derive the mean and variance of the sample mean and the mean of the sample variance as discussed in Section 5.2.

[1] Casella, G. and Berger, R. (2001). Statistical Inference, second edition. Duxbury Press.

- *Definition L3.1* (Def 5.2.1 on p.211): Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, \ldots, X_n)$. Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called a *statistic*. The probability distribution of a statistic Y is called the *sampling distribution of* Y.
- The estimators $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ and $\widehat{\sigma^2} = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ are examples of statistics.
- Some other examples of statistics are median$(X_1, \ldots, X_n)$ and $\max(X_1, \ldots, X_n)$.
- Note that a statistic cannot be a function of a population parameter(s).

- *Definition L3.2* (Def 2.2.1 on p.55): The *expected value* or *mean* of a random variable $g(X)$, denoted by $\mathsf{E}[g(X)]$, is

$$\mathsf{E}[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x)\, dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) & \text{if } X \text{ is discrete} \end{cases}$$

  provided that the integral or the sum exists.
  If $\mathsf{E}|g(X)| = \infty$, we say that $\mathsf{E}[g(X)]$ does not exist.
- *Theorem L3.1* (Thm 2.2.5 on p.57): Let X be a random variable and let a, b, and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,
  a. $\mathsf{E}[ag_1(X) + bg_2(X) + c] = a\mathsf{E}[g_1(X)] + b\mathsf{E}[g_2(X)] + c$.
  b. If $g_1(x) \geq 0$ for all x, then $\mathsf{E}[g_1(X)] \geq 0$.
  c. If $g_1(x) \geq g_2(x)$ for all x, then $\mathsf{E}[g_1(X)] \geq \mathsf{E}[g_2(X)]$.
  d. If $a \leq g_1(x) \leq b$ for all x, then $a \leq \mathsf{E}[g_1(X)] \leq b$.

Review

- *Definition L3.3* (Def 2.3.1 on p.59): For each integer n, the *nth moment of X* is $\mu'_n = \mathsf{E}[X^n]$.
  The *nth central moment of X* is $\mu_n = \mathsf{E}[(X-\mu)^n]$, where $\mu = \mu'_1 = \mathsf{E}[X]$ is referred to as the *mean*.
- *Definition L3.4* (Def 2.3.2 on p.59): The *variance* of a random variable X is its second central moment,
  $\mathsf{Var}[X] = \mathsf{E}[(X - \mathsf{E}[X])^2]$.
  The *standard deviation* of X is $\sqrt{\mathsf{Var}[X]}$.
- An useful alternative formula for the variance is

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

- *Theorem L3.2* (Thm 2.3.4 on p.60): If X is a random variable with finite variance, then for any constants a and b, $\mathsf{Var}[aX+b] = a^2\mathsf{Var}[X]$.

- *Theorem L3.3* (Thm 4.6.6 on p.183): Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be mutually independent random vectors. Let $g_1, \ldots, g_n$ be real-valued functions such that $g_i(\boldsymbol{x}_i)$ is a function only of $\boldsymbol{x}_i, i = 1, \ldots, n$. Then

$$\mathsf{E}[g_1(\boldsymbol{X}_1) \bullet \cdots \bullet g_n(\boldsymbol{X}_n)] = \mathsf{E}[g_1(\boldsymbol{X}_1)] \bullet \cdots \bullet \mathsf{E}[g_n(\boldsymbol{X}_n)].$$

- *Proof of Theorem L3.3*: Without loss of generality, assume X_1 and X_2 are two continuous random variables.

$$\begin{aligned}
\mathsf{E}[g_1(X_1)g_2(X_2)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1)g_2(x_2)f(x_1,x_2)dx_1dx_2 \\
&\overset{2.4}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x_1)g_2(x_2)f(x_1)f(x_2)dx_1dx_2 \\
&= \int_{-\infty}^{\infty} g_2(x_2)f(x_2)\int_{-\infty}^{\infty} g_1(x_1)f(x_1)dx_1dx_2 \\
&= \int_{-\infty}^{\infty} g_2(x_2)f(x_2)\left(\int_{-\infty}^{\infty} g_1(x_1)f(x_1)dx_1\right)dx_2
\end{aligned}$$

