Stats 135, Fall 2019
Lecture 8, Monday, 9/16/2019

1 Review

Last time, we went over §8.5, the method of Maximum Likelihood, which is an optimization problem and typically involves calculus. There are some technical assumptions we need to make about our random variables in order for an MLE to exist. These are called **regularity assumptions**. Our text treats these very lightly, and Lucas emphasizes the same points as the book:

R0. The pdfs are distinct (i.e. $\theta \neq \theta' \implies f(x_i \mid \theta) \neq f(x_i \mid \theta')$) That is, if we have two different parameters, then their pdfs are distinct.

R1. The pdfs have common support for all θ. Support is where the function equals zero. (i.e. $\{x \mid f(x \mid \theta) \neq 0\} = \{x \mid f(x \mid \theta') \neq 0\}$.

R2. The true parameter θ_0 is an interior point in the parameter space.

Theorem 1.1. Let θ_0 be the true parameter. Under assumptions (R0) and (R1),

$$\lim_{n\to\infty} \ldots$$

In other words, for large enough sample size, we can find the MLE.

For example, we look at the normal distribution with the following log-likelihood:

$$l(\mu, \sigma) = -n \log(\sigma) - \frac{n}{2} \ldots$$

Now unlike in Gamma, we can actually solve for our MLE estimators. We simply have:

$$\hat{\mu} = \overline{X}$$

and

$$\hat{\theta} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2},$$

where we do NOT have an $n-1$ in the denominator. Often we want an estimator for σ^2, and this is very easy via the **equivariance** property of MLE estimators. That is,

$$\overline{\sigma}^2 = \overline{\sigma}^2_{ML} = \frac{1}{n}\sum(X_i - \overline{X})^2.$$

Topics Today:

- §8.5 - Equivariance
- §8.5 - Consistency
- §6.2 - t distribution and $\chi^2(n)$ distribution

In the interest of time Lucas notes that we'll go through consistency rather quickly, so that we may do t-distribution and $\chi^2(n)$ distributions more carefully (as they may be in exams or homework).

2 Equivariance

A nice property of the MLE $\hat{\theta}_{ML}$ is **equivariance** in that

$$g(\hat{\theta}) = ...$$

Theorem 2.1. If $\hat{\theta}_{ML}$ is a MLE of θ and g is a function (need not be monotonic) then $g(\hat{\theta}_{ML})$ is an MLE of $g(\theta)$.

3 Consistency

Recall that this means that our estimator converges in probability to the true parameter. The picture that we should have in our heads is that we have a list of parameters

$$\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n,$$

where for larger sample sizes, our estimator will be more and more pointed and converge to the true value θ_0. Lucas puts it as that we become more and more certain of what that value is.

Theorem 3.1. (Consistency.) Assume that $X_1, \dots, X_n$ satisfy ergularity conditions (R0), (R1), (R2), where θ_0 is the true parameter, and further assume that $f(x \mid \theta)$ is differentiable with respect to θ. Then $l'(\theta) = 0$ has at least a solution $\hat{\theta}_n$ with the property that $\hat{\theta}_n \xrightarrow{p} \theta$.
If $l'(\theta) = 0$ has a unique solution, then it is a consistent estimator θ, so we need not check that our parameter has a maximum.

If n is infinitely large, there still may be several solutions. We would check each solution and see which gives the maximum likelihood. Lucas notes that for our purposes, we need only understand how this proof works (we will not be held responsible for the proof's details), but for time we will put it in the appendix and skip onwards.

4 §6.2: t distribution and χ^2 distribution

Suppose $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

Motivation: In chapter 7, we showed that for large n,

$$\overline{x} \pm 1.96z(.025)$$

is a 95% confidence interval for μ. We needed the averages to be approximately normal. However, to talk about a confidence interval for small n, we need a t distribution to make this precise. Furthermore, to find a 95% CI for σ^2, we will need a χ^2 distribution.

The goal for today is to provide background on these two distributions.

4.1 χ^2 distribution

We define this by starting off with a standard normal distribution:

$$Z \sim N(0,1).$$

Now define:

$$Z^2 \sim \chi_1^2,$$

where the subscript 1 is the 'degree(s) of freedom', here showing that there is only one variable.
Now we should be able to verify: $Z^2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$. If we know the density for Z, we can find the density of $g(Z)$ via the change of variable probability theorem. Lucas will add the details to this later.

