Section 10, 11/5/19

Bayesian Testing

Suppose we have observed data and hypothesize two potential distributions from which the data could have come. Often the distribution is the same and we hypothesize two potential sets of parameters H_0 and H_1. In this case, our posterior probabilities are

$$P(H_0|x) \quad \text{and} \quad P(H_1|x)$$

By Bayes law

$$P(H_0|x = \frac{P(H_0,x)}{P(x)}) = \frac{P(x|H_0P(H_0))}{P(x)}$$

Thus, we can express the ratio of the posterior probabilities as

$$\frac{P(H_0|x)}{P(H_1|x)} = \frac{P(H_0)}{P(H_1)}\frac{P(x|H_0)}{P(x|H_1)}$$

the product of the priors and likelihoods. Once we define our priors, this ratio is a defined number. In order to choose a hypothesis, it makes sense to choose H_0 if the ratio is greater than 1, i.e. the likelihood of H_0 is greater than that of H_1.

$$\frac{P(H_0|x)}{P(H_1|x)} > 1$$

Since the ratio of the priors is some constant, this is equivalent to

$$\frac{P(x|H_0)}{P(x|H_1)} > c$$

for some threshold value c based on our prior beliefs. That is $P(x|H_0) > cP(x|H_1)$. It turns out that the choice of c balances the tradeoff between Type 1 and Type 2 errors.

Type 1 Error: P(reject $H_0|H_0$)

Type 2 Error: P(accept $H_0|H_1$)

Neyman Pearson

This theory of hypothesis testing frames it as a decision problem based on the two types of errors. There is no need to specify prior distributions, this is not a Bayesian setting. There are two hypotheses, the null H_0 and the alternative H_A. Each of these are simple hypotheses in that we completely specify a probability distribution for each. In this way, we can construct a likelihood ratio.

- **Type I Error**: Rejecting H_0 when it is true
- **Type II Error**: Accepting H_0 when it is false
- **Power**: Probability that H_0 is rejected when it is false (1 - Type II Error)

Their decision rule rejects H_0 whenever the likelihood ratio is less than some constant c and significance level α.

$$d(X_1, \dots, X_n) = \begin{cases} 0 & \text{fail to reject } H_0 \\ 1 & \text{reject } H_0 \end{cases}$$

The Neyman-Pearson Lemma tells us that the likelihood ratio test (LRT) at level α is most-powerful level-α test of H_0 vs. H_A. There are many possible decision rules one can choose, but this is telling us that among all possible ones for which the Type I Error probability is less than or equal to $1 - \alpha$, the likelihood ratio test is the one with the greatest power. This is simply a binary decision rule not related to the truth. Thus it is not rejecting when H_0 is false nor accepting when it is true.

The level-α test relies on a critical inequality. So let d be the likelihood ratio test. We reject the null if the likelihood under the null is less than some constant times the likelihood under the alternative. Says the data is sufficiently less likely.

Proof

Our decision rule is $d(x) = 1$ if $f_0(x) < c\, f_A(x)$. We set our Type I Error, $E_0(d(X)) = \alpha$ and consider **any other binary decision rule** d^* where $E_0(d^*(X)) \leq E_0(d(X)) = \alpha$.

We want to now show that $E_A(d^*(X)) \leq E_A(d(X))$, that is that the power of the LRT is always greater. We will rely on showing the following inequality.

$d^*(x_1, \dots, x_n)[cf_A(x_1, \dots, x_n) - f_0(x_1, \dots, x_n)] \leq d(x_1, \dots, x_n)[cf_A(x_1, \dots, x_n) - f_0(x_1, \dots, x_n)]$

We use that fact that both rules are binary.

Case 1: $cf_A(x_1, \dots, x_n) - f_0(x_1, \dots, x_n) > 0$

- If this is the case, then $d(x) = 1$. Since $d^*(x) \in \{0, 1\}$this holds.

Case 2: $cf_A(x_1, \dots, x_n) - f_0(x_1, \dots, x_n) \leq 0$

- Here, $d(x) = 0$ and so RHS = 0. Since $d^*(x) \in \{0, 1\}$ LHS is less than or equal to 0.

