Stats 135, Fall 2019
Lecture 5, Monday, 9/9/2019

CLASS ANNOUNCEMENTS: Quiz 1 on Friday. There will be 3 questions total: 2 problems similar to those in HW 1 and 2, and 1 MOM calculation.

1 Review

Last time, in §8.4, we found that MOM estimators are **consistent** when h is continuous. That is,

$$\hat{\theta}_{MOM} = h(\hat{\mu}_1, \ldots, \hat{\mu}_l).$$

Because the sample moment converges in probability to the true moment,

$$\underbrace{\hat{\mu}_k}_{\frac{1}{n}\sum_{i=1}^{n} x^k} \xrightarrow{p} \underbrace{\mu_k}_{\mathbb{E}(X^2)}$$

(that is the sample parameter converges in probability to the true parameter) via generalized weak law of large numbers, we have:

$$\hat{\theta} = h(\hat{\mu}_1, \ldots, \hat{\mu}_l) \xrightarrow{p} \theta = h(\mu_1, \ldots, \mu_l)$$

when h is continuous.

Topics Today:

- Example of MOM calculation
- p 264 : nonparametric bootstrap for 95% confidence interval (done in R last Friday in lab)
- p 262 : finding SE of $(\hat{\theta})$ by hand

2 §8.4 MOM

Recall that the density of Gamma is given by:

$$f(x) = \frac{\lambda^r}{\Gamma(r)} \underbrace{x^{r-1}e^{-\lambda x}}_{\text{variable}},$$

where $\Gamma(r) = r - 1$ when $r \in \mathbb{Z}^+$.

$$\int_0^\infty f(x)\ dx = 1 \implies \int_0^\infty \frac{\lambda^r}{\Gamma(r)} x^{r-1}e^{-\lambda x}\ dx = 1$$
$$\implies \boxed{\int_0^\infty x^{r-1}e^{-\lambda x}\ dx = \frac{\Gamma(r)}{\lambda^r}},$$

which follows from a useful identity when r is an integer (so we don't need integration by parts).
Consider an iid sample of random variables with density

$$f(x|\sigma) = \frac{1}{2\sigma} e^{\left(\frac{-|x|}{\sigma}\right)}, \quad \sigma > 0.$$

We want to find the MOM estimator $\hat{\sigma}$ of σ, so we calculate the fist moment $\mathbb{E}(x)$. That is,

$$\mathbb{E}(x) = \int_{-\infty}^{\infty} xf(x|\sigma)\ dx = \int_{-\infty}^{\infty} \overbrace{x\frac{1}{2\sigma}e^{\left(-\frac{|x|}{\sigma}\right)}}^{g(x)}\ dx = 0$$

where $g(-x) = -g(x)$ shows g is an odd function, and hence this integral equals zero. More precisely,

$$\int_{-\infty}^{0} xf(x|\sigma)\ dx = -\int_{0}^{\infty} xf(x|\sigma)\ dx$$

Our function is 0, so this isn't helpful. We try the next moment, μ_2, $\mathbb{E}(x^2)$. Lucas tasks us to compute this.

$$\begin{aligned}
\mathbb{E}(x^2) &= \int_{-\infty}^{\infty} x^2 f(x|\sigma)\ dx \\
&= \int_{-\infty}^{\infty} x^2 \frac{1}{2\sigma}e^{\left(-\frac{|x|}{\sigma}\right)}\ dx \\
&= 2\frac{1}{2\sigma}\int_{0}^{\infty} x^2 e^{-\frac{1}{\sigma}x}\ dx \\
&= \frac{1}{\sigma}\cdot\frac{\Gamma(3)}{(1/\sigma)^3} = \boxed{2\sigma^2}.
\end{aligned}$$

Here we used the boxed formula we found above, $\int_0^\infty x^{r-1}e^{-\lambda x}\ dx = \frac{\Gamma(r)}{\lambda^r}$, with $r = 3$ and $\lambda = \frac{1}{\sigma}$.
Recall that the second step to this Method of Moments calculation is to rewrite the parameter in terms of the moments. That gives us:

$$\sigma = \sqrt{\frac{\mu_2}{2}},$$

and for Step 3, we plut in the sample moment (put hats on) to get:

$$\hat{\sigma} = \sqrt{\frac{\hat{\mu}_2}{2}},$$

where $\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$.

3 Nonparametric Bootstrapping and Confidence Interval of Population Parameter

See p284-285 in Rice. Recall that to calculate the 95% CI of a population mean **by hand**, we need two things:

- a Simple Random Sample (We need an SRS so that we know $\mathrm{Var}(\overline{x})$, and we have a formula for that. It may be unrealistic to assume sampling without replacement.)
- a large sample size (so that the distribution is approximately normal so that we know the 2.5th and 97.5th quantiles of $\overline{x}$.)

