Statistics 210A Lecture 15 Notes

Daniel Raban

October 14, 2021

1 One-Sided and Two-Sided Tests

1.1 Recap: Basics of hypothesis testing

Last time, we introduced hypothesis testing, where we have a model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ and want to distinguish between $H_0 : \theta \in \Theta$ and $H_1 : \theta \in \Theta$ (usually, $\Theta_1 = \Theta \setminus \Theta_0$). The tests were described by a **critical function** $\phi : \mathcal{X} \to [0, 1]$ given by

$$\phi(x) = \begin{cases} 1 & \text{reject} \\ \pi & \text{flip a (biased) coin} \\ 0 & \text{fail to accept.} \end{cases}$$

We defined the **rejection region** $R = \{x : \phi(x) = 1\}$ (ignoring randomization), the **power function** $\beta_\phi(\theta) = \mathbb{E}_\theta[\phi(X)] = \mathbb{P}_\theta(\text{Reject } H_0)$, and the **significance level** $\sup_{\theta \in \Theta_0} \beta_\phi(\theta)$.

Our goal is to obtain the maximum power for $\theta \in \Theta_1$, relative to the constraint that the significance level is at α. There are two types of errors in this setting:

Definition 1.1. A **Type I error** is rejecting the null hypothesis when H_0 is true. A **Type II error** is failing to reject the null hypothesis when H_1 is true.

We introduced the **Likelihood Ratio Test (LRT)** in the case of a simple null $H_0 : \theta = \theta_0$ vs a simple alternative hypothesis $H_1 : \theta = \theta_1$. This test is given by

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

where we choose c, γ such that $\mathbb{E}_0[\phi(X)] = \alpha$. There is a bit of ambiguity because any test of the form (for $c \geq 0$)

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

maximizes $\mathbb{E}_1[\phi(X)] - c\,\mathbb{E}_0[\phi(X)] = \int (p_1 - cp_0)\,d\mu$, as long as we keep the constraint that the significance level is α.

Last time, we had a proposition that said that any test of this form maximizes $\mathbb{E}_1[\phi(X)]$ subject to $\mathbb{E}_\theta[\phi(X)] = \alpha =: \mathbb{E}_1[\phi^*]$. A corollary to this

Example 1.1. If $X \sim p_\eta(x) = e^{\eta T(x) - A(\eta)} h(x)$ is an exponential family with $H_0 : \eta = \eta_0$ and $H_0 : \eta = \eta_1 > \eta_0$, then the LRT gave

$$LR(X) = e^{(\eta_1 - \eta_0)T(X) - (A(\eta_1) - A(\eta_0))},$$

which was monotone in $T(X)$. So we saw that the LRT was dependent only on $T(X)$ and not on the particular value of η_1. So the same exact test is the best for all alternative hypotheses of this form.

1.2 Uniformly most powerful (UMP) tests

Definition 1.2. If $\phi^*(X)$ has significance level α, and for any other level-α test ϕ,

$$\mathbb{E}_\theta[\phi^*(X)] \geq \mathbb{E}_\theta[\phi(X)] \qquad \forall \theta \in \Theta_1,$$

we say that ϕ^* is **uniformly most powerful (UMP)**.

Definition 1.3. A model $\mathcal{P}$ is **identifiable** if $\theta_1 \neq \theta_2 \implies P_{\theta_1} \neq P_{\theta_2}$.

This is just saying that the different values of θ actually mean different things in our model.

Definition 1.4. Assume $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ is identifiable and has densities p_θ for P_θ with respect to μ. We say $\mathcal{P}$ has **monotone likelihood ratios (MLR)** in $T(x)$ if $\frac{p_{\theta_2}}{p_{\theta_1}}(x)$ is a nondecreasing function of $T(x)$ for all $\theta_2 > \theta_1$.

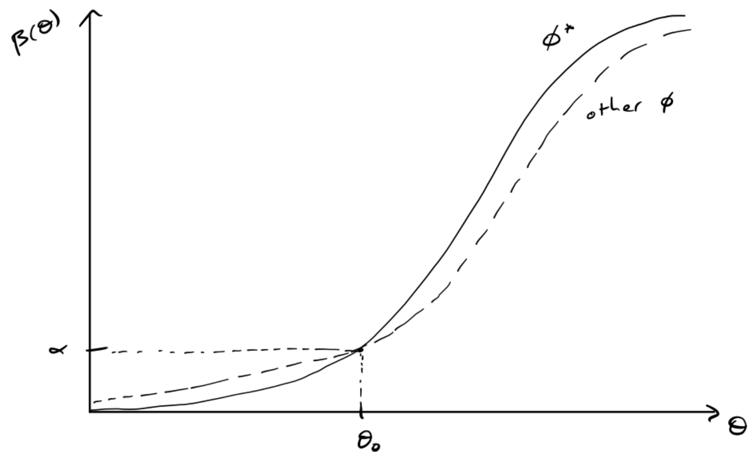
Remark 1.1. This is different from $T(X)$ being **stochastically increasing** in θ, which says that $\mathbb{P}_\theta(T(X) > c)$ is increasing in θ. This condition is enough to construct a valid one-sided test that rejects when T is large, but it will not necessarily be uniformly most powerful.

