

Session 12

The decision problem

A pub owner in Charlottesville plans to sell beer outside on St. Patrick's Day, March 17. The pub owner must decide whether to arrange to rent a supplemental refrigeration system for the day.

Supplemental refrigeration offers a form of insurance. If temperatures outside on March 17 are high and the pub owner has not arranged supplemental refrigeration, she will be left with warm beer that she will have difficulty selling, leading to financial losses. Conversely, if she pays for supplemental refrigeration when temperatures are low, she will have incurred an unnecessary expense.

The analysis here uses statistical decision theory to generate a recommendation about whether or not to pay to rent a supplemental refrigeration system, in order to insure against the potential of financial losses in case of high temperatures.

Formalism

Action set

The action set \mathbb{A} includes just two elements: $\mathbb{A} = \{a_1, a_2\}$, where a_1 and a_2 correspond to the decisions to not purchase insurance, and to purchase insurance, respectively.

Payoffs: The loss function

Let $L(a, y)$ denote the losses incurred by the publican. These depend on the action chosen, and on y , the high temperature in Charlottesville on March 17, measured in degrees Celsius.

Suppose we have two possible actions $a \in a_1, a_2$, and we want to predict the weather temperature y_{n+1} , and based on that we want to decide about the two following actions.

The payoff only depends on the true value of some $L(a, y_{n+1})$

a_1 : not purchase insurance

a_2 : purchase insurance

Historical data on $y = y_1, y_2, \dots, y_n$. This is the history of the temperature in this example.

We want:

- probabilistic forecast of y_{n+1}
- use prediction to choose optimal a^*

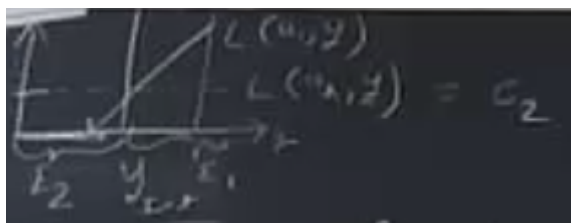


Figure 1: 1

What to do?

We want the probabilistic forecast; we want to know the distribution probability for y_{n+1} .

Sometimes the payoff is not symmetric. If the temperature is high, the loss is high. If the temperature is low, the loss is low.

Based on the loss function which action should be chosen?

$$\hat{L}_{emp} = \frac{\sum_{i=1}^n (a_1, y_i)}{n}$$

Define events:

$$E_1 = \{y | y_{n+1} \leq y_{int}\} \quad a^* = a_1$$

$$E_2 = \{y | y_{n+1} \geq y_{int}\} \quad a^* = a_2$$

$$\text{Want } Pr\{E_1\} = Pr\{y \leq y_{int}\} = \frac{1}{n} \sum_{y \leq y_{int}} (1) \quad \text{Estimate of } Pr(E_1)$$

$$\text{If } \hat{L}_{emp} \geq c_2, \text{ then } a^* = a_2$$

Need some model of the data generating process

$$y_i = y_0 + \varepsilon_i, \varepsilon_i \sim N(0, \sigma_y^2)$$

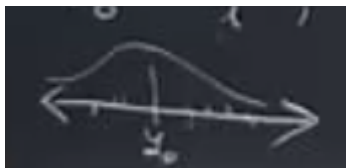


Figure 2: 2

Here, the mean and variance are unknown.

$$\text{Equivalently } y_i \sim N(y_i, \sigma_y^2)$$

We need to find the distribution of the losses over the different actions. To do that, we need the distribution of the y

Want the probability forecast of y_{n+1}

To get it, estimate the y_0, σ_y^2 from y_1, y_2, \dots, y_n

There are multiple ways to do that.

Standard approach: OLS

mean: $\min_{\hat{y}_0} \sum (\hat{y}_0 - y_i)^2 = \sum (\hat{\varepsilon}_i^2)$

So, we will get the y_0, σ_y^2

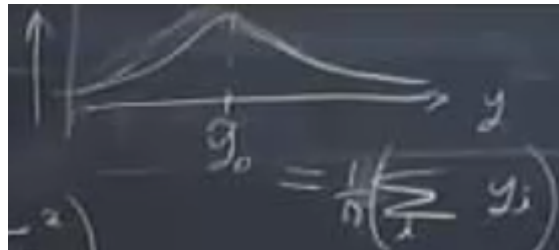


Figure 3: 3

This is the distribution y . This describes the probability forecast over the next value of y . Using it, we can get the probability forecast for losses.

Suppose, you plot your estimated errors $\hat{\varepsilon}_i = y_i - \hat{y}_0$

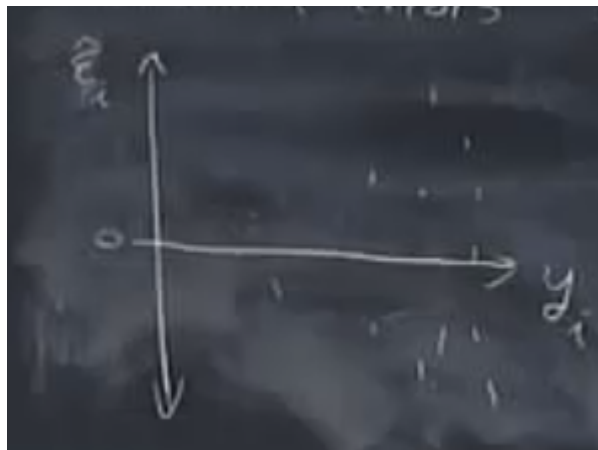


Figure 4: 4

The variance of the error tends to be growing. It violates the assumption of the independent identically distribution. This is called the heteroskedasticity. When you work with data, you need to look at your data. In overall, if heteroskedasticity happens, we would have the wrong estimate of probabilities.