However, it is often the case that the sample is small and perhaps we don't have an SRS. Instead, we perform a boostrap technique. Instead, we can find a 95% CI by bootstrapping in R **without any assumptions on the sampling distribution**, which is useful.

To talk about this, Lucas gives some background on the 95% CI of a population parameter. We did this in detail for the mean before, and now we want to generalize. Let θ be a (generalized) parameter of our population (for example it could be the population average, population median, or the population SD, etc).
Let $\hat{\theta}$ be an estimator of θ. Of course, $\hat{\theta}$ is a random variable, and θ is an unknown constant. We look at:

$$\hat{\theta} - \theta,$$

which is also a random variable. Its distribution is **not necessarily normal**. Suppose we want to compute points a, b that give the 2.5th and 97.5th percentiles of $\hat{\theta} - \theta$. In other words,

$$\mathbb{P}(a < \hat{\theta} - \theta < b) = 95\%.$$

As justification for this formatting, Lucas notes that we can compute $a, b, \hat{\theta}$ from our sample. It only takes a little algebra to get:

$$\iff \mathbb{P}(-\hat{\theta} + a < -\theta < -\hat{\theta} + b) = 95\%$$
$$\iff \mathbb{P}(\hat{\theta} - b < \theta < \hat{\theta} - a) = 95\%,$$

which we call the 95% CI of θ. This distribution of $\hat{\theta} - \theta$ is only known to 'Tyche', the god of fortune (we don't know anything about θ), so there is no way for us to know what a, b are. Lucas notes that this is okay, because we are going to make a simulation (nonparametric bootstrap) and estimate these. What this is is that we'll take a sample once from our population and compute our estimator for that sample, and set this value to θ. Once we have this, we will sample from that same sample (same size) "many many times" **with replacement** to fill out our distribution. We can approximate $\hat{\theta} - \theta$ via bootstrapping:

4 Bootstrapping a 95% CI of θ:

(Step 1) Take a sample $x_1, \ldots, x_n$ of size n one time from your population. We will calculate $\hat{\theta}$ (which is θ above).

(Step 2) Now we **resample** from this same sample (size n), say $B = 1000$ times, but now **with replacement**. Now we compute the estimator of θ each time. This will give us a list of $B = 1000$ numbers, $\theta_1^*, \theta_2^*, \ldots, \theta_B^*$.

(Step 3) Now we subtract $\hat{\theta}$ (from step 1) from each θ_i^*, which gives us the list:

$$\theta_1^* - \hat{\theta},\ \theta_2^* - \hat{\theta},\ \ldots,\ \theta_B^* - \hat{\theta}$$

This is $\hat{\theta} - \theta$ above. Lucas apologizes there are multiple different $\hat{\theta}$ 'running around'. This will fill our distribution for $\hat{\theta} - \theta$ (and we can sketch the graph).

(Step 4) Now we find the 25th and 975th largest values from your (ordered) list of 1000 resulting numbers in step 3. This is an approximation of the 2.5th and 97.5th percentile of $\hat{\theta} - \theta$.

(Step 5) The 95% CI of θ then is:

$$(\hat{\theta} - b, \hat{\theta} - a),$$

where $\hat{\theta}$ is our original sample estimator of θ, and a is our 25th largest number in the list of $\theta_i^* - \hat{\theta}$, and b is our 975th largest. The question arises for a rule of thumb for the value of B, and Lucas holds off answering because we'll talk about why this method works and produces good results.

5 §8.4: SE of $\hat{\theta}$

Recall that θ is a parameter **determining** the distribution of x. So the $SD(x) = \theta$ is a continuous function of θ, so we can write $\sigma(\theta)$.
For example, if $x \sim \text{Poisson}(\lambda)$, then the $SD(\lambda) = \sqrt{\lambda}$, and so:

$$\sigma(\lambda) = \sqrt{\lambda},$$

which is a continuous function. Now because σ is continuous, then we've found that

$$\hat{\theta} \xrightarrow{p} \theta \implies \sigma(\hat{\theta}) \xrightarrow{p} \sigma(\theta).$$

So for large n, unknown $\sigma(\theta)$ is very closely approximated by $\sigma(\hat{\theta})$, which is known. We don't know the true SD of x, but we can bootstrap an estimator.

Lecture ends here.

We'll look at an example of this at the start of Wednesday's lecture.