Note that in terms of expectations,

$E_0[d(X)] = 1 \times P_0(d(X) = 1) + 0 \times P_0(d(X) = 0) = P(d = 1 | H_0) = P(\text{Type I Error})$.

$E_A[d(X)] = P_A(d(X) = 1) = \text{Power}$.

So we now take expectations in the inequality, which correspond to probabilities since the decision rules are binary.

$cE_A[d^*(X)] - E_0[d^*(X)] \leq cE_A[d(X)] - E_0[d(X)]$

Rearranging,

$E_0[d(X)] - E_0[d^*(X)] \leq c(E_A[d(X)] - E_A[d^*(X)])$

The left side is the difference in the probability Type one error, which we have said is greater than or equal to 0. Thus, the LRT always has a equivalent or greater power since the RHS is nonnegative.

NOT BAYESIAN SETTING. Either the null is true or it isn't, no probabilities.

Example A (page 333)

Let $X_1, \ldots, X_n$ be a random sample from a normal distribution having known variance σ^2. Consider two simple hypotheses:

$$H_0\colon \mu = \mu_0$$
$$H_A\colon \mu = \mu_1$$

where μ_0 and μ_1 are given constants. Let the significance level α be prescribed. The Neyman-Pearson Lemma states that among all tests with significance level α, the test that rejects for small values of the likelihood ratio is most powerful. We thus calculate the likelihood ratio, which is

$$\frac{f_0(\mathbf{X})}{f_1(\mathbf{X})} = \frac{\exp\left[\dfrac{-1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i - \mu_0)^2\right]}{\exp\left[\dfrac{-1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i - \mu_1)^2\right]}$$

since the multipliers of the exponentials cancel. Small values of this statistic correspond to small values of $\sum_{i=1}^{n}(X_i - \mu_1)^2 - \sum_{i=1}^{n}(X_i - \mu_0)^2$. Expanding the squares, we see that the latter expression reduces to

$$2n\overline{X}(\mu_0 - \mu_1) + n\mu_1^2 - n\mu_0^2$$

Now, if $\mu_0 - \mu_1 > 0$, the likelihood ratio is small if $\overline{X}$ is small; if $\mu_0 - \mu_1 < 0$, the likelihood ratio is small if X is large. To be concrete, let us assume the latter case. We then know that the likelihood ratio is a function of $\overline{X}$ and is small when $\overline{X}$ is large. The Neyman-Pearson lemma thus tells us that the most powerful test rejects for $\overline{X} > x_0$ for some x_0, and we choose x_0 so as to give the test the desired level α. That is, x_0 is chosen so that $P(\overline{X} > x_0) = \alpha$ if H_0 is true. Under H_0 in this example, the null distribution of $\overline{X}$ is a normal distribution with mean μ_0 and variance σ^2/n, so x_0 can be chosen from tables of the standard normal distribution. Since

$$P(\overline{X} > x_0) = P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{x_0 - \mu_0}{\sigma/\sqrt{n}}\right)$$

we can solve

$$\frac{x_0 - \mu_0}{\sigma/\sqrt{n}} = z(\alpha)$$

for x_0 in order to find the rejection region for a level α test. Here, as usual, $z(\alpha)$ denotes the upper α point of the standard normal distribution; that is, if Z is a standard normal random variable, $P(Z > z(\alpha)) = \alpha$. ■

Uniformly Most Powerful (UMP)

After example A, we determine a region for the test based on the distribution under the null. But the actual value of the point of rejection is **independent of μ_1 as long as $\mu_1 > \mu_0$.**

Rejection region can be determined just by alpha and standard normal critical values, NOT μ_1, the exact alternative doesn't not matter. The Neyman-Pearson Lemma required both hypotheses to be *simple*, but here our alternative can be *composite* in that it need not fully specify the distribution. Because this test is the most powerful **and** is the same for every alternative of the form

$H_0 : \mu \leq \mu_0$
$H_1 : \mu > \mu_0$

it is **uniformly most powerful** for this hypothesis test. This is called a **one-sided alternative**. Note that a two sided alternative wouldn't work.