INSE6300	Jia Yuan Yu
6: Hypothesis Testing	
Concordia	November 26, 2017

We have seen different estimators for the unknown parameter of an unknown probability distribution that belongs to a known set.

In hypothesis testing, we have observations of a random variable, whose distrubition is unknown, up to a set of distributions $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$. A hypothesis is a subset H of $\mathcal{P}$. There are two decisions: accept or reject the hypothesis.

Example 0.1 (Pride and Prejudice)**.** This situation is like that of the book Pride and Prejudice: one hypothesis is that Darcy is despicable, the other is that Darcy is lovely. You observe samples throughout the book, and must make a decision; the initial samples may lead you astray.

1 Finite-valued distributions

Suppose that the observations $X_1, \dots, X_n$ takes values in a finite set $M = \{1, \dots, m\}$. Consider $\mathcal{P} = \{P_0, P_1\}$, and the hypothesis $H_0 = \{P_0\}$. The complement of the hypothesis is called the alternative and denoted H_1.

Let d_0 and d_1 denote accepting H_0 or H_1 respectively. A nonrandomized decision rule is a mapping $\delta : M^n \to \{d_0, d_1\}$. As in the study of control charts, there are two types of errors associated with the two decisions and two hypotheses:

- When $X_1, \dots, X_n$ are distributed according to P_0, but $\delta(X_1, \dots, X_n) = d_1$;
- When $X_1, \dots, X_n$ are distributed according to P_1, but $\delta(X_1, \dots, X_n) = d_0$.

One objective of quality assurance or performance guarantees is to find decision rules that tradeoff the two types of errors.

Let us write $X = (X_1, \dots, X_n)$.

For a given $\alpha \geq 0$, one objective in the choice of δ is to minimize the probability of one type of errors:

$$\mathbb{P}_1(\delta(X) = d_0),$$

subject to the contraint that the probability of error of the other type is below $\alpha > 0$:

$$\mathbb{P}_0(\delta(X) = d_1) \leq \alpha.$$

In other words, we want to minimize false-alarms, subject to a constraint on missed detections. This is equivalent to:

$$\begin{aligned} \max_{\delta} \quad & \mathbb{P}_1(\delta(X) = d_1) \\ \text{s.t.} \quad & \mathbb{P}_0(\delta(X) = d_1) \leq \alpha. \end{aligned}$$

Observe that a nonrandomized decision rule δ is described by a subset S_δ of M: it is a lookup table of the form

x	$\delta(x)$	$P_0(X=x)$	$P_1(X=x)$	$P_0(X=x)/P_1(X=x)$
1	d_0	0.1	0.2	0.5
2	d_1	0.04	0.02	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m^n	d_0	0.05	0.05	1

The subset S_δ contains the elements of M where $\delta(x) = d_1$ (reject H_0). Observe that

$$\mathbb{P}_1(\delta(X) = d_1) = \sum_{x \in S_\delta} \mathbb{P}_1(X = x),$$
$$\mathbb{P}_0(\delta(X) = d_1) = \sum_{x \in S_\delta} \mathbb{P}_0(X = x).$$

Therefore, we can find the optimal decision rule δ by finding the subset A which solves the following:

$$\begin{aligned} \max_{A \subseteq M} \quad & \sum_{x \in A} P_1(X = x) \\ \text{subject to} \quad & \sum_{x \in A} P_0(X = x) \leq \alpha. \end{aligned}$$

Remark 1 (Continuous-valued random variables). If X is a continuous-valued measurement, and the probability distributions P_0 and P_1 have densities f_0 and f_1, then the optimization problem becomes

$$\begin{aligned} \max_{A \subseteq M} \quad & \int_{x \in A} f_1(x) \mathrm{d}x \\ \text{subject to} \quad & \int_{x \in A} f_0(x) \mathrm{d}x \leq \alpha. \end{aligned}$$

2 Solution approach

One method to solve the above optimization is to rank all $x \in M$ according to

$$\frac{P_0(X = x)}{P_1(X = x)},$$

and adding elements to S_δ until the threshold α is reached. More precisely, consider the following decision rule. Given a threshold $\lambda > 0$,

$$\delta_\lambda(x) = \begin{cases} d_0 & \text{if } \frac{P_0(X=x)}{P_1(X=x)} > \lambda, \\ d_1 & \text{otherwise.} \end{cases}$$

