Statistics 210A Lecture 14 Notes

Daniel Raban

October 12, 2021

1 Introduction to Hypothesis Testing

1.1 Null and alternative hypotheses

Suppose we have a model $\mathcal{F} = \{P_\theta : \theta \in \Theta\}$ with data $X \sim P_\theta$, and we want to distinguish between two submodels, the **null hypothesis** $H_0 : \theta \in \Theta_0 \subseteq \Theta$, and the **alternative hypothesis** $H_1 : \theta \in \Theta_1$. If unspecified, $\Theta_1 = \Theta \setminus \Theta_0$.

There is an asymmetry here, where H_0 is considered the "default assumption." We either

1. reject H_0 (conclude $\theta \notin \Theta_0$)
2. fail to reject[1] H_0 (no definite conclusion).

Example 1.1. If $X \sim N(\theta, 1)$, here are common hypothesis tests:

- $H_0 : \theta = 0$ vs $H_1 : \theta > 0$.
- $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$.
- $H_0 : |\theta| \leq \delta$ vs not.

We can also consider nonparametric tests.

Example 1.2. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} P$ and $Y_1, \dots, T_m \overset{\text{iid}}{\sim}$. We can consider the hypothesis test

$$H_0 : P = Q, \qquad H_1 : P \neq Q.$$

[1] We might slip up and say "accept" the null, but really what we are doing is failing to reject the null. Don't say "accept" around non-statisticians.

1.2 The power function of a hypothesis test

How can we tell how good our hypothesis test is? We can formally describe a test by its critical function.

Definition 1.1. The **critical function** (or **test function**) of a hypothesis test is

$$\phi(x) = \begin{cases} 0 & \text{fail to reject } H_0 \\ \pi \in (0,1) & \text{reject with probability } \pi \\ 1 & \text{reject } H_0 \end{cases}$$

The power function tells us how good the test is.

Definition 1.2. The **power function** of a hypothesis test is

$$\beta_\phi(\theta) = \mathbb{E}_\theta[\phi(x)] = \mathbb{P}_\theta(\text{Reject } H_0).$$

Definition 1.3. For nonrandomized ϕ, the **rejection region** is

$$R = \{x : \phi(x) = 1\},$$

and the **acceptance region** is

$$A = \mathcal{X} \setminus R.$$

So the power function is $\mathbb{P}_\theta(X \in R)$. We want the power to be large on the alternative hypothesis and small on the null hypothesis. Usually, people refer to the power under the alternative hypothesis, so you want more power for your test.

Definition 1.4. The **significance level** of ϕ is

$$\sup_{\theta \in \Theta_0} \beta_\phi(\theta).$$

We'll say ϕ is a **level-α test** if its significance level is $\leq \alpha$.

The ubiquitous choice is $\alpha = 0.05$.[2]

Example 1.3. Let $X \sim N(\theta, 1)$, where we are testing $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$. Let $z_n = \Phi^{-1}(1-\alpha)$, where Φ denotes the normal CDF. The usual 2-sided test is

$$\phi_2(X) = \mathbb{1}_{\{|X| > z_{\alpha/2}\}}.$$

We could also do a 1-sided test

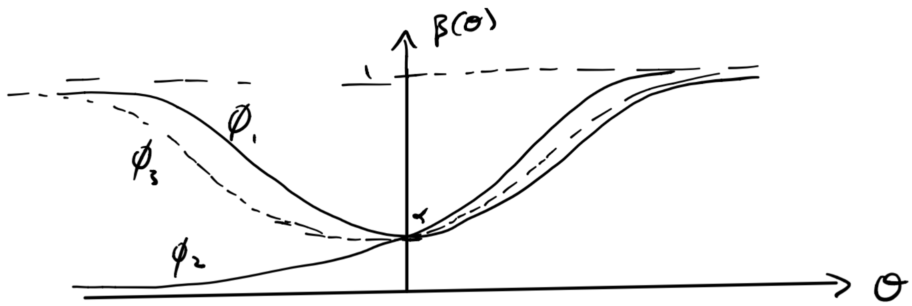
$$\phi_1(X) = \mathbb{1}_{\{X > z_\alpha\}}.$$

[2] This is probably ubiquitous because when Fisher came up with the idea of hypothesis testing, he said that he sometimes likes to use the value 0.05. This is probably this most influential offhand remark in the history of science.