- *Proof of Theorem L3.3 continued*:

$$\begin{aligned}
\mathsf{E}[g_1(X_1)g_2(X_2)] &= \int_{-\infty}^{\infty} g_2(x_2)f(x_2)\left(\int_{-\infty}^{\infty} g_1(x_1)f(x_1)dx_1\right)dx_2 \\
&= \int_{-\infty}^{\infty} g_1(x_1)f(x_1)dx_1 \int_{-\infty}^{\infty} g_2(x_2)f(x_2)dx_2 \\
&= \mathsf{E}[g_1(X_1)]\mathsf{E}[g_2(X_2)]
\end{aligned}$$

- *Corollary to Thm L3.3*: Let $X_1, \ldots, X_n$ be mutually independent random variables. For sets $A_1 \subset \mathbb{R}, \ldots, A_n \subset \mathbb{R}$,

$$P\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{i=1}^{n} P\left(X_i \in A_i\right).$$

  (This is a generalization of Theorem 4.2.10(a) on p.154.)

- *Proof of Corollary*: Apply Thm L3.3 with $g_i(x_i) = I(x_i \in A_i)$ where I is the indicator function equal to 1 is the statement is true and equal to 0 if the statement is false.

- *Theorem L3.4* (Thm 4.6.11 on p.184): Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors. Then $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are mutually independent random vectors if and only if there exist functions $g_i(x_i), i = 1, \ldots, n$, such that the joint pdf/pmf of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ can be written as

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = g_1(\boldsymbol{x}_1) \bullet \cdots \bullet g_n(\boldsymbol{x}_n).$$

- *Proof of Theorem L3.4*: Here we prove the statement for the case when $X_1, \ldots, X_n$ are continuous random variables. If $X_1, \ldots, X_n$ are independent, then

$$f(x_1, \ldots, x_n) \overset{2.4}{=} \prod_{i=1}^{n} f_{X_i}(x_i)$$

so $g_i(x_i) = f_{X_i}(x_i)$ for $i = 1, \ldots, n$ satisfies the condition.

- *Proof of Theorem L3.4 continued*: Conversely, suppose $f(x_1, \ldots, x_n) = \prod_{i=1}^n g_i(x_i)$.
- Let $c_i = \displaystyle\int_{-\infty}^{\infty} g_i(x_i)\ dx_i$.
- Note that
$$\prod_{i=1}^n c_i = \prod_{i=1}^n \int_{-\infty}^{\infty} g_i(x_i)\ dx_i = \int_{-\infty}^{\infty} \prod_{i=1}^n g_i(x_i)\ dx_1 \cdots dx_n$$
$$= \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\ dx_1 \cdots dx_n = 1.$$

- *Proof of Theorem L3.4 continued*: Next, note that the marginal pdf of X_i is

$$
\begin{aligned}
f_{X_i}(x_i) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\ dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{n} g_j(x_j)\ dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n \\
&= g_i(x_i) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j \neq i} g_j(x_j)\ dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n \\
&= g_i(x_i) \prod_{j \neq i} \int_{-\infty}^{\infty} g_j(x_j)\ dx_j \\
&= g_i(x_i) \prod_{j \neq i} c_j.
\end{aligned}
$$

- *Proof of Theorem L3.4 continued*: Now, the product of the marginals is
$$\prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n}\left\{g_i(x_i)\prod_{j\neq i} c_j\right\} = \prod_{i=1}^{n} g_i(x_i)\left\{\prod_{i=1}^{n}\prod_{j\neq i} c_j\right\}$$
$$= \prod_{i=1}^{n} g_i(x_i)\left\{\prod_{k=1}^{n} c_k^{n-1}\right\} = \prod_{i=1}^{n} g_i(x_i)\left\{\prod_{k=1}^{n} c_k\right\}^{n-1}$$
$$= \prod_{i=1}^{n} g_i(x_i) = f(x_1,\ldots,x_n).$$
- Thus, $X_1,\ldots,X_n$ are independent.

