IEOR 151 – Lecture 3
Risk in Decision Making

1 Motivating Example

Suppose there is a chain of fast food restaurants that only offers lunch/dinner options on its menu. Given the potential for a new revenue source, the chain is deciding whether it should begin to offer breakfast options for its menu. The introduction of breakfast is not without risk, though, because of the costs associated with expanding the menu. There are both additional fixed and recurring costs associated with this change. Based on these costs, the chain has determined that offering breakfast is financially beneficial only if the additional revenue in each restaurant exceeds \$7500.

To assist in this decision, the chain has tested the new menu in a limited number of randomly selected locations. Specifically, the chain has implemented the new menu in 20 restaurant locations and found that the average increase in revenue is \$8214. Furthermore, an analyst has conducted a one-sided hypothesis test (with null hypothesis that the revenue increase is \$7500 or lower) to determine if the revenue increase is from the breakeven point of \$7500 is statistically significant, and the resulting p-value is $p = 0.045$.

A common approach to decision-making using hypothesis tests is to (i) choose a significance level α and then (ii) reject the null hypothesis whenever the p-value is smaller. One choice is $\alpha = 0.05$, and so we would reject the null hypothesis in this case based on this analysis. However, the choice of $\alpha = 0.05$ is somewhat arbitrary. More importantly, this approach does not explicitly incorporate the notion of risk of making incorrect decisions in either direction.

2 Loss Functions

In order to be able to specify risk, it is useful to make a few definitions. One framework for decision making incorporates the following three elements

1. X is the set of data we have collected, and whose distribution $f_X(u; \theta)$ is partly known; the θ is a parameter that labels the different possible distributions, and it belongs to some known set Ω call the *parameter space*;

2. $\delta(u)$ is a function known as a *decision rule*, and it inputs data and outputs a decision d; the set of possible decisions is $\mathcal{D}$;

3. There is a loss associated with making decision d when the distribution of X is $f_X(u;\theta)$; this is expressed as a nonnegative real number $L(\theta, d)$, and it is known as a *loss function*.

The significance of a loss function $L(\theta, d)$ is that if we were to use our decision rule δ to make many decisions in a case where $f_X(u;\theta)$ is the true distribution of X, then the sample average of the loss will converge to the expected loss

$$R(\theta, \delta) = \mathbb{E}(L(\theta, \delta(X))).$$

This function $R(\theta, \delta)$ is known as a risk function. In the statistics literature, it is convention to indicate undesirable results with high values of loss and risk.

2.1 Point Gaussian Example

Suppose $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, 1)$ is iid data drawn from a normal distribution with mean μ and variance $\sigma^2 = 1$; here, the mean is unknown. This specifies the first element for decision making.

Consider a situation in which the decision we would like to make is whether the mean is μ_0 or μ_1. There are two possible decisions

1. d_0 is the decision that the mean is μ_0;
2. d_1 is the decision that the mean is μ_1.

For convenience, we will assume here that $\mu_0 < \mu_1$.

Specifying a loss function is difficult because it is situation dependent, but one loss that applies to this example is

- $L(\mu_0, d_0) = 0$ and $L(\mu_0, d_1) = a$; in the situation where the true mean is μ_0, there is no loss if we correctly decide the mean is μ_0, otherwise we have a loss of a if we decide the mean is μ_1;
- $L(\mu_1, d_0) = b$ and $L(\mu_1, d_1) = 0$; in the situation where the true mean is μ_1, there is no loss if we correctly decide the mean is μ_1, otherwise we have a loss of b if we decide the mean is μ_0.

We will assume that these a, b values are positive: $a > 0$ and $b > 0$.

3 Minimax Procedures

The complication associated with this formulation is that constructing a decision rule $\delta(u)$ to minimize the risk function $R(\theta, \delta)$ requires knowing the parameter θ, which is unknown. And so we need some way to compare decision rules so that we can select a "best" decision rule. There are multiple approaches to this, and one approach is the minimax procedure.

In this method, we would like to pick a decision rule to that solves the following optimization problem

$$\min_{\delta(u)} \max_{\theta \in \Omega} R(\theta, \delta).$$

In general, solving this optimization problem is very difficult. However, it can be solved in a few special cases.

3.1 Point Gaussian Example

Recall our Gaussian example, and suppose we would like to compute a minimax hypothesis test. This means that we would like to solve

$$\min_{\delta(u)} \max_{\mu \in \{\mu_0, \mu_1\}} \mathbb{E}(L(\mu, \delta(X))).$$

Note that if $\mu = \mu_0$, then

$$\begin{aligned}\mathbb{E}(L(\mu, \delta(X))) &= \mathbb{E}(L(\mu_0, \delta(X))) \\ &= L(\mu_0, d_0)\mathbb{P}_{\mu_0}(d_0 = \delta(X)) + L(\mu_0, d_1)\mathbb{P}_{\mu_0}(d_1 = \delta(X)) \\ &= a \cdot \mathbb{P}_{\mu_0}(d_1 = \delta(X)),\end{aligned}$$

since $L(\mu_0, d_0) = 0$ and $L(\mu_0, d_1) = a$. A similar computation gives

$$\mathbb{E}(L(\mu, \delta(X))) = b \cdot \mathbb{P}_{\mu_1}(d_0 = \delta(X)).$$

Combining everything, the optimization problem we would like to solve is

$$\min_{\delta(u)} \max\{a \cdot \mathbb{P}_{\mu_0}(d_1 = \delta(X)), b \cdot \mathbb{P}_{\mu_1}(d_0 = \delta(X))\}.$$

One reason this is a difficult to solve is that it is an infinite-dimensional optimization problem because $\delta(u)$ is a function to be optimized over. For simplicity, suppose the decision rule is given by

$$\delta(u) = \begin{cases} d_0, & \text{if } \frac{1}{n}\sum_{i=1}^n X_i \leq \gamma \\ d_1, & \text{if } \frac{1}{n}\sum_{i=1}^n X_i > \gamma \end{cases}$$

for some value of γ that we will compute. By imposing this specific decision rule, we have converted our infinite-dimensional optimization problem over $\delta(u)$ into a finite-dimensional problem over γ.

For notational simplicity, we will denote the sample average by $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Now suppose we choose γ such that

$$a \cdot \mathbb{P}_{\mu_0}(\overline{X} > \gamma) = b \cdot \mathbb{P}_{\mu_1}(\overline{X} \leq \gamma).$$

Since each $X_i \sim \mathcal{N}(\mu, 1)$ are iid, the sample average $\overline{X}$ is also Gaussian with mean μ and variance $1/n$. As a result, we can restate the condition we use to choose γ so that it satisfies

$$a \cdot (1 - \Phi(\sqrt{n}(\gamma - \mu_0))) = b \cdot \Phi(\sqrt{n}(\gamma - \mu_1)),$$

where $\Phi(\cdot)$ is the cdf of a normal distribution and can be found from a standard z-table or using a computer. If we call this resulting value γ^*, then our decision rule is

$$\delta(X) = \begin{cases} d_0, & \text{if } \overline{X} \leq \gamma^* \\ d_1, & \text{if } \overline{X} > \gamma^* \end{cases}$$

It turns out that this decision rule is a minimax procedure (meaning that it solves the minimax optimization problem above), though showing this is beyond the scope of the class.

4 Further Details

More details about these concepts can be found in the book *Testing Statistical Hypotheses* by Lehmann and Romano, from which much of the above material is found.