Both of these are valid hypothesis tests at level α; the 1-sided test has lower power for $\theta < 0$. We could also try any number of hypothesis tests, such as

$$\phi_3(X) = \mathbb{1}_{\{x < -z_{\alpha/3} \text{ or } X > z_{2\alpha/3}\}}.$$

We can plot the power of these tests against θ:



Can we tell which hypothesis test is the best? In some situations, there is a best test.

Example 1.4. Let $X \sim (0, 1)$ with $H_0 : \theta \leq 0$ vs $H_1 : \theta > 0$. Then the test ϕ_1 is the best possible test (called uniformly most powerful). We will discuss this in detail next time.

So 1-sided tests have a best test. We'll start simple and work our way up to more complicated tests.

Definition 1.5. A **simple hypothesis** is a singleton. A **composite hypothesis** is one that isn't simple.

1.3 Likelihood ratio tests and the Neyman-Pearson lemma

Suppose we test $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$. Without loss og generality, we may assume $\theta_0 = 0$ and $\theta_1 = 1$. Without loss of generality, assume P_0 and P_1 have densities p_0, p_0 (which we may do because P_0 and P_1 are both absolutely continuous with respect to $P_0 + P_1$). The optimal test rejects for large values of $\frac{p_1(x)}{p_0(x)}$.

Definition 1.6. The **likelihood ratio test (LRT)** is of the form

$$\phi^*(x) = \begin{cases} 1 & \frac{p(x)}{p_0(x)} > c \\ \gamma & \frac{p_1(x)}{p_0(x)} = c \\ 0 & \frac{p_1(x)}{p_0(x)} < c, \end{cases}$$

where c, γ are chosen so $\mathbb{P}_0(\text{Reject}) = \alpha$.

We will prove that this is the best test, but first, here is some intuition. The power under the alternative hypothesis H_1 is

$$\int_{\mathbb{R}} p_1(x)\, d\mu(x),$$

and the significance level is

$$\int_{\mathbb{R}} p_0(x)\, d\mu(x).$$

We want to maximize the first integral subject to constraint that the second integral equals α. Think of the first integral as the bang, and the second integral as the buck; you want to get the most bang for your buck. If you think about wanting to buy flour from the grocery store with a fixed budget, you'll try to buy bags of flour with the lowest cost per unit until you run out of money. Here, the cost per unit is $\frac{p_1(x)}{p_0(x)}$, and the γ corresponds to the little bit of change you have left over, which you use to buy a fractional bag of flour.

To carry out the proof that the likelihood ratio test is the best test, we would like to use Lagrange multipliers. Since this is over infinitely many parameters, here is a lemma which lets us carry out this optimization.

Proposition 1.1 (12.1 in Keener)**.** *Suppose $c \geq 0$ and ϕ^* maximizes*

$$\mathbb{E}_1[\phi(X)] - c\,\mathbb{E}_0[\phi(X)]$$

among all critical functions. If $\mathbb{E}_0[\phi(X)] = \alpha$, then ϕ^* maximizes $\mathbb{E}_1[\phi(X)]$ among all level-α tests ϕ.

Proof. Suppose $\mathbb{E}_0[\phi(X)] \leq \alpha$. Then

$$\begin{aligned}
\mathbb{E}_1[\phi(X)] &\leq \mathbb{E}_1[\phi(X)] + c(\alpha - \mathbb{E}_0[\phi(X)]) \\
&\leq \mathbb{E}_1[\phi^*(X)] - c\,\mathbb{E}_0[\phi^*(X)] + c\alpha \\
&= \mathbb{E}_1[\phi^*(X)].
\end{aligned}$$

□

Theorem 1.1 (Neyman-Pearson[3])**.** *The likelihood ratio test wit h signigicance level $= \alpha$ is optimal for testing $H_0 : X \sim P_0$ vs $H_1 : X \sim P_1$ (maximizes $E_1[\phi(X)]$ such that $\mathbb{E}_0[\Phi(X)] \leq \alpha$).*

Proof. We want to maximize the Lagrangian

$$\begin{aligned}
\mathcal{L}(\phi; c) &:= \mathbb{E}_1[\phi(X)] - c\,\mathbb{E}_0[\phi(X)] \\
&= \int_{\mathcal{X}} (p_1(x) - cp_0(x))\phi(x)\, d\mu(x)
\end{aligned}$$

[3]This important theorem is often referred to as a lemma.

$$= \int_{\{p_1 > cp_0\}} |p_1 - cp_0| \phi \, d\mu - \int_{p_1 < cp_0} |p_1 - cp_0| \phi \, d\mu.$$

