IEOR 151 – Lecture 4
Composite Minimax

1 Numerical Example for Point Gaussian Example

1.1 Computing γ

Suppose $X_i \sim \mathcal{N}(\mu, \sigma^2)$ (for $n = 20$ data points) is iid data drawn from a normal distribution with mean μ and variance $\sigma^2 = 20$. Here, the mean is unknown, and we would like to determine if the mean is $\mu_0 = 0$ (decision d_0) or $\mu_1 = 4$ (decision d_1). Lastly, suppose our loss function is

- $L(\mu_0, d_0) = 0$ and $L(\mu_0, d_1) = a = 3$;
- $L(\mu_1, d_0) = b = 2$ and $L(\mu_1, d_1) = 0$.

Recall that the minimax hypothesis test is given by

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

where γ_n^* is the value of γ that satisfies

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

The $\Phi(\cdot)$ denotes the cdf of a normal distribution and can be found from a standard z-table or using a computer. The trick to finding this γ value when using a z-table is to observe that the left hand side (LHS) decreases as γ increases, while the right hand side (RHS) increases while γ increases.

In our case, we would like to find the γ that satisfies

$$3 \cdot (1 - \Phi(\sqrt{20}(\gamma - 0)/\sqrt{20})) = 2 \cdot \Phi(\sqrt{20}(\gamma - 4)/\sqrt{20}),$$

or equivalently

$$3 \cdot (1 - \Phi(\gamma - 0)) = 2 \cdot \Phi(\gamma - 4),$$

We will do a search by hand to find the corresponding value of γ. For instance, if our first guess is $\gamma = 2$, then we find from the z-table that $\Phi(2) = 0.9773$ and $\Phi(2-4) = \Phi(-2) = 1 - \Phi(2) = 0.0227$. Thus, we have $3 \cdot (1 - \Phi(2)) = 3 \cdot (1 - 0.9733) = 0.0801$ and $2 \cdot \Phi(-2) = 0.0454$. Since the LHS is larger, this means we should increase γ.

Now suppose our second guess is $\gamma = 2.5$. Then, $\Phi(2.5) = 0.9938$ and $\Phi(2.5 - 4) = \Phi(-1.5) = 1 - \Phi(1.5) = 1 - 0.9332$. Thus, we have that $LHS = 3 \cdot (1 - 0.9938) = 0.0186$ and $RHS = 2 \cdot (1 - 0.9332) = 0.1336$. Now, the RHS is larger and so we should decrease our guess of γ. Since we know that $\gamma = 2$ is too small, we could try half-way in between with $\gamma = 2.25$.

We can summarize the steps We conclude the process when we have sufficient accuracy

Step	γ	LHS	RHS
1	2	0.0801	0.0454
2	2.5	0.0186	0.1336
3	2.25	0.0367	0.0801
4	2.13	0.0498	0.0615
5	2.06	0.0591	0.0524
6	2.09	0.0549	0.0561
7	2.07	0.0577	0.0536
8	2.08	0.0563	0.0549

in our computed value of γ. In this case, we know that γ should be between 2.08 and 2.09, and so we set $\gamma_n^* = 2.085$. Computing γ to more precision would require a computer.

1.2 Differences in γ as n Changes

In the above example, we computed γ_n^* for a single value of $n = 20$. A natural question to ask is what happens to γ as n increase. In the table below, the value of γ_n^* for different values of n is given. These values were computed using a computer.

n	γ_n^*
5	2.2658
10	2.1536
20	2.0855
100	2.0194
200	2.0099
300	2.0068

The trend is clear: As n increases, γ_n^* is decreasing towards 2. The intuition is that when we have little data, we err on the side of deciding d_0 since otherwise we incur a larger loss if we incorrectly decide d_1. As we gather more data, we are more confident that the sample average is close to the true average, and so we can use a less biased threshold. Effectively, with large amounts of data the threshold converges to $(\mu_0 + \mu_1)/2$.

2 Composite Gaussian Example

In our discussion so far, we have considered a situation in which the two hypothesis each represent a distribution with a single mean. However, another class of interesting and more general hypotheses are those in which we would like to discrimination between $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$. And suppose we keep a similar loss function of $L(H_0, d_0) = 0$, $L(H_0, d_1) = a$, $L(H_1, d_0) = b$, and $L(H_1, d_1) = 0$. In the minimax framework, this class of hypotheses are not well-posed because the worst case scenario occurs when "nature" selects $\mu = \mu_0$, because then it is not possible to distinguish between d_0 and d_1.

More rigorously, what happens is that the minimax procedure is defined by solving the optimization problem

$$\inf_{\delta(u)} \sup_{\mu} R(\mu, \delta).$$

And nature will choose

$$\begin{aligned}\sup_{\mu:\mu\leq\mu_0} R(\mu, \delta) &= \sup_{\mu:\mu\leq\mu_0} a \cdot \mathbb{P}_\mu(d_1 = \delta(X))\} \\ &= a \cdot \mathbb{P}_{\mu_0}(d_1 = \delta(X))\}\end{aligned}$$

and

$$\begin{aligned}\sup_{\mu:\mu>\mu_0} R(\mu, \delta) &= \sup_{\mu:\mu>\mu_0} b \cdot \mathbb{P}_\mu(d_0 = \delta(X))\} \\ &= b \cdot \mathbb{P}_{\mu_0}(d_0 = \delta(X))\}.\end{aligned}$$

This last step is subtle: Even though nature is constrained to choose $\mu > \mu_0$, it can choose μ to be arbitarily close to μ_0. Thus, in the worst case scenario described by the minimax framework, nature will effectively set $\mu = \mu_0$ even though we are in the case $H1 : \mu > \mu_0$.

Because of this pathological behavior, the best we can do with a minimax procedure is to make a purely probabilistic decision:

$$\delta(X) = \begin{cases} d_0, & \text{with probability } a/(a+b) \\ d_1 & \text{w.p. } b/(a+b) \end{cases}$$

This is not a useful rule for decision making, and so we must consider alternative classes of hypotheses.

One interesting class of hypotheses is the following: Suppose $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ is iid data drawn from a normal distribution with mean μ and known variance σ^2; here, the mean is unknown. The decision we would like to make is whether the mean is $H_0 : \mu \leq \mu_0$ or $H_1 : \mu \geq \mu_1$. We are *indifferent* in the case where $I : \mu \in (\mu_0, \mu_1)$. Our loss function also encodes this region of indifference

- $L(H_0, d_0) = 0$, $L(H_0, d_1) = a$;
- $L(H_1, d_0) = b$, $L(H_1, d_1) = 0$;
- $L(I, d_0) = 0$, $L(I, d_1) = 0$.

It turns out that this *composite* hypothesis test has the same minimax procedure as the point hypothesis test. To summarize, we choose γ so that it satisfies

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

where $\Phi(\cdot)$ is the cdf of a normal distribution. If we call this resulting value γ^*, then our decision rule is

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

Showing that this decision rule is a minimax procedure is beyond the scope of the class.