Theorem 1.1. *Assume $\mathcal{P}$ has MLR in $T(x)$, and test $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$. Let let $\phi^*(x)$ reject for large $T(x)$, where c, γ are chosen so $\mathbb{E}_{\theta_0}[\phi^*(X)] = \alpha$.*

(a) ϕ^* is a UMP level-α test.

(b) $\beta_{\phi^*}(\theta) = \mathbb{E}_\theta[\phi^*(X)]$ is non-decreasing in θ and strictly increasing if $\mathbb{E}_\theta[\phi^*(X)] \in (0, 1)$.

(c) If $\theta_1 < \theta_0$, ϕ^* minimizes $\mathbb{E}_{\theta_1}[\phi(X)]$ among all tests ϕ with power $= \alpha$ at θ.



Proof.

(b): if $\theta_1 < \theta_2$, then $\frac{p_{\theta_2}}{p_{\theta_1}}(x)$ is nondecreasing i n$T(X)$. So ϕ^* is a LRT for $H_0 : \theta = \theta_1$ vs $H_1 : \theta = \theta_2$ (at level $\widetilde{\alpha} := \mathbb{E}_{\theta_1}[\phi^*(X)]$). Then the corollary from last time says that $\mathbb{E}_{\theta_2}[\phi^*(X)] > \widetilde{\alpha} = \mathbb{E}_{\theta_1}[\phi^*(X)]$.

(a): If $\theta > \theta_0$, then ϕ^* is the LRT for $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.

(c): If $\theta_1 < \theta_0$, assume $\mathbb{E}_{\theta_0}[\widetilde{\phi}(X)] = \alpha$. Then both $1 - \phi^*$ and $1 - \widetilde{\phi}$ are level $1 - \alpha$ tests of $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$. But $1 - \phi^*$ is the LRT for this test. Indeed, $\frac{p_{\theta_1}}{p_{\theta_0}}(x) = [\frac{p_{\theta_0}}{p_{\theta_1}}(x)]^{-1}$ is decreasing in T, and $1 - \phi^*$ rejects for small $T(X)$. So ϕ^* maximizes $\mathbb{E}_{\theta_1}[1 - \phi]$ scuh that $\mathbb{E}_{\theta_0}[1 - \phi] \leq 1 - \alpha$. □

1.3 Two-sided tests

What about two-sided alternative hypotheses? Suppose $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}\}$ with $\theta_0 \in \Theta^0$, where want to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ (this can be generalized to $H_0 : \theta \in [\theta_1, \theta_2]$).

Definition 1.5. $T(X)$ is **stochastically increasing** in θ if $\mathbb{P}_\theta(T(X) \leq t)$ is nonincreasing in θ for all t.

Assume $T(X)$ is a stochastically increasing summary test statistic.

Example 1.2. For example, this applies to $X_i \overset{\text{iid}}{\sim} p_0(x - \theta)$ where $T(X)$ is the sample mean or median.

Example 1.3. This also applies to $X_i \overset{\text{iid}}{\sim} \frac{1}{\theta} p_1(x/\theta)$ where $T(X) = \sum_i X_i^2$.

Definition 1.6. The **two-tailed test** rejects when $T(X)$ is extreme in any direction:

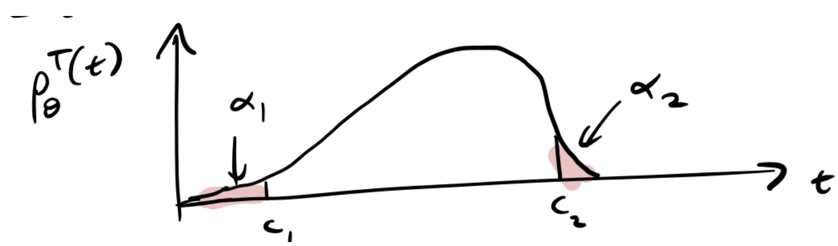
$$\phi(x) = \begin{cases} 1 & T(x) < c_1 \text{ or } T(x) > c_2 \\ 0 & T(X) \in (c_1, c_2) \\ \gamma_i & T(x) = c_i, i = 1, 2. \end{cases}$$

In this setting, we will not usually be able to get a UMP test. We usually have a tradeoff between allocating our type I error to values where θ is large or values where θ is small. Let

$$\alpha = \mathbb{P}_{\theta_0}(T(X) < c_1) + \gamma_1 \mathbb{P}(T(X) = c_1)$$
$$\alpha_2 = \mathbb{P}_{\theta_0}(T(X) > c_2) + \gamma_2 \mathbb{P}(T(X) = c_2).$$

