

MATH 2B RECITATION 2/16/12

ALDEN WALKER

1. TESTING HYPOTHESES SETUP

I highly recommend the notes on the course website.

Just like with estimators, we are going to try to figure something out about the parameters of a distribution just from observing the distribution. In this setup, we pick two hypotheses which completely fill the possible parameter space. In other words, we know the mean is either 0 or 1, so we pick the hypotheses “it’s 1” and “it’s 0”. Or perhaps we don’t know the mean at all, but we’re interested in 0, so we pick “it’s 0” and “it’s not 0”.

However, the situation isn’t symmetric: there is a special (“null”) hypothesis on which the test is focused, so we will either accept or reject the null hypothesis.

1.1. **Some Definitions.** You pick a null hypothesis H_0 :

- The probability of rejecting H_0 when it is true is the **significance level**, usually denoted α .
- If the probability of accepting H_0 when it is false is β , then the **power** of the test is $1 - \beta$ (the probability of rejecting it when it is, in fact, false). It is common to ask about the power of the test “at” a value of the parameter you are estimating. This means that assuming the true value of the parameter is given (and it’s not in the null hypothesis), calculate the probability of rejecting the null (and take one minus that).
- The test **statistic** is the number that you compute from your data (like the sample mean, for example), and determining whether or not that number lies in a certain range is your “test.”
- To “determine critical values” for a test with a given test statistic T and significance α means to find the values of T which should indicate acceptance or rejection of the null hypothesis in such a way that the probability of rejecting the null hypothesis when it is true is α .
- To choose which hypothesis gets to be the null hypothesis, you pick the one that you would want to favor when you are unsure. For example, typically you set α quite small. This means that near the null hypothesis you will default to the null hypothesis. That is, it is the one that you assume unless it is proven wrong.

It is often the case that you want to choose the “bad” thing as the null hypothesis because the consequences of incorrectly assuming something is good are worse. For example, if you are testing a new drug to see if it is safe, the null hypothesis is: “it’s not safe.” That way, you can control (and make small) the probability of incorrectly rejecting the null (saying it’s safe incorrectly). This is clearly better because you want to be really sure that your new drug incur any lawsuits and such.

1.2. **Recipe.**

- (1) Get a parameter you are supposed to draw conclusions about
- (2) Choose hypotheses and a significance level
- (3) Choose a test statistic (likelihood ratio?)
- (4) Choose a rejection rule
- (5) Determine the constants in the rejection rule such that the significance level is what it’s supposed to be.

2. EXAMPLES (BASIC NORMAL VARIABLES)

Your test statistic is a single normal $(\mu, 1)$ variable X . The null hypothesis is “ $\mu = 0$ ” and the alternate is “ $\mu \neq 0$ ”. What are critical values of X if we want $\alpha = 0.01$?

This is not that hard: find C and D so that the probability that X lies outside the range (C, D) is 0.01, given that the null hypothesis is true. Thus, assume $\mu = 0$. Then using a table or Mathematica¹, we see that setting $C = -2.57$ and $T = 2.57$ is what we want, since $P(|X| \geq 2.57) \approx 0.01$.

See page 4 in the course notes for an example involving more than a single trial.

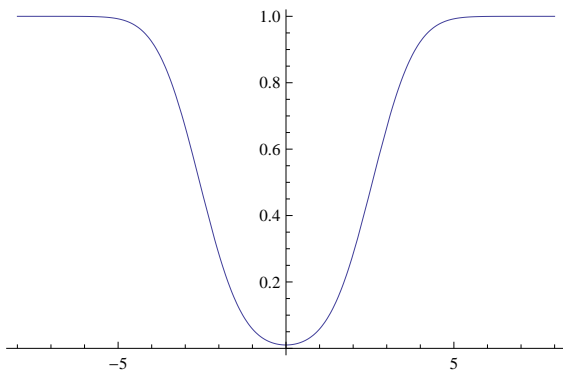
3. POWER

When setting up hypotheses, we control the probability α , which is the probability of a type 1 error, i.e. the probability of rejecting the null when it is actually true. Often, however, you would like to reject the null with reasonable probability when it is actually false. See the section on α - β specification, which talks about how to make sure you reject when the null is false. Here we will talk about just how to determine the power of a test. Let's look again at the above example.