Lemma 2.1 (Neyman-Pearson)**.** *Let X be a random variable taking a finite set M of values. Suppose that the rule δ_λ has error probability $\mathbb{P}_0(\delta_\lambda(X) = d_1) = \alpha$. Then for every other decision rule δ with $\mathbb{P}_0(\delta(X) = d_1) \leq \alpha$, the probability of correct decision is not higher than δ_λ:*

$$\mathbb{P}_1(\delta(X) = d_1) \leq \mathbb{P}_1(\delta_\lambda(X) = d_1).$$

Proof. Let S denote the subset of M where δ_λ decides d_1. Since $\lambda P_1(X = x) - P_0(X = x) \geq 0$ for $x \in S$ and $\lambda P_1(X = x) - P_0(X = x) < 0$ for $x \in S$, we conclude that for every other set $A \subseteq M$:

$$\sum_{x \in S}(\lambda P_1(X = x) - P_0(X = x)) \geq \sum_{x \in A}(\lambda P_1(X = x) - P_0(X = x)).$$

By algebra, we obtain:

$$\begin{aligned}\lambda\left(\sum_{x \in S} P_1(X = x) - \sum_{x \in A} P_1(X = x)\right) &\geq \sum_{x \in S} P_0(X = x) - \sum_{x \in A} P_0(X = x) \\ &= \mathbb{P}_0(\delta_\lambda(X) = d_1) - \sum_{x \in A} P_0(X = x) \geq 0,\end{aligned}$$

where the last inequality is by assumption. Hence, we can conclude that $\sum_{x \in S} P_1(X = x) \geq \sum_{x \in A} P_1(X = x)$. □

Remark 2 (Randomized rules). Mapping $x \in M$ to a probability distribution $\phi(x)$, then flip a coin with probability $\phi(x)$ to determine accept or reject.

3 Example: finite distribution

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli distributed. Let H_0 correspond to the Bernoulli distribution with $p = 1/2$. Let H_1 correspond to the Bernoulli distribution $N(\mu, 1)$ with $p \neq 1/2$. The likelihood ratio is

$$\begin{aligned}\frac{f_0(x_1, \ldots, x_n)}{f_1(x_1, \ldots, x_n)} &= \frac{0.5^n}{\prod_i p^{x_i}(1-p)^{1-x_i}} \\ &= \frac{0.5^n}{(1-p)^n}\left(\frac{1-p}{p}\right)^{\sum_i x_i}\end{aligned}$$

Given a $\lambda > 0$, the decision rule δ_λ has error probability

$$\mathbb{P}_0(\delta_\lambda(X) = d_1) = \mathbb{P}_0\left(\frac{0.5^n}{(1-p)^n}\left(\frac{1-p}{p}\right)^{\sum_i x_i} \leq \lambda\right).$$

We can calculate this probability using the fact that the sum of Bernoulli random variables $\sum_i x_i$ is a binomial random variable.

If we want $\mathbb{P}_0(\delta_\lambda(X) = d_1) \leq 0.1$, what value should λ take?

4 Example: continuous distribution

Let $X_1, \ldots, X_n$ be i.i.d. normally distributed. Let H_0 correspond to the distribution $N(0,1)$. Let H_1 correspond to the distribution $N(\mu, 1)$ for a given $\mu > 0$. The likelihood ratio is

$$\begin{aligned}\frac{f_0(x_1,\ldots,x_n)}{f_1(x_1,\ldots,x_n)} &= \frac{e^{-x_1^2/2}\ldots e^{-x_n^2/2}}{e^{-(x_1-\mu)^2/2}\ldots e^{-(x_n-\mu)^2/2}} \\ &= \exp\left(-\frac{1}{2}\sum_i x_i^2 + \frac{1}{2}\sum_j (x_j-\mu)^2\right) \\ &= \exp\left(\frac{n\mu^2 - 2\mu\sum_i X_i}{2}\right).\end{aligned}$$

Given a $\lambda > 0$, the decision rule δ_λ has error probability

$$\mathbb{P}_0(\delta_\lambda(X) = d_1) = \mathbb{P}_0\left(\exp\left(\frac{n\mu^2 - 2\mu\sum_i X_i}{2}\right) \leq \lambda\right).$$

If we want $\mathbb{P}_0(\delta_\lambda(X) = d_1) \leq 0.1$, what value should λ take?

5 References

- Lehmann and Romano's Testing Statistical Hypotheses.
- Robert W. Keener's "Theoretical Statistics: Topics for a Core Course."