STA 250/MTH 342 – Intro to Mathematical Statistics

Lecture 16

Beyond simple hypotheses

- Last class we see that an elegant result by Neyman and Pearson provides a general solution to the problem of testing *simple hypotheses*.
- It states that for a given level α for the Type I error rate, the LR test minimizes β, the Type II error rate, or equivalently maximizes π the power.
- In other words, the LR test is the *most powerful* test at the given level α.
- But what if one or both of the hypotheses are composite?
- It turns out that in *some* (not all!) of such problems, the LR test is also the *most powerful* test.
- What does "most powerful" mean here?

Example: Testing normal mean with known variance

- Suppose the data $X_1, X_2, \ldots, X_n$ are i.i.d. observations from $\mathrm{N}(\mu, \sigma_0^2)$.
- Consider the problem of testing

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0$$

  This is, more explicitly

$$H_0 : \theta = \theta_0 = (\mu_0, \sigma_0^2) \quad \text{vs} \quad H_1 : \theta = \theta_1 = (\mu_1, \sigma_0^2) \textit{ for any } \mu_1 > \mu_0.$$

- So we are interested in testing whether $\mu = \mu_0$, or some (unspecified) larger value.
- The alternative hypothesis here is composite. Which test should we use?

- Recall that for each *specific* value $\mu_1 > \mu_0$, we have found that the LR test takes the form

$$\text{Reject when } \bar{X} > C.$$

- *The form of the test does not depend on the value of μ_1!*
- When determining the form of the test we only used the fact that $\mu_1 > \mu_0$, not the specific value of μ_0.

$$\frac{L(\theta_1)}{L(\theta_0)} = \frac{(2\pi)^{-n/2}\sigma_0^{-n} e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i-\mu_0)^2}}{(2\pi)^{-n/2}\sigma_1^{-n} e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i-\mu_1)^2}}$$

$$= e^{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^n (X_i-\mu_1)^2 - \sum_{i=1}^n (X_i-\mu_0)^2\right]}$$

$$= e^{\frac{1}{\sigma_0^2}\left[(\mu_1-\mu_0)\sum_{i=1}^n X_i\right]} e^{-\frac{n}{2\sigma_0^2}(\mu_1^2-\mu_0^2)}$$

- The constant C is chosen so that the Type I error rate is α:

$$C = \mu_0 + \Phi^{-1}(1-\alpha)\frac{\sigma_0}{\sqrt{n}}$$

  which also does *not* depend on the value of μ_1.
- Therefore, the test that rejects when

$$\bar{X} > \mu_0 + \Phi^{-1}(1-\alpha)\frac{\sigma_0}{\sqrt{n}}$$

  is the most powerful level α test for *all* $\mu_1 > \mu_0$.
- The Type II error is given by

$$\beta(\mu_1) = \Phi\left(\frac{C-\mu_1}{\sigma_0/\sqrt{n}}\right) = \Phi\left(\frac{\mu_0-\mu_1}{\sigma_0/\sqrt{n}} + \Phi^{-1}(1-\alpha)\right),$$

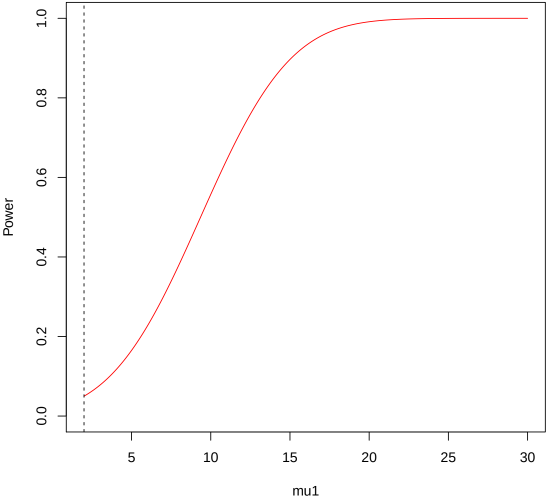
  which *does* depend on the alternative hypothesis μ_1.

- What is the power of the test?
- Now the alternative hypothesis is composite, the power is a function of the specific alternative parameter value.

$$\pi(\mu_1) = 1 - \beta(\mu_1) = 1 - \Phi\left(\frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + \Phi^{-1}(1-\alpha)\right).$$

- Draw a figure of the sampling distribution under H_0 and that under H_1. Illustrate how the power change as $\mu_1 - \mu_0$, σ_0, n, and α change.
- The *power function* $\pi(\mu)$ gives the power of the test when $\mu_1 = \mu$.
- Let us plot the *power function* of this test $\pi(\mu)$ versus μ.

Power function for $\mu_0 = 2$, $\sigma_0 = 20$, $n = 20$, and $\alpha = 0.05$.



- We see that the power increases as μ_1 is farther away from μ_0.
- What is the power at μ_0, $\pi(\mu_0)$?

