Stats 135, Fall 2019

Lecture 3, Wednesday, 9/4/2019

1 Review

Adam Lucas hinted that our last result from lecture 2 will be used in the next lecture for a proof, which briefly gloss over now.
Let $X_1, \ldots, X_n$ be iid with $E(x) = \mu$, with $\mathrm{Var}(x) = \sigma^2$.

Theorem 1.1. The sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ is an unbiased estimator of the true population variance r^2.

Proof. Note that $E\left(\sum x_i\right) = \sum E(x_i)$ and $E(cx) = cE(x)$. Recall that $\mathrm{Var}(x) = x = E(x^2) - \underbrace{E(x)^2}_{\mu^2}$ which implies

$$E(x^2) = \sigma^2 + \mu^2.$$

Now, $\mathrm{Var}(\overline{x}) = E(\overline{x}^2) - E(\overline{x})^2$ implies

$$E(\overline{x}^2) = \frac{\sigma^2}{n} + \mu^2. \tag{1}$$

To show that the sample variance S^2 is an unbiased estimator, we show that $E(S^2) = \sigma^2$. Consider:

$$\begin{aligned}
E\left(\sum (x_i - \overline{x})^2\right) &= E\left(\sum (x_i^2 - 2x_i\overline{x} + \overline{x}^2)\right) \\
&= E\left(\sum x_i^2 - \sum 2x_i\overline{x} + \sum \overline{x}^2\right) \\
&= E\left(\sum x_i^2 - 2\overline{x}\sum x_i + n\overline{x}^2\right) \quad \text{because } \overline{x} \text{ is a constant} \\
&= E\left(\sum x_i^2 - 2\overline{x}n\overline{x} + n\overline{x}^2\right) \quad \text{because } \sum x_i = n\overline{x} \\
&= E\left(\sum x_i^2 - n\overline{x}^2\right) \\
&= \sum E(x_i^2) - nE(\overline{x}^2) \\
&= \sum (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \quad \text{by (1) above} \\
&= n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2} \\
&= (n-1)\sigma^2
\end{aligned}$$

Hence

$$\frac{1}{n-1}E\left(\sigma(x_i - \overline{x})^2\right) = \sigma^2,$$

as required. □

Adam Lucas notes that he's not particularly interested in the proof, but we should be able to understand every step of it.
A question arises in the audience: why, intuitively is it $\frac{1}{n-1}$? Lucas notes that $1/n$ is a little bit too small, and to make it a little bit bigger, we use $1/(n-1)$. He says 'honestly it's just what we need to make it unbiased', and it simply appears from the algebra; he laments he does not have a super intuitive explanation for this outside of the algebraic steps of the proof.

2 §7.3.1: Expectation and Variance of Sample Mean

We now prove:

$$\mathrm{Var}(\overline{x}) = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right],$$

if $x_1, \ldots, x_n$ are identically distributed with mean μ and variance σ^2.
We give a different proof from what was given in the book (we will follow Pitman page 441-443).

Proof. Assume $x_1, \ldots, x_n$ are SRS (simple random samples, without replacement), each with mean μ and variance r^2.
For example, x_i can be the annual income of the ith household in the US.
We take a look at the sum of these incomes:

$$T := x_1 + \cdots + x_n$$

and then

$$\mathrm{Var}(T) = \mathrm{Var}(x_1 + \cdots + x_n) = \sum_{i,j=1}^{n} \mathrm{Cov}(x_i, x_j),$$

and we can break up this sum into where $i = j$ and $i \neq j$, where equality simply nets variance:

$$\mathrm{Var}(T) = \sum_{j=1}^{n} \mathrm{Var}(x_i) + \sum_{i \neq j} \mathrm{Cov}(x_i, x_j).$$

Because all these x_i are identically distributed, this will simply give:

$$= \underbrace{n \cdot \mathrm{Var}(x_i)}_{\sigma^2} + n(n-1)\mathrm{Cov}(x_1, x_2),$$

and we employ a trick to find out this covariance. Recall that n is a sample of some population, so assume there are N of these (i.e. US households). The trick we use is to consider $n := N$ (our sample will be the entire population). In this case, the variance of T will simply be 0 because there will be no variation when we take our sample to be the entire population! So we have:

$$\mathrm{Var}(T) = 0 \implies 0 = N\sigma^2 + N(N-1)\mathrm{Cov}(x_1, x_2),$$

which implies

$$\mathrm{Cov}(x_1, x_2) = \frac{-N\sigma^2}{N(N-1)} = \boxed{\frac{-\sigma^2}{N-1}}.$$

Now for general n, we have:

$$\mathrm{Var}(T) = n\sigma^2 + n(n-1)\left(\frac{-\sigma^2}{N-1}\right) = n\sigma^2\left[\frac{N-n}{N-1}\right],$$

so

$$\mathrm{Var}(\overline{x}) = \mathrm{Var}\left(\frac{T}{n}\right) = \frac{1}{n^2}\mathrm{Var}(T) = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right].$$