The right question to ask is: suppose that the true value of μ is *not* 0 (so ideally we would reject). For a given true value of μ , what is the probability that we do, in fact, reject? This is the power at μ . We can just write down a formula for this probability; if we assume the true mean is μ , then let $\rho(\mu)$ be the probability of rejecting the null. Then

$$\begin{aligned} \rho(\mu) &= P(|X| \geq 2.57 \mid X \text{ is } (\mu, 1)) \\ &= P(X < -2.57 \mid X \text{ is } (\mu, 1)) + P(X > 2.57 \mid X \text{ is } (\mu, 1)) \\ &= \Phi(-2.57 - \mu) + (1 - \Phi(2.57 - \mu)) \end{aligned}$$

Here is a plot of $\rho(\mu)$:



Note that the height of the curve at 0 is α .

4. EXAMPLE (LIKELIHOOD RATIO)

The question of what statistic to use is often difficult to answer. It turns out that a very good statistic to use is the generalized likelihood ratio. Basically, this fraction measures how the probability under the null and the probability under the alternate of your observations compare. If we have samples $\{X_i\}_{i=1}^n$ from a distribution with density $f(x; \Theta)$ (i.e., with parameter Θ), then recall the definition of the likelihood function $L(\theta)$, and we define the likelihood ratio λ to be:

$$\lambda = \frac{\max_{\Theta \in \omega} L(\Theta)}{\max_{\Theta \in \Omega} L(\Theta)}$$

Where ω is the set of Θ values in the null hypothesis, and Ω is the full set of Θ values.

The generalized likelihood ratio test is the test which rejects the null if λ is too small, i.e. “reject if $\lambda < \lambda^*$ ”, where λ^* is chosen so that $P(\lambda < \lambda^* \mid \text{null}) = \alpha$.

¹You will find that you want to know the inverse of Φ to answer questions like: for what value x is $\Phi(x) = 0.5$. You can use the table in the back of Larsen and Marx, or you can ask Mathematica for `InverseCDF[NormalDistribution[], x]`.

4.1. **Example (6.5.3).** The question is: Let $\{y_i\}_{i=1}^n$ be random samples from a normal pdf with unknown mean μ and variance 1. Find the form of the GLRT for $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$.

Our statistic will be the likelihood ratio. We first need to maximize the likelihood function over the null parameter space ω . This is easy, since $\omega = \{\mu_0\}$ (it's just a single point). Next, we maximize the likelihood function over all possible values. We've already done this, though, and the maximum occurs at the mle! Thus, the maximum likelihood function value occurs when we just plug the mle into the likelihood function. This means that our likelihood ratio is (recall that the mle for the mean of a normal is the sample average):

$$\begin{aligned} \lambda &= \frac{\max_{\mu \in \omega} L(\mu)}{\max_{\mu \in \Omega} L(\mu)} \\ &= \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \mu_0)^2\right)}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \bar{y})^2\right)} \\ &= \exp\left(\frac{1}{2} \sum_{i=1}^n ((y_i - \bar{y})^2 - (y_i - \mu_0)^2)\right) \end{aligned}$$

So a good test would be to reject the null if this value were too small. To match this up with your intuition, note that to get a small value out of this, the thing inside the exp must be negative, which means that many of the y_i must be far from μ_0 (which indicates that μ_0 is probably not the mean).

5. EXAMPLE $((\alpha, \beta)$ SPECIFICATION)

You always want to control the significance α , but sometimes you want to set a value of the parameter in the alternative hypothesis space and say that if the true value is close to this chosen value, you want to make sure to reject the null. Let's see an example:

Let's say we have the distribution

$$f_{\Theta}(x) = \begin{cases} \Theta x^{\Theta-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

And we want to find Θ , or at least draw some conclusions about it. Our null hypothesis will be $\Theta < \Theta_0$, so H_1 is " $\Theta \geq \Theta_0$ ". Suppose that we want a significance level of $\alpha = 0.05$, but we also want the probability of accepting the null (incorrectly) when $\Theta = \Theta_1 > \Theta_0$ to be small (say $\beta = 0.05$). This is our α, β specification. How many samples do we need to take, and what should our critical point be? Also, what should our statistic be?

