Prob & Stats - Section 6.1: Unbiased Estimators

The whole point of statistical inference is to see what statistics tell us about parameters.

Statistics are numerical characteristics of random samples.

Parameters are numerical characteristics of populations.

Our interest is in figuring out which statistics give us good information about certain parameters. There are many ways to do this. We just have to decide what we think "good" means. One such choice is "unbiasedness."

Defn: A statistic $\hat{\theta}$ is called an unbiased estimator of a parameter θ if θ is the mean of $\hat{\theta}$'s sampling distribution. i.e. $E(\hat{\theta}) = \theta$

Otherwise, an estimator is called biased.

We saw in class that:

a) $\overline{X}$ is an unbiased estimator for μ.
Why? $E(\overline{X}) = E(\Sigma X/n) = 1/n * \Sigma E(X_i) = 1/n * \Sigma\mu = 1/n * n\mu = \mu$

b) $\hat{p}$ is an unbiased estimtor for p.
Why? $E(\hat{p}) = E(\Sigma X_i/n) = 1/n * \Sigma E(X_i) = 1/n * \Sigma p = 1/n * np = p$

One other thing we mentioned in class (but didn't prove) was that:

c) s^2 is an unbiased estimator for σ^2
Why? $E(s^2) = \sigma^2$

(I've put the proof for this at the end of the document, just for your information or curiosity. You will not have to understand the proof, or really even any of the steps.)

This last one answers a question we asked earlier in the course: When calculating the variance (and standard deviation) using the summation formula, why do we divide by n for the population, but we divide by n-1 for the sample?

The answer? If we divided by n with the sample, then the estimator s^2 would not be unbiased! i.e. It wouldn't be a good estimator.

Okay, so we know $\overline{X}$ is unbiased for μ. Are there any other unbiased estimators for μ?

Example: Suppose X_1, ... X_n is a random sample of size n from some population with mean μ and standard deviation σ.
Prove that the following are (or are not) unbiased.

Stat 1: X_1 Just the first one in the sample. We ignore the rest.

Proof: We must show that $E(X_1) = \mu$.
But X_1 has that same pdf that the population has,
so $E(X_1)$ is just asking for the mean of that pop,
which we know to be μ. Therefore, unbiased!

Stat 2: $(X_1+X_2)/2$

Proof: $E[(X_1+X_2)/2] = 1/2*E(X_1+X_2) = 1/2*(E(X_1)+E(X_2))$
$= 1/2*(\mu + \mu) = \mu$. Unbiased.

<u>You check the rest of these</u>. Are they unbiased or not?
(Answers on the next page.)

Stat 3: $(X_1-X_2+X_3)$

Stat 4: $\frac{X_1+2X_2+3X_3+4X_4}{10}$

Stat 5: $\frac{X_1-2X_2+3X_3-4X_4}{-2}$

Stat 6: $\frac{X_1+X_2+X_{n-1}+X_n}{4}$

Answers for the previous page: All 6 are unbiased.

So, it should be clear that if we can make up these six unbiased estimators, we could probably make up quite a few more. In fact, there must be infinitely many unbiased estimators for any parameter. So how do we decide which one is best, or at least better than the others we have?

Defn: If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased for θ, then we say $\hat{\theta}_2$ is a more efficient estimator of θ if $\text{Var}(\hat{\theta}_2) < \text{Var}(\hat{\theta}_1)$.

That is, the smaller the variance, the more accurate the estimation. So let's find some variances of our unbiased estimators. Just remember of our two rules about variances: $\text{Var}(aX) = a^2\text{Var}(X)$ and $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$, provided X & Y are independent, which will be the case on our random samples. Combining those two properties we also get $\text{Var}(aX+bY) = a^2\text{Var}(X)+b^2\text{Var}(Y)$. That is, when the coefficients factor out, you square them.

Stat 1: X_1 $\quad \text{Var}(X_1) = \sigma^2$, since X_1 has that pdf with μ and σ.