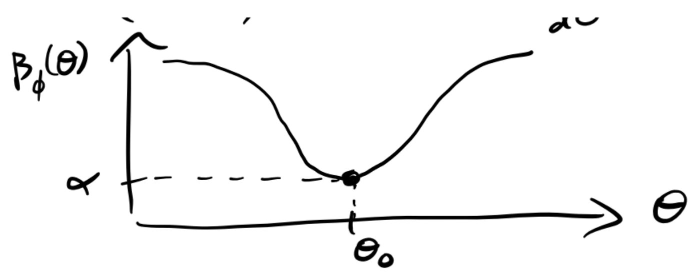
We need $\alpha_1 + \alpha_2 = \alpha$, and we have to balance these considerations. Here are some ideas:

One natural way to do this is to do an **equal-tailed test**, i.e. set $\alpha_1 = \alpha_2 = \alpha/2$.



Definition 1.7. $\phi(x)$ is **unbiased** if

$$\inf_{\theta \in \Theta_1} \mathbb{E}_\theta[\phi(X)] \geq \alpha.$$

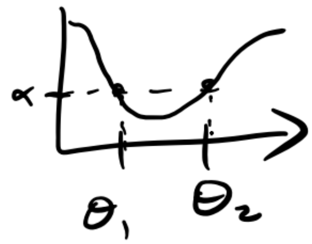


The second idea is to choose an unbiased test.

Theorem 1.2. *Assume $X_i \overset{\text{iid}}{\sim} e^{\theta T(x) - A(\theta)} h(x)$, so the sufficient statistic $\sum_{i=1}^n T(X_i)$. Test $H_0 : \theta \in [\theta_1, \theta_2]$ (with possibly $\theta_1 = \theta_2$) vs the alternative $H_1 : \theta \notin [\theta_1, \theta_2]$. Let $\phi(x)$ be the two-tailed test based on $\sum_{i=1}^n T(X_i)$.*

(a) *The unbiased two-tailed test for $\sum_{i=1}^n T(X_i)$ with significance level $= \alpha$ is UMP among all unbiased tests (UMPU).*

(b) If $\theta_1 < \theta_2$, the UMPU test solves $\mathbb{E}_{\theta_1}[\phi(X)] = \mathbb{E}_{\theta_2}[\phi(X)] = \alpha$.



(c) If $\theta_1 = \theta_2 = \theta_0$, the UMPU test solves $\mathbb{E}_{\theta_0}[\phi(X)] = \alpha$ and

$$\mathbb{E}_{\theta_0}\left[\sum_{i=1}^{n} T(X_i)(\phi(X) - \alpha)\right] = \frac{d}{d\theta}\,\mathbb{E}_\theta[\phi(X)]\Big|_{\theta=\theta_0} = 0.$$

Proof. Proof is in Keener. □

1.4 p-values

Here is an informal definition (if $\phi(x)$ rejects for large $T(x)$): The p-value is

$$\begin{aligned} p(x) &= \text{“}\mathbb{P}_{H_0}(T(X) \geq T(x))\text{.”} \\ &= \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T(X) \geq T(x)). \end{aligned}$$

Example 1.4. Let $X \sim N(\theta, 1)$, and test $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$. The two-sided test rejects for large $|X|$, and the two-sided p-value is

$$p(x) = \mathbb{P}_\theta(|x| > |x|) = 2(1 - \Phi(|x|)).$$

We could instead test $H_0 : |\theta| \leq \delta$ against $H_1 : |\theta| > \delta$. It turns out that we will get

$$\begin{aligned} p(x) &= \mathbb{P}_\delta(|X| > |x|) \\ &= 1 - \Phi(|x| - \delta) + \Phi(-|x| - \delta). \end{aligned}$$

Not every test will look like this, so we want a more formal definition.

Definition 1.8. Let $\phi_\alpha(x)$ be a family of tests with $\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\phi_\alpha(x)] \leq \alpha$ and $\phi_\alpha(x)$ monotone in α. Then the p**-value** is

$$\begin{aligned} p(x) &= \inf\{\alpha : \phi_\alpha(x) = 1\} \\ &= \inf\{\alpha : x \in R_\alpha\}. \end{aligned}$$

This is the α for which the corresponding test just barely rejects.

For $\theta \in \Theta_0$,

$$\mathbb{P}_\theta(p(X) \le \alpha) \le \inf_{\widetilde{\alpha} > \alpha} \mathbb{P}_\theta(\phi_{\widetilde{\alpha}}(X) = 1) \le \alpha,$$

so the p-value is stochastically larger than $U[0,1]$ under H_0.

Remark 1.2. The p-value is dependent on not just the data but also the null hypothesis and the hypothesis test we use! This is something many people misunderstand in practice.