

STATS300A - Lecture 12

Dominik Rothenhaeusler
Scribed by Michael Howes

11/01/21

Contents

1 Setting from now on	1
2 Set up	1
3 Level α tests	3
4 Optimal simple tests	3

1 Setting from now on

From now on we will be doing hypothesis testing (binary decision problems). As with estimation we will have to add constraints. The two constraints we will consider are risk bounds and unbiasedness. There is a version of equivariance used in hypothesis testing but we will not study this.

2 Set up

As before we have data $X \sim \mathbb{P}_\theta$ and a model $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Omega\}$. We now also have a partition $\Omega = \Omega_0 \cup \Omega_1$ where $\Omega_0 \subseteq \Omega$ and $\Omega_1 = \Omega \setminus \Omega_0$. From data X we wish to determine if $\theta \in \Omega_0$ or $\theta \in \Omega_1$. We call these statements hypotheses. We have

$$\begin{aligned} H_0 &= \text{null hypothesis: } \theta \in \Omega_0, \\ H_1 &= \text{alternative hypothesis: } \theta \in \Omega_1. \end{aligned}$$

Our decision space D is $\{\text{reject } H_0, \text{accept } H_0\}$. We use 0 – 1 loss which we can visualize as a grid.

	$\theta \in \Omega_0$	$\theta \in \Omega_1$
reject H_0	1 (Type I error)	0
accept H_0	0	1 (Type II error)

Remark 1. In this case the loss is symmetric and a Type I error is equally as bad as a Type II error. There may be times when we wish to have an asymmetrical loss. For example if we are banning a substance that may cause cancer. In this case we might worry less about Type I error. This is because failing to ban a harmful substance is much worse than banning a harmless substance.

In the context of hypotheses testing we will call decision procedures tests. We will allow for randomized tests $\delta(X, U)$ but we will use a new representation for randomized procedures.

Definition 1. Given a randomized test $\delta(X, U)$, the *test function* of δ is a function $\phi : \mathcal{X} \rightarrow [0, 1]$ given by

$$\phi(x) = \mathbb{P}(\delta(x, U) = 1).$$

That is $\phi(x)$ is the probability that we reject H_0 after seeing data x .

Note that $\phi(x)$ does not depend on the unknown parameter θ . This is because $U \perp\!\!\!\perp X$ and we have fixed x . Given a test function ϕ we can construct a randomized test δ_ϕ that has test function ϕ . We define δ_ϕ in the following way:

- If $\phi(x) = 1$, then $\delta_\phi(x, U) = 1$ with probability 1.
- If $\phi(x) = 0$, then $\delta_\phi(x, U) = 0$ with probability 1.
- If $\phi(x) \in (0, 1)$, then $\delta_\phi(x, U) = 1$ with probability $\phi(x)$ and $\delta_\phi(x, U) = 0$ otherwise.

The advantage of test functions is that there may be multiple randomized test with the same test function. These different tests are the same from a risk/decision theoretic perspective. The test function contains all the information about the risk of a test. Thus we will work with test functions rather than working with randomized tests directly.

Definition 2. The *power function* of a test function ϕ is a function $\beta : \Omega \rightarrow [0, 1]$ that gives the probability that ϕ rejects under the distribution \mathbb{P}_θ . That is,

$$\beta(\theta) = \mathbb{E}_\theta[\phi(X)] = \mathbb{P}_\theta(\delta_\phi \text{ rejects } H_0).$$