Stat 2: $(X_1 + X_2)/2$

$\text{Var}[(X_1 + X_2)/2] = 1/4*\text{Var}(X_1 + X_2) = 1/4*(\text{Var}(X_1)+\text{Var}(X_2))$

$= 1/4*(\sigma^2 + \sigma^2) = \sigma^2/2$

So Stat 2 is better than Stat 1.

Stat 3: $(X_1-X_2+X_3)$

$\text{Var}[X_1-X_2+X_3] = 1^2*\text{Var}(X_1) + (-1)^2*\text{Var}(X_2) + 1^2*\text{Var}(X_3)$

$= \text{Var}(X_1) + \text{Var}(X_2) + \text{Var}(X_3) = \sigma^2+\sigma^2+\sigma^2 = 3\sigma^2$.

So Stat 3 is much worse than Stat 1 and Stat 2.

Your turn: Find the variances of the remaining unbiased estimators. Then identify which of the six is the best. (Answers on the next page.)

Stat 4: $\frac{X_1 + 2X_2 + 3X_3 + 4X_4}{10}$

Stat 5: $\frac{X_1 - 2X_2 + 3X_3 - 4X_4}{-2}$

Stat 6: $\frac{X_1 + X_2 + X_{n-1} + X_n}{4}$

Answers for the previous page: Stat 4 is $3\sigma^2/10$, Stat 5 is $15\sigma^2/2$, and Stat 6 is $\sigma^2/4$. So the best of all is Stat 6.

Now let's include one more unbiased estimator: $\bar{X}$. We know that the variance for $\bar{X}$ is σ^2/n. So, if our random sample size is at least 4 (which it would have to be for some of those stats to exist), then $\bar{X}$ must be the best of all seven we have.

But <u>which unbiased estimator is the best</u>? How can we know?

Defn: Among all unbiased estimators of θ, the one with the least variance is called the minimum variance unbiased estimator (MVUE).

From the theory of statistics we can find a number or expression called the Rao-Cramer Lower Bound (RCLB). This RCLB is the very <u>smallest variance any unbiased estimator could possibly have</u>. If we can find an unbiased estimator that actually hits this RCLB, then it must be the best unbiased estimator of all. (The RCLB formula and one example is shown on a following page, but you will not be responsible for any information about it.)

It turns out that, if our population is ND(μ,σ), then the RCLB is σ^2/n. That is, for all unbiased estimators of μ, the very least variance any of them could have is σ^2. What does that tell you about the MVUE in this case?

<u>Your homework for today</u>. You should be able to email your answers to me. No need to show work.

1) Answer that last question above. In the case where our population is ND(μ,σ), what unbiased estimator has the least variance of all unbiased estimators?

2) From your textbook, do 6.1 #13. Report to me either: "I was able to do it" or "I got partway there but couldn't quite get the anwer" or "I didn't know how to start."

3) From your textbook, do 6.1 #15a. Report the same way as on the previous one.

NOTE: There are hints for the last two on the homework page.

Proof that $E(s^2) = \sigma^2$

Consider a random sample of size n from a ?D(μ,σ) distribution.
So our random sample would be $X_1, X_2, \ldots X_n$.

What is $E(s^2)$?

Since $s^2 = \Sigma(X_i-\bar{X})^2/(n-1)$, then $E(s^2) = E[\Sigma(X_i-\bar{X})^2/(n-1)]$

We're going to factor that constant denominator n-1 out of there and just focus on the numerator for the moment. Then we will divide by n-1 at the end.