Definition: χ_n^2 **-**

Let $Z_1, \ldots, Z_n \overset{iid}{\sim} N(0,1)$.
Then:

$$Z_1^2 + \cdots + Z_n^2 \sim \chi_n^2,$$

which is the chi-squared with n degrees of freedom.

Fact:

$$Z_1^2 + \cdots + Z_n^2 \sim \text{Gamma}(\frac{\text{n}}{2}, \frac{1}{2}).$$

To see this, recall that the sum of n iid Gamma(r, α) is also Gamma(nr, α).

Lucas adds:

$$\mathbb{E}(\chi_k^2) = k$$

As $k \to \infty$, this χ_k^2 distribution looks more and more normal.

Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Then

$$\sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2 \sim \chi_n^2,$$

so

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \sim \chi_n^2.$$

We see (this is very important) that:

$$\frac{n-1}{\sigma^2} s^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Xi - \overline{X})^2 \sim \chi_{n-1}^2.$$

Somehow, we lose a degree of freedom. Why is this the case? Basically, what this says is that the standard **variance is a multiple** of χ_{n-1}^2. Lucas notes that this is a simple way to think about the sample variance.

Intuitively, we would expect to lose a degree of freedom, whereas

$$\{X_1 - \mu, X_2 - \mu, \ldots, X_n - \mu\}$$

are independent. However,

$$\{X_1 - \overline{X}, x_2 - \overline{X}, \ldots, X_n - \overline{X}\}$$

ARE dependent, because if we add them all up and distribute the sum, we have:

$$\sum_{i=1}^{n}(X_i - \overline{X}) = \sum_{i=1}^{n} X_i - n \cdot \overline{X} = 0.$$

There is a dependency between these guys, so it somewhat makes sence that we'll lose one degree of freedom. It's a good thing to make this rigorous. We provide a sketch of a more rigorous proof.

Theorem 4.1. Let $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Then

$$\frac{(n-1)}{\sigma^2} s^2 \sim \chi^2_{n-1}.$$

Proof. The proof of this is given in Rice Chapter 6 on page 197. Basically,

$$\begin{aligned}
\frac{(n-1)s^2}{\sigma^2} &= \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 \\
&= \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma} - \frac{\overline{X} - \mu}{\sigma}\right)^2 \\
&= \sum_{i=1}^{n}\left(Z_i - \overline{Z}\right)^2, \text{ where } Z_i := \frac{x_i - \mu}{\sigma^2}, \text{ and } \overline{Z} := \frac{1}{n}\sum Z_i = \frac{\overline{X} - \mu}{\sigma} \\
&= \sum_{i=1}^{n} Z_i^2 - 2\overline{Z}\underbrace{\sum_{i=1}^{n} Z_i}_{=n\overline{Z}} + \underbrace{\sum_{i=1}^{n} \overline{Z}^2}_{=n\overline{Z}^2} \\
&= \sum_{i=1}^{n} Z_i^2 - n\overline{Z}^2.
\end{aligned}$$

Then rearranging gives:

$$\sum_{i=1}^{n} Z_i^2 = \frac{(n-1)s^2}{\sigma^2} + n\overline{Z}^2,$$

where the LHS is a chi-squared distribution with n degrees of freedom. In other words, $\chi^2_n = \text{Gamma}(\frac{\text{n}}{2}, \frac{1}{2})$. Now for the second term on the RHS, $n\overline{Z}^2$, recall that $\overline{Z} = \frac{\overline{X} - \mu}{\sigma}$. Then $\sqrt{n}\overline{Z}$ takes on the standard normal. Hence $n\overline{Z}^2$ takes on the $\chi^2_1 = \text{Gamma}(\frac{1}{2}, \frac{1}{2})$ distribution.
Additionally, by Theorem A, p. 195 of Rice, we have that:

$$\overline{X} \text{ and } s^2 \text{ are independent random variables for } X_i \sim N(\mu, \sigma^2).$$

It follows that $\frac{(n-1)s^2}{\sigma^2}$ and $n\overline{Z}^2$ are independent. Now because the sum of independent gamma is gamma, and $\sum_{i=1}^{n} Z_i^2$ is Gamma$(\frac{\text{n}}{2}, \frac{1}{2})$, we conclude:

$$\frac{(n-1)s^2}{\sigma^2}$$

must be Gamma$(\frac{\text{n}-1}{2}, \frac{1}{2}) = \chi^2_{\text{n}-1}$.

□

Lecture ends here.

Next time, we'll look at the t-distribution.