Stats 135, Fall 2019
Lecture 6, Wednesday, 9/11/2019

CLASS ANNOUNCEMENTS: Quiz 1 on Friday will have 3 problems: 2 problems similar to HW1,2 or up to lecture 5. There will be 1 problem of MOM calulation.

1 Review:

Last time, we covered §8.4. We showed how to compute SE by hand. To find $\mathrm{Var}(\hat{\theta})$, we often need to know $\sigma^2 = \mathrm{Var}(x)$ which is a function of θ. We have $\sigma^2(\theta) \approx \sigma^2(\hat{\theta})$.

Example:
Given an iid sample, we collect data:

$$x_1 4, x_2 = 7, x_3 = 4, x_4 = 2, x_5 = 3,$$

which follows a Poisson(λ) distribution. Then we find a MOM estimator of λ and approximate the SE of our estimate.

Solution.

$$\begin{aligned} X &\sim \text{Poisson}(\lambda) \\ \mu_1 &= \mathbb{E}(X) = \lambda \\ \lambda = \mu_1 \implies \hat{\lambda} &= \overline{x} = 4 \end{aligned}$$

Using properties of the Poisson distribution (variance is simply λ), we have:

$$\begin{aligned} SE(\hat{\lambda}) &= \sqrt{\mathrm{Var}(\hat{\lambda})} \\ &= \sqrt{\mathrm{Var}(\overline{x})} \\ &= \sqrt{\frac{\mathrm{Var}(x)}{n}} \\ &= \sqrt{\frac{\lambda}{n}} \\ &\approx \sqrt{\frac{\hat{\lambda}}{n}} \\ &= \sqrt{4/5}. \end{aligned}$$

Alternatively, at the 3rd equality we could have taken the sample variance instead of $\hat{\lambda}$ as an estimate of $\mathrm{Var}(x)$ above. That is,

$$S^2 = \cdots$$

□

Topics Today:

- Empirical cdf
- §8.4 Example
- §8.4, 8.4.6.

2 Empirical CDF (p 378 Rice)

This is actually in Chapter 10, but we talk about it briefly for justification of the bootstrap method.

(insert example here)

Facts about ECDF.
(1) F_n is an unbiased estimator of F.
(2) $\text{Var}(F_n) \to 0$ as $n \to \infty$.
So for large sample size, the empirical cdf is a good approximation of the population cdf. Now when we bootstrap, we take from the 'staircase' as opposed to the smooth curve population.

3 Another example of MOM estimator and computing the SE

This is #4ab from Chapter 8. Suppose X is a discrete random variable with:

$$\begin{aligned}
\mathbb{P}(X=0) &= \frac{2}{3}\theta \\
\mathbb{P}(X=1) &= \frac{1}{3}\theta \\
\mathbb{P}(X=2) &= \frac{2}{3}(1-\theta) \\
\mathbb{P}(X=3) &= \frac{1}{3}(1-\theta).
\end{aligned}$$

Then the following 10 iid observations are taken, giving us:

$$3, 0, 2, 1, 3, 2, 1, 0, 2, 1.$$

We are tasked to find the MOM estimator of θ and approximate the SE of our estimate. We have one parameters, so our first step is to compute the first moment:

$$\begin{aligned}
\mathbb{E}(X) &= 0 \cdot \mathbb{P}(X=0) + 1 \cdot \mathbb{P}(X=1) + 2 \cdot \mathbb{P}(X=2) + 3 \cdot \mathbb{P}(X=3) \\
&= \frac{1}{3}\theta + 2 \cdot \frac{2}{3}(1-\theta) + 3 \cdot \frac{1}{3}(1-\theta) \\
&= \frac{7}{3} - 2\theta
\end{aligned}$$

This implies:

$$\begin{aligned}
\theta &= \frac{1}{2}\left(\frac{7}{3} - \mu_1\right) \\
\hat{\theta} &= \frac{1}{2}\left(\frac{7}{3} - \underbrace{\hat{\mu}_3}_{=\overline{X}=\frac{3}{2}}\right) = \frac{5}{12}.
\end{aligned}$$

Next we find the variance $\text{Var}(\hat{\theta})$. From our finding earlier,

$$\begin{aligned}
\text{Var}(\hat{\theta}) &= \text{Var}\left(\frac{1}{2}\left(\frac{7}{3} - \overline{X}\right)\right) \\
&= \frac{1}{4}\text{Var}(\overline{X}) \\
&= \frac{1}{4}\frac{\text{Var}(X)}{10}
\end{aligned}$$

Recall that in our previous example, we had a Poisson distribution, and we knew the variance is just λ. Now here we have an unknown distribution, and we don't know the variance of the top of our head. There are two ways to find the variance.
(1) Approximate $\text{Var}(X)$ by s^2. Taking this approach, we get that $SE(\hat{\theta}) = .171$ (done in R).

```
> sqrt( 1/(4*10) * var( c(3,0,2,1,3,2,1,0,2,1)) )
[1] 0.1707825
```

(2) Analytically compute $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, in terms of θ. Lucas jokes that he takes out his Pitman book for this. Here we get that $SE(\hat{\theta}) = .173$.

