

IEOR 165 – Lecture 25

Power Analysis

1 Problem Statement

In the Neyman-Pearson framework, it is common to begin the analysis before any data is collected. The reason is that it is natural to ask if there is a methodological approach for determining how much data should be gathered? This can be quite important in practice because gathering data can be monetarily expensive or costly in other factors like time.

Suppose the size of the test α is fixed at a desired value, then the power of the test becomes fixed. However, by choosing the amount of data n , we can adjust the power of the test. However, implicit in this is a notion of *effect size*. This is a subtle notion, and so it best illustrated by an example: Suppose iid data is from $X_i \sim \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known, and consider the hypothesis

$$\begin{aligned}H_0 : \mu &\leq 0 \\H_1 : \mu &> 0.\end{aligned}$$

In this example, the power of the test will be

$$\inf_{\mu > 0} \beta(\mu) = \beta(0) = \alpha,$$

since, as discussed previously, the power

$$\beta(\mu) = \mathbb{P}_{\mu} \left(\sqrt{n} \frac{(\bar{X} - \mu)}{\sigma} \geq c - \sqrt{n} \frac{\mu}{\sigma} \right).$$

is increasing in μ . This is problematic because the power of the test must necessarily depend on the size of the test α , and it does not depend on the sample size n .

The issue in the above example is that there we have ignored the effect size in computing the power. It is often the case that if the true mean is small (for example $\mu = 0.001$), then we are not interested having a large power for this instance. We will often have a minimum difference for which differences larger than this are more interesting to detect. The notion of effect size is how much difference would we like to have high power to detect between the null and alternative hypothesis? In the case of the example, we might say that there is some known value Δ such that we are interested in determining if $\mu \geq \Delta$ (versus the null hypothesis). In this case, the power is now given by

$$\inf_{\mu \geq \Delta} \beta(\mu) = \beta(\Delta).$$

The region of $0 < \mu < \Delta$ is called the *indifference region* because, though it is part of the alternative hypothesis, we are not interested in having a high power for this region of possible values of μ .

It turns out that the ability to methodologically choose the sample size is an advantage of the Neyman-Pearson framework. We cannot do this for the setting of null hypothesis testing because there is no notion of an alternative hypothesis.

2 Example: One-Sided Test for Gaussian with Known Variance

Suppose iid data is from $X_i \sim \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known, and consider the hypothesis

$$\begin{aligned} H_0 : \mu &\leq 0 \\ H_1 : \mu &> 0. \end{aligned}$$

Now imagine that we have fixed the size of the test to α , and we would like the power of the test to be some value of β that we have chosen to be close to one. Recall that the test has size α if we reject the test whenever $\sqrt{n}\bar{X}/\sigma > z(1 - \alpha)$. And so we would like

$$\begin{aligned} \beta = \beta(\Delta) &= \mathbb{P}\left(Z \geq z(1 - \alpha) - \sqrt{n}\frac{\Delta}{\sigma}\right) \\ &= \mathbb{P}\left(-Z \leq -z(1 - \alpha) + \sqrt{n}\frac{\Delta}{\sigma}\right) \\ &= \mathbb{P}\left(Z \leq -z(1 - \alpha) + \sqrt{n}\frac{\Delta}{\sigma}\right) \end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$, and we have used the fact that Z and $-Z$ have the same distribution. Consequently, by definition we have that

$$-z(1 - \alpha) + \sqrt{n}\frac{\Delta}{\sigma} = z(\beta) \Rightarrow n = (\Delta/\sigma)^{-2}(z(1 - \alpha) + z(\beta))^2.$$

3 Monte Carlo Algorithm

Finding an equation to choose the sample size can be difficult, and so one possible approach is to use a Monte Carlo algorithm to help choose the sample size. Suppose we have a parametric hypothesis class \mathbb{P}_θ with a simple null and alternative hypothesis

$$\begin{aligned} H_0 : \theta &= \theta_0 \\ H_1 : \theta &= \theta_1. \end{aligned}$$

The algorithm will take as input

- Desired size α ;
- Desired power β ;

- Likelihood ratio $L(X, \theta_0, \theta_1)$;
- Function to generate random samples from \mathbb{P}_θ ;
- Set of thresholds $\{c_1, c_2, \dots, c_m\}$;
- Set of sample sizes $\{n_1, n_2, \dots, n_q\}$;
- Repetition count M , which is a large number (say $M = 1000$).

The algorithm is as follows:

1. For $j = 1, \dots, q$
 - (a) For $k = 1, \dots, m$
 - i. For $s = 1, \dots, M$
 - A. Randomly choose $X_i \sim \mathbb{P}_{\theta_0}$, for $i = 1, \dots, n_j$;
 - B. Randomly choose $Y_i \sim \mathbb{P}_{\theta_1}$, for $i = 1, \dots, n_j$;
 - C. If $L(X, \theta_0, \theta_1) > c_k$, then set $\delta_s = 1$;
 - D. If $L(X, \theta_0, \theta_1) < c_k$, then set $\delta_s = 0$;
 - E. If $L(Y, \theta_0, \theta_1) > c_k$, then set $\gamma_s = 1$;
 - F. If $L(Y, \theta_0, \theta_1) < c_k$, then set $\gamma_s = 0$;
 - ii. Set $a_k = \delta_s/M$;
 - iii. Set $b_k = \gamma_s/M$;
 - (b) Define λ to be the index of the largest value of $a_\lambda \in \{a_1, \dots, a_m\}$ such that a_λ is less than or equal to α . (If no such value exists, then the input needs to be modified to allow smaller thresholds);
 - (c) If $b_\lambda \geq \beta$, then return the values: sample size should be $n = n_j$ and threshold should be $c = c_\lambda$.

If no value is returned, then this means that the input needs to be modified to allow larger sample sizes.

4 More Information and References

The material in these notes follows that of the textbook “Mathematical Statistics, Basic Ideas and Selected Topics, Vol. 1, (2nd Edition)” by Peter Bickel and Kjell Doksum.