□

Remark: Also, we have:

$$E\left(\overline{x}^2\right) = \text{Var}(\overline{x}) + E(\overline{x})^2 = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right] + \mu^2$$

3 §7.3.2: Estimation of Population Variance

We showed that the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$ is an unbiased estimator of the true population variance r^2 (i.e. $E(s^2) = \sigma^2$).
However, this is only true if $x_1, \ldots, x_n$ are iid. More generally, if $x_1, \ldots, x_n$ is a SRS (simple random sample), then

$$\left(1 - \frac{1}{N}\right) S^2$$

is an unbiased estimator of r^2. We won't prove this in class, but see notes posted by Lucas.

Theorem 3.1. Let $x_1, \ldots, x_n$ be SRS with mean μ and variance σ^2. Then with the 'fudge factor' $\left(\frac{N-1}{N}\right)$, we have:

$$E\left[\left(\frac{N-1}{N}\right)\frac{1}{n-1}\left(\sigma(x_i - \overline{x})^2\right)\right] = \sigma^2.$$

With this, we essentially finish Chapter 7. The results we need to know are highlighted in Lucas' notes, from page 214 in Rice.

Population Parameter	Estimate	Variance of Estimate	Estimated Variance
μ	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$	$s_{\overline{X}}^2 = \frac{s^2}{n}\left(1 - \frac{n}{N}\right)$
p	$\hat{p}$ = sample proportion	$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}\left(\frac{N-n}{N-1}\right)$	$s_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n-1}\left(1 - \frac{n}{N}\right)$
τ	$T = N\overline{X}$	$\sigma_T^2 = N^2\sigma_{\overline{X}}^2$	$s_T^2 = N^2 s_{\overline{X}}^2$
σ^2	$\left(1 - \frac{1}{N}\right)s^2$		

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

4 Chapter 8: Estimation of Parameters, Fitting of a Probability Distribution

For motivation, the idea is that we start with observed and recorded data in the real world. We may believe our data set obeys some probability distribution (say Poisson) for a certain parameter θ. We wish to find an estimate for our parameter θ, and we usually write the estimate as $\hat{\theta}$ which has good properties: perhaps it is an unbiased estimator, or other properties. The canonical example is radioactive decay. Suppose in an experiment we observe α-partical decay for 12,070 seconds. We break this time interval (axis) into 10 second intervals and count the number of arrivals in each interval.

Performing the experiment and counting the total of 10,129 α-particle arrivals, we let X be the number of arrivals in a 10-second interval. Then X is a good candidate for a Poisson distribution because α-particles obey the three properties that a Poisson process has. Let λ be the rate of arrival in 10 seconds.

(1) λ is constant (Americium has a long half-life)
(2) the numbers (counts) of arrivals in disjoint intervals are independent
(3) the arrivals do not coincide (no simultaneous arrivals)

Hence we can model this as 1207 iid Poisson(λ) random variables (RV). Recall that $X \sim \text{Poisson}(\lambda)$ implies:

$$\pi_k = P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

We don't know λ, so we need to estimate. We're told there are 10,129 arrivals, so we take the average:

$$\hat{\lambda} := \frac{10,129 \text{ arrivals}}{1207 \text{ seconds}} = 8.392 \text{ arrivals / 10 sec intervals.}$$

We have 1207 independent intervals, so what is the probability that an interval gets 3 arrivals? This is simple: We know X is the number of arrivals and $X \sim \text{Poisson}(8.4)$, so:

$$p = P(X = 3) = \frac{e^{-8.4}(8.4)^3}{3!} = 0.022$$

Now we need to know the number of intervals that get 3 arrivals. The probability of getting 3 arrivals in any interval is given by the probability 0.022. We can think of this as a bunch of Bernoulli trials (either get 3 or not), where the chance of getting 3 is 0.022.

Let Y be the number of intervals that get 3 arrivals. Then we have:

$$Y \sim \text{Binomial}(1207, .022),$$

so

$$E(Y) = np = 1207 \cdot 0.022$$

using this, we fill out this table for expected counts:

n	Observed	Expected
0–2	18	12.2
3	28	27.0
4	56	56.5
5	105	94.9
6	126	132.7
7	146	159.1
8	164	166.9
9	161	155.6
10	123	130.6
11	101	99.7
12	74	69.7
13	53	45.0
14	23	27.0
15	15	15.1
16	9	7.9
17+	5	7.1
	1207	1207

For example, to get $n = 4$, we have:

$$56.5 = \text{Binomial}\left(1207, \frac{e^{-8.4}(8.4)^4}{4!}\right)$$

In Chapter 9, we perform a χ-squared test to see how well our model fits the data. Next time we'll look at the method of moment estimating (§8.4).