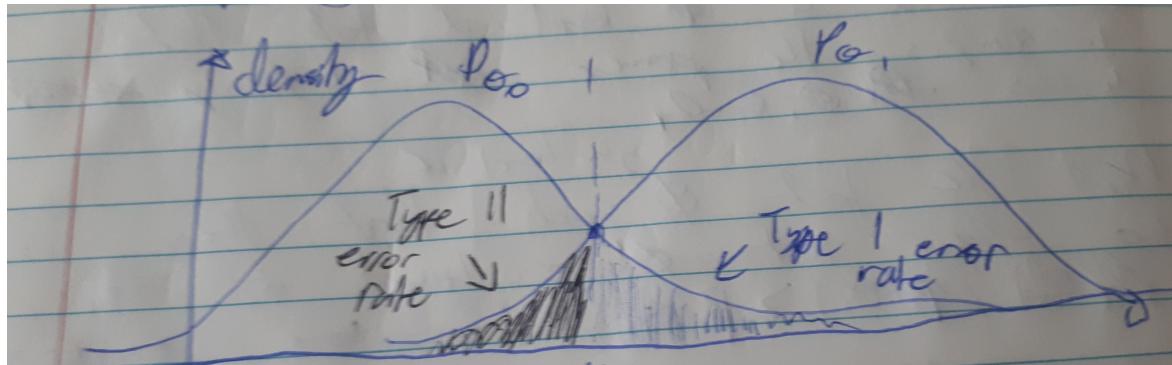
Note that if $\theta \in \Omega_0$, then

$$\beta(\theta) = \mathbb{P}_\theta(\delta_\phi \text{ rejects } H_0) =: \text{Type I error rate.}$$

Likewise if $\theta \in \Omega_1$, then

$$1 - \beta(\theta) = 1 - \mathbb{P}_\theta(\delta_\phi \text{ rejects } H_0) = \mathbb{P}_\theta(\delta_\phi \text{ accepts } H_0) =: \text{Type II error rate.}$$

Thus an optimal test would minimize $\beta(\theta_0)$ uniformly for $\theta_0 \in \Omega_0$ and maximize $\beta(\theta_1)$ uniformly for $\theta_1 \in \Omega_1$. Thus there is going to be a fundamental trade off between minimizing/maximizing β in these two different regions. Consider the below picture where we have plotted the densities p_{θ_0} and p_{θ_1} for fixed $\theta_0 \in \Omega_0$ and $\theta_1 \in \Omega_1$. Since the two densities overlap, the two error rates (corresponding to the shaded areas) trade off against each other.



3 Level α tests

How do we get around this tradeoff? Like with estimation, we will constrain the set of decision procedures.

Definition 3. The *size* of test function ϕ is

$$\sup_{\theta_0 \in \Omega_0} \mathbb{E}_{\theta_0} \phi(X) \in [0, 1].$$

Thus the size is the largest possible Type I error rate. For $\alpha \in [0, 1]$, a test ϕ is said to be *level α* if

$$\sup_{\theta_0 \in \Omega_0} \mathbb{E}_{\theta_0} \phi(X) \leq \alpha.$$

We will now constrain our class of tests to the level α tests and ask which of these tests are optimal.

Definition 4. A test, ϕ , is *uniformly most powerful (UMP)* at level α if ϕ is a level α -test and among the class of level α -tests, ϕ maximizes the power $\beta(\theta_1)$ uniformly across all $\theta_1 \in \Omega$.

That is, a UMP test is a level α test ϕ such that for all level α tests ϕ' and parameters $\theta_1 \in \Omega_1$, we have

$$\beta'(\theta_1) \leq \beta(\theta_1),$$

where β' and β are the power functions of ϕ and ϕ' respectively.

4 Optimal simple tests

Definition 5. The hypothesis H_0 is called *simple* if $|\Omega_0| = 1$ and otherwise H_0 is called *composite*. Likewise, H_1 is called *simple* if $|\Omega_1| = 1$ and otherwise H_1 is called *composite*.

For now suppose that both H_0 and H_1 are simple and so $|\Omega_0| = |\Omega_1| = 1$. Consider the measure $\mu = \mathbb{P}_0 + \mathbb{P}_1$ where $H_0 : X \sim \mathbb{P}_0$ and $H_1 : X \sim \mathbb{P}_1$. The measures \mathbb{P}_0 and \mathbb{P}_1 are both uniformly continuous with respect to μ and thus, by Radon-Nikodym they have densities p_0 and p_1 with respect to μ . Thus without loss of generality we can assume that \mathbb{P}_0 and \mathbb{P}_1 have densities with respect to a common measure.