- *Theorem L3.5* (Thm 4.6.12 on p.184): Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random vectors. Let $g_i(\boldsymbol{x}_i)$ be a function only of $\boldsymbol{x}_i, i = 1, \ldots, n$. Then the random variables $U_i = g_i(\boldsymbol{X}_i), i = 1, \ldots, n$, are mutually independent.
- *Proof of Theorem L3.5*: Here we prove the statement for the continuous case.
- Let $A_{i,u} = \{\boldsymbol{x}_i : g_i(\boldsymbol{x}_i) \leq u\}$.
- Then the cumulative distribution function (cdf) of $U_1, \ldots, U_n$ is

$$F_{U_1,\ldots,U_n}(u_1, \ldots, u_n) = P\left(\bigcap_{i=1}^{n} \{U_i \leq u_i\}\right)$$

$$= P\left(\bigcap_{i=1}^{n} \{\boldsymbol{X}_i \in A_{i,u_i}\}\right) = \prod_{i=1}^{n} P\left(\boldsymbol{X}_i \in A_{i,u_i}\right).$$

- *Proof of Theorem L3.5 continued*: Differentiating with respect to each u_i (see Equation 4.1.4 on p.147), the joint pdf of $U_1, \ldots, U_n$ is

$$
\begin{aligned}
f_{U_1,\ldots,U_n}(u_1,\ldots,u_n) &= \frac{\partial^n}{\prod_{i=1}^n \partial u_i} F_{U_1,\ldots,U_n}(u_1,\ldots,u_n) \\
&= \prod_{i=1}^n \frac{d}{du_i} P\left(\boldsymbol{X}_i \in A_{i,u_i}\right).
\end{aligned}
$$

Since the ith term is a function only of u_i for $i = 1, \ldots, n$, $U_1, \ldots, U_n$ are independent by Theorem L3.4.

- *Definition L3.5* (Def 4.5.1 on p.169): Assume $\mathsf{E}[X]$ and $\mathsf{E}[Y]$ exist. The *covariance of* X and Y is defined by

$$\mathsf{Cov}[X,Y] = \mathsf{E}\left[(X - \mathsf{E}[X])(Y - \mathsf{E}[Y])\right].$$

- *Theorem L3.6* (Thm 4.5.3 on p.170): Provided all expectations exist, $\mathsf{Cov}[X,Y] = \mathsf{E}[XY] - \mathsf{E}[X]\mathsf{E}[Y]$.
- *Theorem L3.7* (Thm 4.5.5 on p.171): If X and Y are independent random variables, then $\mathsf{Cov}[X,Y] = 0$.

- *Theorem L3.8* (Thm 4.5.6 on p.171): If X and Y are any two random variables, and a and b are any two constants, then

$$\mathsf{Var}[aX + bY] = a^2\mathsf{Var}[X] + b^2\mathsf{Var}[Y] + 2ab\mathsf{Cov}[X, Y].$$

  If X and Y are independent random variables, then

$$\mathsf{Var}[aX + bY] = a^2\mathsf{Var}[X] + b^2\mathsf{Var}[Y].$$

- A more general formula for covariances of sums is given below.
- *Theorem L3.9*: For random variables $X_1, \dots, X_m, Y_1, \dots, Y_n$ and constants $a_1, \dots, a_m, b_1, \dots, b_n$,

$$\mathsf{Cov}\left[\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \mathsf{Cov}[X_i, Y_j].$$

- *Theorem L3.10* (Lem 5.2.5 on p.213): Let $X_1, \ldots, X_n$ be a random sample from a population and let $g(x)$ be a function such that $\mathsf{E}[g(X_1)]$ and $\mathsf{Var}[g(X_1)]$ exist. Then
$$\mathsf{E}\left[\sum_{i=1}^n g(X_i)\right] = n\mathsf{E}[g(X_1)] \text{ and}$$
$$\mathsf{Var}\left[\sum_{i=1}^n g(X_i)\right] = n\mathsf{Var}[g(X_1)].$$
- *Proof of Theorem L3.10*: First, we have

$$\begin{aligned}\mathsf{E}\left[\sum_{i=1}^n g(X_i)\right] &\overset{3.4}{=} \sum_{i=1}^n \mathsf{E}[g(X_i)] \\ &\overset{2.5}{=} \sum_{i=1}^n \mathsf{E}[g(X_1)] = n\mathsf{E}[g(X_1)].\end{aligned}$$