In this case, the likelihood ratio is

$$\prod_i \frac{\Theta_0 X_i^{\Theta_0-1}}{X_i^{\Theta_1-1} \Theta_1} = \left(\frac{\Theta_0}{\Theta_1}\right)^n \left(\prod_i X_i\right)^{\Theta_0-\Theta_1}$$

Warning: this is not exactly the GLR above – to make things simpler, I've put just the alternate hypothesis on the bottom (not the max over all possible parameters). Either you can convince yourself that this is essentially the same, or you can just take as given that we're going to use this statistic.

Note this depends bijectively with simply $\prod_i X_i$, so that is our test statistic. However, since $\Theta_1 > \Theta_0$, it's an orientation-reversing bijection. That is, we will reject the null if our statistic is *larger* than a constant C . Unfortunately, the distribution of $\prod_i X_i$ isn't too easy. However, now let's take the log and define our test function $T(X_i) = \sum_i \log(X_i)$. Of course, we haven't made the distribution easier, but now we can use the normal approximation (assuming n is fairly large). We need to know the mean and variance of $\log(X_i)$. The mean is $\int_0^1 \log(x) \Theta x^{\Theta-1} dx$. You can do this fairly easily with integration by parts and get $E(\log(X_i)) = -1/\Theta$. We can also compute $E(\log(X_i)^2) = \int_0^1 \log(x)^2 \Theta x^{\Theta-1} dx$, which comes out to $2/\Theta^2$. Therefore,

$$Var(\log(X_i)) = E(\log(X_i)^2) - E(\log(X_i))^2 = \frac{2}{\Theta^2} - \frac{1}{\Theta^2} = \frac{1}{\Theta^2}$$

Therefore, we can approximate the distribution of $\sum_i \log(X_i)$ by a normal $(n \cdot (-\frac{1}{\Theta}), n \cdot \frac{1}{\Theta^2})$ distribution. This holds for any Θ , so we now have two equations: for α , we have $P(\text{reject}|\text{null}) = P(\sum_i \log(X_i) > C|\text{null}) < 0.05$, and from β we get $P(\sum_i \log(X_i) < C | \Theta = \Theta_1) < 0.05$. This gives two equations in the unknowns n and C , which you can solve for, given any two particular Θ values.

To make this concrete, suppose $\Theta_0 = 1/4$ and $\Theta_1 = 3/4$. Then the null distribution is (approximately) $N(-4n, 16n)$, and the distribution for Θ_1 is $N(-(4/3)n, (16/9)n)$. Therefore

$$P\left(\sum_i \log(X_i) > C \mid \text{null}\right) \approx 1 - \Phi\left(\frac{C + 4n}{4\sqrt{n}}\right)$$

and

$$P\left(\sum_i \log(X_i) < C \mid \Theta = \Theta_1\right) \approx \Phi\left(\frac{C + (4/3)n}{(4/3)\sqrt{n}}\right)$$

Note that you will have to use the inverse of Φ here. Now you can either just go and experimentally try it, but a better way is to use Φ^{-1} to get two equalities for C , and then solve for n . That is, from the first and second equations, respectively, we get:

$$C = (4\sqrt{n})\Phi^{-1}(1 - 0.05) - 4n$$

and

$$C = (4/3)\sqrt{n}\Phi^{-1}(0.05) - (4/3)n$$

Setting these equal we get $(4 + (4/3))\Phi^{-1}(0.95) = (4 - (4/3))\sqrt{n}$, which gives us $n = (2\Phi^{-1}(0.95))^2 = 10.82$, which means that we need 11 trials to be able to satisfy this (C turns out to be about -21.6).