Our goal thus becomes the constrained maximization problem

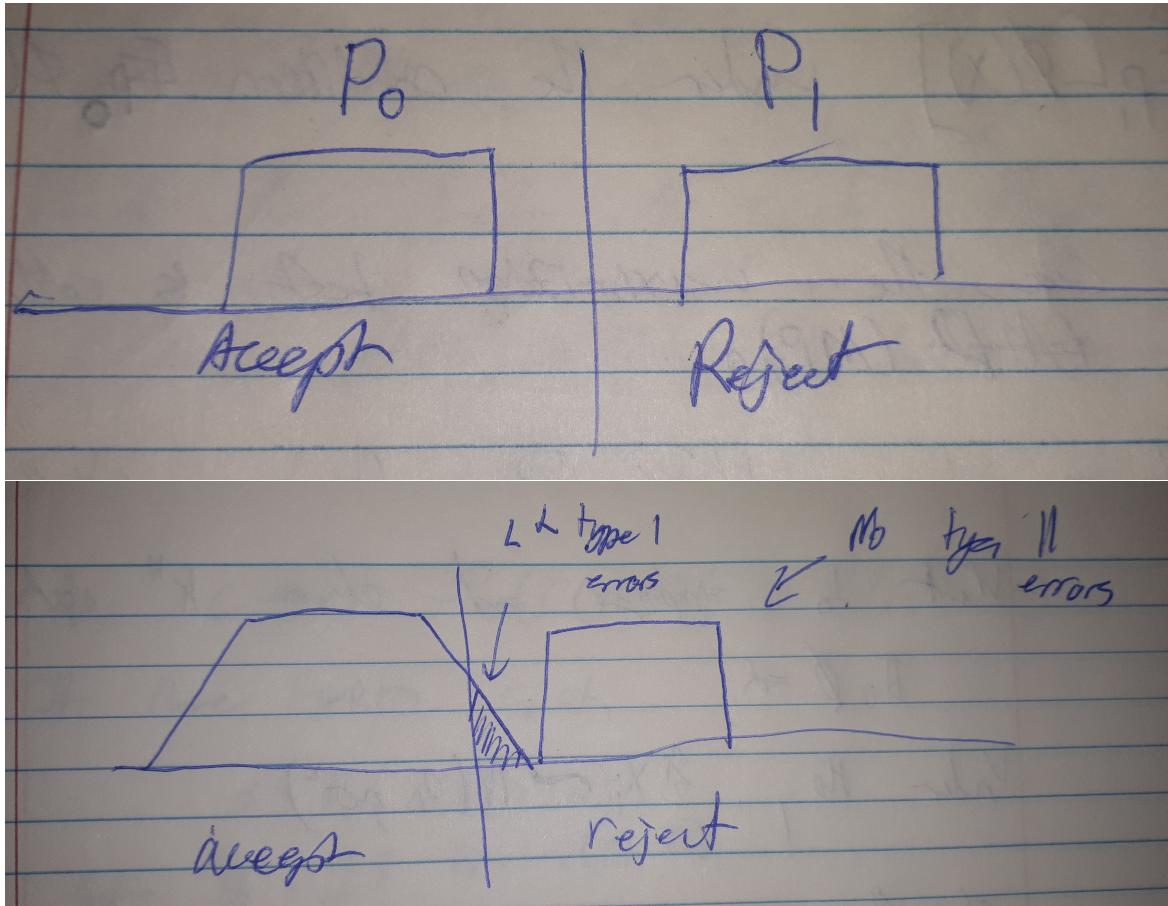
$$\max_{\phi} \mathbb{E}_{p_1} [\phi(X)], \text{ s.t } \mathbb{E}_{p_0} [\phi(X)] \leq \alpha.$$

In the case of simple testing, the maximizing test is called *most powerful (MP)*.