- *Proof of Theorem L3.10 continued*: Next, we have

$$
\begin{aligned}
\mathsf{Var}\left[\sum_{i=1}^{n} g(X_i)\right] &= \mathsf{Cov}\left[\sum_{i=1}^{n} g(X_i), \sum_{j=1}^{n} g(X_j)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathsf{Cov}[g(X_i), g(X_j)]
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathsf{Cov}[g(X_i), g(X_i)] + \sum_{i=1}^{n}\sum_{j\neq i} \mathsf{Cov}[g(X_i), g(X_j)] \\
&\overset{3.14}{=} \sum_{i=1}^{n} \mathsf{Var}[g(X_i)] + 0 \\
&\overset{2.5}{=} \sum_{i=1}^{n} \mathsf{Var}[g(X_1)] = n\mathsf{Var}[g(X_1)].
\end{aligned}
$$

- *Theorem L3.11* (Thm 5.2.4(b) on p.212): Let $x_1, \ldots, x_n$ be any numbers. Then $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$.
- *Proof of Theorem L3.11*:

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})^2 &= \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} 2x_i\bar{x} + \sum_{i=1}^{n} \bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2
\end{aligned}
$$

Some Moments of $\bar{X}$ and $\widehat{\sigma^2}$

- *Theorem L3.12* (Thm 5.2.6 on p.213): Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. Then
  a. $\mathsf{E}[\bar{X}] = \mu$,
  b. $\mathsf{Var}[\bar{X}] = \dfrac{\sigma^2}{n}$,
  c. $\mathsf{E}\left[\widehat{\sigma^2}\right] = \dfrac{(n-1)\sigma^2}{n}$.

- *Proof of Theorem L3.12*:

  (a) $$\mathsf{E}[\bar{X}] = \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \overset{3.4}{=} \frac{1}{n}\mathsf{E}\left[\sum_{i=1}^{n} X_i\right] \overset{3.16}{=} \frac{1}{n}(n\mathsf{E}[X_1]) = \mu$$

- (b) $$\mathsf{Var}[\bar{X}] = \mathsf{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \overset{3.5}{=} \frac{1}{n^2}\mathsf{Var}\left[\sum_{i=1}^{n} X_i\right] \overset{3.16}{=} \frac{1}{n^2}(n\mathsf{Var}[X_1]) = \frac{\sigma^2}{n}$$

- *Proof of Theorem L3.12 continued*:

$$
\begin{aligned}
\text{(c) } \mathsf{E}\left[\widehat{\sigma^2}\right] &= \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] \\
&\overset{3.18}{=} \mathsf{E}\left[\frac{1}{n}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)\right] \\
&\overset{3.4}{=} \frac{1}{n}\left(\mathsf{E}\left[\sum_{i=1}^{n}X_i^2\right]-n\mathsf{E}\left[\bar{X}^2\right]\right) \\
&\overset{3.16}{=} \frac{1}{n}\left(n\mathsf{E}\left[X_1^2\right]-n\mathsf{E}\left[\bar{X}^2\right]\right) \\
&\overset{3.5}{=} \frac{1}{n}\left(n\left\{\sigma^2+\mu^2\right\}-n\left\{\frac{\sigma^2}{n}+\mu^2\right\}\right) \\
&= \frac{1}{n}\left(n\sigma^2-\sigma^2\right) = \frac{(n-1)}{n}\sigma^2
\end{aligned}
$$

Sample Mean and Sample Variance

- *Definition L3.6* (Def 5.2.2): The *sample mean* is the arithmetic average of the values in a random sample. It is usually denoted by $\bar{X} = \dfrac{X_1 + \cdots + X_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.
- *Definition L3.7* (Def 5.2.3): The *sample variance* is the statistic defined by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

  The *sample standard deviation* is the statistic defined by $S = \sqrt{S^2}$.
- *Theorem L3.13*: $\mathsf{E}\left[S^2\right] = \sigma^2$