(1) $E[\Sigma(X_i-\bar{X})^2]$

(2) $= E[\Sigma(X_i-\mu+\mu-\bar{X})^2]$ (just put a μ in and took it back out)

(3) $= E[\Sigma((X_i-\mu)-(\bar{X}-\mu))^2]$ (rearranged the stuff inside the parentheses)

(4) $= E[\Sigma[(X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2]]$ (FOILed the parentheses)

(5) $= E[\Sigma(X_i-\mu)^2 - 2\Sigma(X_i-\mu)(\bar{X}-\mu) + \Sigma(\bar{X}-\mu)^2]$ (spread the summation)

(6) $= E[\Sigma(X_i-\mu)^2 - 2(\bar{X}-\mu)*\Sigma(X_i-\mu) + \Sigma(\bar{X}-\mu)^2]$ (factored constants out of the middle sum, since $(\bar{X}-\mu)$ is constant)

(7) $= E[\Sigma(X_i-\mu)^2 - 2(\bar{X}-\mu)*(\Sigma X_i-\Sigma\mu) + \Sigma(\bar{X}-\mu)^2]$ (spread the sum through the mid part)

(8) $= E[\Sigma(X_i-\mu)^2 - 2(\bar{X}-\mu)*(n\bar{X}-n\mu) + \Sigma(\bar{X}-\mu)^2]$ (rewrote the mid part)

(9) $= E[\Sigma(X_i-\mu)^2 - 2n(\bar{X}-\mu)^2 + \Sigma(\bar{X}-\mu)^2]$ (factored the mid part)

(10) $= E[\Sigma(X_i-\mu)^2 - 2n(\bar{X}-\mu)^2 + n(\bar{X}-\mu)^2]$ (evaluated the last part, since $(\bar{X}-\mu)$ is still constant)

(11) $= E[\Sigma(X_i-\mu)^2 - n(\bar{X}-\mu)^2]$ (combined mid part and last part)

(12) $= \Sigma E[(X_i-\mu)^2] - nE[(\bar{X}-\mu)^2]$ (spread the E through the sum, and then swapped the E & sum on the first part)

(13) $= \Sigma Var(X_i) - n\ Var(\bar{X})$

(14) $= n*\sigma^2 - n*\sigma^2/n$ (the var of each Xi is σ^2, while the var of $\bar{X}$ is σ^2/n)

(15) $= (n-1)\ \sigma^2$

Plugging back in we get: $E(s^2) = (n-1)\ \sigma^2/(n-1) = \sigma^2$.

The End of this proof.

The Rao-Cramer Lower Bound (RCLB)

For all unbiased estimators $\hat{\theta}$ of θ,

$$\text{Var}(\hat{\theta}) \geq \frac{1}{n*E\left[\left(\frac{\partial \ln(f(x;\theta))}{\partial\theta}\right)^2\right]}$$

n is the sample size; θ is the parameter being estimated; x is the random variable

Let's find the RCLB for μ from a ND(μ,σ) population. We start with the normal pdf, f(x;μ), then we work our way inside-out on that expression above.

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

2) $\ln(f(x;\mu) = -1/2*\ln(2\pi) - \ln(\sigma) - (x-\mu)^2/(2\sigma^2)$

3) $\partial(\ln(f(x;\mu)))/\partial\mu = 0 + 0 - 2(x-\mu)(-1)/(2\sigma^2) = (x-\mu)/\sigma^2$

4) The square of that thing right above = $(x-\mu)^2/\sigma^4$

5) The exp value of that thing right above = $E[(x-\mu)^2/\sigma^4] = 1/\sigma^4*E(x-\mu)^2$
But that very last thing is just asking for the variance of X, which is σ^2, since our population was ND(μ,σ). So the expression = $(1/\sigma^4)\sigma^2 = 1/\sigma^2$

6) Multiply the thing right above by n, so n/σ^2 .

7) But that thing right above is in the denominator, so flip it over to get

RCLB = σ^2/n.

Conclusion: The very least variance any unbiased estimator can have is σ^2/n. So if we can find some unbiased estimator whose variance equals σ^2/n, then we have found the minimum variance unbiased estimator: the best of all unbiased estimators.

But we know that $\bar{X}$ has a variance of σ^2/n, so $\bar{X}$ must be the MVUE!

Final note: While s^2 is unbiased for σ^2, it is not in fact the MVUE. In fact, it is still an open question: What unbiased estimator for σ^2 is the MVUE?
(It is possible that this has changed recently and I haven't received the news yet, but I'm not sure this is a problem anyone is working on.)