Theorem 1. [Neyman-Pearson Lemma (NPL)] Let \mathbb{P}_0 and \mathbb{P}_1 be probability distributions with densities p_0 and p_1 w.r.t to μ and consider testing $H_0 : X \sim \mathbb{P}_0$ against $H_1 : X \sim \mathbb{P}_1$. Then,

1. Existence there exists a test of H_0 against H_1 and a constant k such that
 - (a) $\mathbb{E}_{\mathbb{P}_0} \phi(X) = \alpha$.
 - (b) $\phi(x) = \begin{cases} 1 & \text{if } \frac{p_1(x)}{p_0(x)} > k, \\ 0 & \text{if } \frac{p_1(x)}{p_0(x)} \leq k. \end{cases}$
2. Sufficiency If a test ϕ satisfies (a) and (b), then ϕ is MP for testing H_0 against H_1 .
3. Necessity If ϕ is MP at level α for H_0 vs H_1 , then ϕ satisfies (b) for some $k \in [0, \infty]$. The test ϕ also satisfies (a) unless there exists a test with size strictly smaller than α and power equal to 1 for all $\theta_1 \in \Omega_1$.

Note that in 1.(b) we do not specify the value of $\phi(x)$ when $\frac{p_1(x)}{p_2(x)} = k$. We will discuss this case during the proof. The condition in part 3. where (a) does not hold is rather rare. It corresponds to when we have a rejection region which contains the entire support of p_1 and p_0 has probability less than α of lying in said region. Thus our densities look like one of the below pictures



In both of these cases the problem of hypothesis testing at level α has become somewhat trivial.

Example 1. Here we discuss a continuous example. For a discrete example - see the scribed notes. Suppose our data is $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known. Consider the hypotheses $H_0 : \mu = 0$ and $H_1 : \mu = \mu_1$ where $\mu_1 > 0$. By NPL we can find the most powerful test by looking at the likelihood ratio. Note that

$$\begin{aligned} r(x) &= \frac{p_1(x)}{p_0(x)} \\ &= \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu_1)^2\right\}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}x_i^2\right\}}. \end{aligned}$$

Thus $r(x) > k$ if and only if $\log(r(x)) > k'$ if and only if $\sum_{i=1}^n x_i > k''$. We wish to choose k'' such that

$$\mathbb{P}_0 \left(\sum_{i=1}^n X_i > k'' \right) = \mathbb{E}_0 \phi = \alpha.$$

Under H_0 we know that $\sum_{i=1}^n X_i \sim \mathcal{N}(0, n\sigma^2)$. Thus the test $\phi = \mathbb{I}(\sum_{i=1}^n X_i > k'')$ is equivalent to $\phi = \mathbb{I}\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n X_i > k'''\right)$. The latter test has level $1 - \Phi(k''')$ where Φ is the CDF of a $\mathcal{N}(0, 1)$ random variable. Thus we choose $k''' = \Phi^{-1}(1 - \alpha)$. Such a test satisfies (b) and (a) and thus is the MP test for the two simple hypothesis H_0 and H_1 .

However, our test ϕ does not depend on μ_1 and thus ϕ is the uniformly most powerful test of $H_0 : \mu = 0$ against $H_1 : \mu > 0$. (Note that the condition $\mu_1 > 0$ was important in the above calculations). This is a common strategy for find UMP tests. First restrict to simple test and then show that the MP test does not depend on the choice of parameter used for the simple test.

In this example, how we go from k' to k'' and the power of the test depend on μ_1 but the final test we constructed does not depend on μ_1 .

Proof of Neyman Pearson. We will start the proof today and finish it next lecture. We will first prove existance (part 1.). Let $r(x) = \frac{p_1(x)}{p_0(x)}$ and define a function $\alpha(c) = \mathbb{P}_0(r(X) > c)$. If there exists c_α such that $\alpha = \mathbb{P}_0(r(X) > c_\alpha)$, then define ϕ by

$$\phi(x) = \begin{cases} 1 & \text{if } r(x) > c_\alpha, \\ 0 & \text{else.} \end{cases}$$

Then ϕ satisfies (b) and

$$\mathbb{P}_0(\text{reject } H_0) = \mathbb{P}_0(r(X) > c_\alpha) = \alpha.$$

Thus ϕ also satisfies (a). We will continue this proof next lecture. □