STA261: Lecture 8

Comparing Likelihoods

Alex Stringer

July 30th, 2018

Disclaimer

The materials in these slides are intended to be a companion to the course textbook, *Mathematical Statistics and Data Analysis, Third Edition*, by John A Rice. Material in the slides may or may not be taken directly from this source. These slides were organized and typeset by Alex Stringer.

A big thanks to Jerry Brunner as well for providing inspiration for assignment questions.

License

Comparing Likelihoods

We have put forward the idea of maximizing the likelihood in order to find good estimates of parameters.

This gives the value of the parameter that is "most likely" to have generated the observed data.

What about other values? Are there other values that are almost as likely to have generated the observed data?

Or: there is a true value θ_0 that *actually did* generate the observed data. We know we didn't get it perfect in the sample, i.e. $\hat{\theta} \neq \theta_0$. But hopefully under our procedure, θ_0 is at least going to be considered *likely* to have generated the observed data.

So we should be able to plug a range of values for θ into the log-likelihood, and find that they are all "pretty likely".

Comparing Likelihoods

Let's start with the finite-finite case (in terms of parameter space and sample space) to illustrate this.

Suppose I have two coins in my pocket, one with $P(\text{heads}) = 0.7$ and one with $P(\text{heads}) = 0.3$.

In probability I'd ask: suppose I pick the first one, what values could we see, and with what probabilities?

...

Comparing Likelihoods

In statistics, we flip this around and say: suppose I saw a heads. What coin did I flip?

...

Well, it could have been either coin, couldn't it? A head still happens with probability 30%, even with the tail-friendly coin.

But the first coin is *more likely* to have been flipped. The coin with the higher probability of heads *had a higher chance of generating the observed data* than the coin with the lower probability of heads.

Comparing Likelihoods

Let $X : \{T, H\} \to \{0, 1\}$ be the random variable taking on value 1 if the coin is heads and 0 if tails. We observe the single value x, and we don't know which coin was flipped.

If the first coin was flipped, the likelihood is $L(\theta_1|x) = 0.7^x 0.3^{1-x}$, a bernoulli with $\theta_1 = 0.7$.

If the second coin was flipped, the likelihood is $L(\theta_2|x) = 0.3^x 0.7^{1-x}$, a bernoulli with $\theta_2 = 0.3$.

Comparing Likelihoods

For our observed data, we have $L(\theta_1|x=1)=0.7$ and $L(\theta_2|x=1)=0.3$.

Form the **likelihood ratio**:

$$\Lambda(x)=\frac{L(\theta_1|x)}{L(\theta_2|x)}$$

The likelihood ratio $\Lambda(x)$ is how many times more likely θ_1 is to have generated the observed data than θ_2 was.

Here, $\Lambda(1)=0.7/0.3\approx 2.33$. Given that we observed a head, it is about 2.33 times more likely that the first coin was the coin that was thrown.

Comparing Likelihoods

That seems pretty conclusive. What if we threw the coin more times?

...

Wait Times

A more realistic setting is where we have one candidate value of a parameter, and we wish to see if the data supports this value.

Suppose the TTC claims 5 minute service at a particular time of day for a bus route. You're waiting for the bus, and it takes 7 minutes.

Is the TTC blowing smoke? A reasonable model for wait times is $X \sim Exp(\theta)$ with $E(X) = \theta$.

...

Wait Times

Okay, maybe once is too little to tell. You wait for the bus every day all week, and it comes in $(7, 5, 3, 9, 6)$ minutes each day. Is the 5 minute claim reasonable?

...

Likelihood Ratios

This is an example of asking whether the observed data supports a particular value of the parameter.

How likely was this value of the parameter to generate the observed data? How much *less* likely is it than the *most* likely value, the MLE?

Likelihood ratio:

$$\Lambda(\mathbf{x}) = \frac{L(\theta_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

where θ_0 is our *candidate value* of θ (now not necessarily the *true* value), and $\hat{\theta}$ is the MLE.

$0 < \Lambda(\mathbf{x}) \leq 1.$

Likelihood Ratios

For the exponential example, we had $\hat{\theta} = \bar{x} = 6$ and we computed

$$\Lambda(\mathbf{x} = (7,5,3,9,6)) = 0.915$$

so the candidate value is about 91% as likely to have generated the observed data as the MLE was.

Is this enough to conclude that the candidate value *could* be the true value?

Distribution of the LRT, one parameter case

Theorem: suppose θ_0 is the true value of parameter $\theta \in \Omega$, and all the regularity conditions for the CLT for the MLE hold. Then for the procedure above,

$$-2\log\Lambda(\mathbf{X}) \xrightarrow{d} \chi_1^2$$

Proof: ...

Distribution of the LR statistic, one parameter case

This gives us an explicit way of deciding whether a candidate value of a parameter might reasonably be said to be the *true value*- the value that generated the data we observed.

We can compute

$$P(-2\log\Lambda(\mathbf{X}) > -2\log\Lambda(\mathbf{x}))$$

that is, the probability of seeing an even larger value of $-2\log\Lambda(\mathbf{X})$, if θ_0 is actually the true value of θ. If this probability is low, then either θ_0 isn't really the true value, or we saw a sample that was very improbable.

For the LRT above...

Example: normal distribution

Suppose $X_i \overset{IID}{\sim} N(\mu, \sigma^2)$, with σ^2 known. Derive a likelihood ratio procedure for evaluating whether $\mu = \mu_0$.

...

Example: normal distribution

Now suppose σ^2 is not known. We estimate it with its MLE, *which depends on* μ, and then plug this in to the top and bottom of the likelihood ratio.

...

Distribution of the LR statistic, multi parameter case

We didn't cover the case where we have more than one parameter.

Let Ω represent the full or *unrestricted* parameter space. Above, $\Omega = \mathbb{R} \times \mathbb{R}^+$. Let $p = \dim\Omega$ be the number of *free parameters* in Ω. Above, $p = 2$.

Let Ω_0 represent the *restricted* parameter space, with parameter(s) fixed at their candidate value(s). Above, $\Omega_0 = \{\mu_0\} \times \mathbb{R}^+$, because we're fixing μ at μ_0, but still estimating σ^2. Let $d = \dim\Omega_0$ be the number of *free parameters* in Ω_0. Above, $d = 1$. Then

$$-2\log\Lambda(\mathbf{X}) \xrightarrow{d} \chi^2_{p-d}$$