To maximize $\mathcal{L}(\phi; c)$, set

$$\phi(x) = \begin{cases} 1 & \text{if } \frac{p_1(x)}{p_0(x)} > c \\ 0 & \text{if } \frac{p_1(x)}{p_0(x)} < c. \end{cases}$$

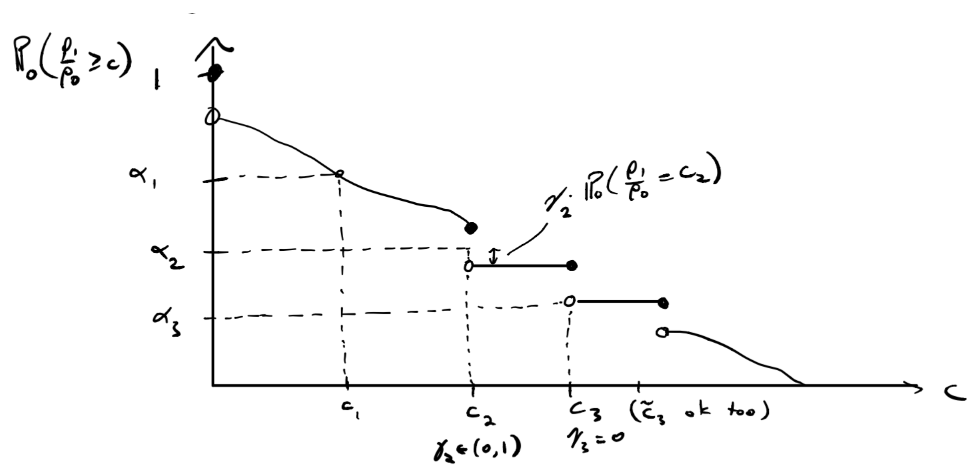
Choose the minimum value of c such that

$$\mathbb{P}_0\left(\frac{p_1}{p_0}(X) > c\right) \leq \alpha \leq \mathbb{P}_0\left(\frac{p_1}{p_0}(X) \geq x\right),$$

and choose γ to "top up" the significance level to α:

$$\mathbb{P}_0\left(\frac{p_1}{p_0}(X) > c\right) + \gamma \mathbb{P}_0\left(\frac{p_1}{p_0}(X) = c\right) = \alpha. \qquad \square$$

Here's a picture of how we can pick c_α and γ_α for ϕ^*:



Corollary 1.1 (12.4 in Keener)**.** *If $p_0 \overset{a.s.}{\neq} p_1$ and ϕ is the LRT with level $\alpha \in (0,1)$, then $\mathbb{E}_1[\phi(X)] > \alpha$.*

Proof. We have $\mu(\{p_1 > p_0\}), \mu(\{p_0 > p_1\}) > 0$. We split into a few cases:

$c \geq 1$: We split

$$\mathbb{E}_1[\phi] - \mathbb{E}_0[\phi] = \int_{\{p_1/p_0 > 1\}} |p_1 - p_0| \phi \, d\mu - \int_{\{p_1/p_0 < 1\}} |p_1 - p_0| \phi \, d\mu$$

$$> 0.$$

$c < 1$: This case is similar. $\square$

Example 1.5. Suppose we have a 1-parameter exponential famil $X \sim p_\eta(x) = e^{\eta T(x) - A(\eta)}$. Test the null hypothesis $H_0 : \eta = \eta_0$ vs the alternative $H_1 : \eta = \eta_1 > \eta_0$. The likelihood ratio is

$$\begin{aligned} \frac{p_1(x)}{p_0(x)} &= \frac{e^{\eta_1 T(x) - A(\eta)}}{e^{\eta_0 T(x) - A(\eta)}} \\ &= e^{(\eta_1 - \eta_0) T(x) - (A(\eta_1) - A(\eta_0))} \end{aligned}$$

So the LRT should be to reject when this is large. Since this is a monotone function in $T(x)$, this is the same as saying we reject when $T(x)$ is large. So we can say the test is

$$\phi^*(x) = \begin{cases} 1 & T(x) > c \\ \gamma & T(x) = c \\ 0 & T(x) < c, \end{cases}$$

where we choose c, γ to make

$$\mathbb{P}_{\eta_0}(T(X) > c) + \gamma \mathbb{P}_{\gamma_0}(T(X) = c) = \alpha.$$

Notice that η_1 is nowhere to be found. So this exact test is the best against any alternative η_1, as long as $\eta_1 > \eta_0$. So the best test only depends on the direction of the alternative.

Next time, we will discuss more situations like this, where we have best tests against any alternative in a range of alternatives.