4 Yet Another Example

Let $X \sim \text{Exponential}(\lambda)$. Fact: $\mathbb{E}(X) = \frac{1}{\lambda}$. We are tasked to find $\hat{\lambda}$ and $SE(\hat{\lambda})$.
Then

$$\mu_1 = \frac{1}{\lambda} \implies \lambda = \frac{1}{\mu_1} \implies \boxed{\hat{\lambda} = \frac{1}{X}}$$

Here we don't have the variance of a linear combination of μ_1, so here we need the δ-method instead.

5 δ-method

Consider a random variable X with a known mean μ and known SD σ. We have some function $Y = g(X)$, smooth around the mean, for example it can be $1/X$. That is, $g'(\mu) \neq 0$. The idea comes from Taylor expansion. Taylor had the idea that all smooth functions are locally linear. We'll make a linear Taylor approximation around μ (truncation) here. Take:

$$Y = g(X) \approx g(\mu) + g'(\mu)(X - \mu),$$

and this is a good approximation when $(X - \mu)$ is very small (and hence exponentials of them, the truncated terms, are even smaller). In other words, this is a good approximation when $X \approx \mu$.
If we can write it this way, then applying Var across the equation gives:

$$\begin{aligned}\text{Var}(Y) &\approx \text{Var}\left(g(\mu) + g'(\mu)(X - \mu)\right) \\ &= \left(g'(\mu)\right)^2 \underbrace{\text{Var}(X)}_{\sigma^2}.\end{aligned}$$

Now we may ask: what random variables have this property? That is, what random variables have a small SD? Surely, the normal or uniform distributions won't work. Instead, take $\overline{X}$ with large n, which Lucas notes is very 'pointy' around the mean.

Theorem 5.1. (δ-method). Let $X_1, \ldots, X_n$ be iid with mean μ and $SD = \sigma$. Take g to be smooth μ, where $g'(\mu) \neq 0$. Then

$$\text{Var}(g(\overline{X})) \approx (g'(\mu))^2 \cdot \frac{\sigma^2}{n}$$

Now for us, take: $\hat{\theta} := g(\overline{X})$. Let's finish the earlier example.

Example: $X \sim \text{Exponential}(\lambda)$. Then the estimator is $\hat{\lambda} = \frac{1}{\overline{X}}$, so let:

$$g(\overline{X}) := \frac{1}{\overline{X}},$$

in the δ-method. By the theorem, this method gives

$$\text{Var}(g(\overline{X})) \approx (g'(\mu))^2 \cdot \frac{\sigma^2}{n}.$$

All we need to do is compute the derivative and plug it in, evaluated at μ.

$$g'(X) = \frac{-1}{X^2} \implies (g'(\mu))^2 = \left(\frac{-1}{\mu^2}\right)^2 = \lambda^4,$$

where we used $\mu = \frac{1}{\lambda}$.
So the variance of our MOM estimator is:

$$\text{Var}(\hat{\lambda}) = \text{Var}(g(\overline{X})) = \lambda^4 \cdot \frac{\frac{1}{\lambda^2}}{n}.$$

We don't know λ, so we plug in $\hat{\lambda} = \frac{1}{\overline{X}}$. Then

$$SE(\hat{\lambda}) \approx \frac{\lambda}{\sqrt{n}} \approx \frac{\frac{1}{\overline{X}}}{\sqrt{n}} = \frac{1}{\sqrt{n} \cdot \overline{X}}.$$

Example: This one is problem 52 from Chapter 8, and it is a bit more complicated. Let's say we have $X_1, \dots, X_n$ iid random variables with density $f(X|\theta) = (\theta+1)X^{\theta}$, where $0 \le X \le 1$. We're tasked to find $\hat{\theta}$ and use the δ-method to approximate $SE(\hat{\theta})$.

Solution. First, we find our MOM estimator $\hat{\theta}$. We have only one parameter, so we need only the first moment:

$$\mathbb{E}(X) = \int_0^1 (\theta+1)X^{\theta+1}\, dx = \frac{\theta+1}{\theta+2}X^{\theta+2}|_0^1 = \frac{\theta+1}{\theta+2}$$

Now this is our $\mu = \frac{\theta+1}{\theta+2}$. We do some algebra on

$$\begin{aligned}
\mu\theta + 2\mu &= \theta + 1 \\
\mu\theta - \theta &= 1 - 2\mu \\
\theta(\mu - 1) &= 1 - 2\mu \\
\theta &= \frac{1-2\mu}{\mu - 1}
\end{aligned}$$

Now plugging in $\mu = \overline{X}$ gives:

$$\hat{\theta} = \frac{1 - 2\overline{X}}{\overline{X} - 1}.$$

Notice this is a function of $\overline{X}$, so we should be able to use the δ method. We'll do this during the next lecture. □