- Note that the value of μ_1 affects the Type II error rate and the power of the test: but *not* the test itself, or the corresponding rejection region.
- Therefore the LR test is not only the most powerful test for a particular μ_1, but rather for *all* $\mu_1 > \mu_0$.
- We say the test is *uniformly most powerful* (UMP) against all alternatives $\mu_1 > \mu_0$.
- This would not have been true had the test (i.e. rejection region) depended on μ_1!

- Can we always find a UMP test for a composite alternative?
- The answer is no. In fact, only for a small class of problems does a UMP test exist.
- For the current example, if the alternative hypothesis is enlarged to a *two-sided* (as opposed to *one-sided*) alternative

$$H_1' : \theta = \theta_1 = (\mu_1, \sigma_0^2) \text{ for any } \mu_1 \neq \mu_0,$$

then no UMP tests exist.

- To see why, recall that for any $\mu_1 > \mu_0$ the most powerful test rejects when
$$\bar{X} > \mu_0 + \Phi^{-1}(1-\alpha)\frac{\sigma_0}{\sqrt{n}}.$$
- One can show (exercise!) by a similar reasoning that for any $\mu_1 < \mu_0$, the most powerful test rejects when
$$\bar{X} < \mu_0 - \Phi^{-1}(1-\alpha)\frac{\sigma_0}{\sqrt{n}}.$$
- Thus neither is the UMP test and so no UMP test exist.
- In fact, the test that is most powerful for $\mu_1 > \mu_0$ is actually the least powerful for $\mu_1 < \mu_0$ and vice versa. (Why? Gold miner analogy.)
- Intuitively, a compromise that combines the above two tests rejects when
$$|\bar{X} - \mu_0| > C' = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma_0}{\sqrt{n}}.$$
- Show that the level of this test is α.

The power functions of the three tests are (exercise!)

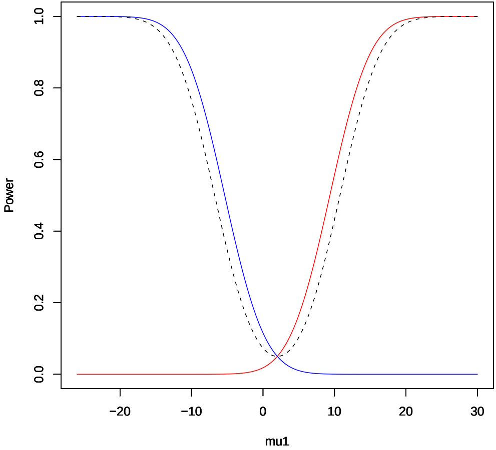
$$\pi_1(\mu_1) = 1 - \Phi\left(\frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + \Phi^{-1}(1-\alpha)\right),$$

$$\pi_2(\mu_1) = \Phi\left(\frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} - \Phi^{-1}(1-\alpha)\right)$$

and

$$\pi_3(\mu_1) = 1 - \Phi\left(\frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) + \Phi\left(\frac{\mu_0 - \mu_1}{\sigma_0/\sqrt{n}} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right)$$

Power functions of the three tests for $\mu_0 = 2$, $\sigma_0 = 20$, $n = 20$, and $\alpha = 0.05$.



- Red: $\pi_1(\mu_1)$; blue: $\pi_2(\mu_1)$; black dashed: $\pi_3(\mu_1)$.
- So the compromise is not most powerful for any particular alternative, but not too much worse.

Example: Testing a binomial proportion

- Our data is the number of successes X out of n independent Bernoulli trials with success probability θ.
- We are interested in testing

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0.$$

- For any particular θ_1, the likelihood ratio is

$$\frac{L(\theta_1)}{L(\theta_0)} = \frac{\theta_1^X (1-\theta_1)^{n-X}}{\theta_0^X (1-\theta_0)^{n-X}} = \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^X \left(\frac{1-\theta_1}{1-\theta_0}\right)^n.$$

- For $\theta_1 > \theta_0$, we have $\frac{\theta_1}{1-\theta_1} > \frac{\theta_0}{1-\theta_0}$, and so

$$\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} > 1.$$

- Therefore the LR test rejects when $X > C$.

- Again we use a pre-specified threshold for Type I error rate to choose the constant C.
- For example, we may want to find a level 0.05 test.
- If $n = 5$ and $\theta_0 = 0.5$, then the distribution fo X under H_0 is

x	0	1	2	3	4	5
$p(x\|\theta_0)$	.03	.16	.31	.31	.16	.03

- If we choose $C = 4$, then

$$P(X > C|\theta_0) = .03 \leq .05.$$

- If we choose $C = 3$, then

$$P(X > C|\theta_0) = .19 > .05.$$

- Therefore in order to have a level 0.05 test, we choose $C = 4$ and reject when $X > 4$.

- What is the power function of this test?
- It is
$$\pi(\theta_1) = P(X > 4|\theta_1) = P(X = 5|\theta_1) = \theta_1^5.$$
- This test is UMP for testing θ_0 vs $\theta_1 > \theta_0$. Why?
- Again, if we are testing θ_0 vs $\theta_1 \neq \theta_0$, is there a UMP test?