CHAPTER 1

PROBABILISTIC MODELS

1.1 INTRODUCTION TO PROBABILISTIC MODELS

Deterministic models assume that outcomes are precisely determined through known relationships among states and events. However, many systems in the real world exhibit inherent randomness and uncertainty. To effectively model and analyze such systems, we turn to **probabilistic models**.

Definition 1.1.1 (**Probabilistic Model**)
A **probabilistic model** is a model that incorporates *random variables* and *probability distrubutions*.

In a a probabilistic model:

- **Random variables** represent the potential outcomes of uncertain events.
- **Probability distrubutions** assign probabilities to the various potential outcomes.

Observation 1.1.1 (**Random Variables and Probability Distrubutions**)
In simple terms:

- A **random variable** *represents the potential outcomes of a certain event*.
- A **probability distrubution** is an *assignation of a probability to a certain outcome*.

Idea 1.1.1
By incorporating the uncertainty explicitly in the model, we can measure the uncertainty asociated with the outputs.

For example, by giving a range to a forecast, which is a more realistic goal than giving just a single outcome.

Definition 1.1.2 (**Forecast**)
Forecast is a *prediction or estimate of future events, especially coming weather or a financial trend*.

Observation 1.1.2 (**Uncertainty in Business Processes**)
In business, by incorporating **uncertainty** is synonymus with understanding and quantifying the **risk** in a business process, and ideally leads to better management decisions.

A key idea is to understand that in some parts of a business process, we have uncertainty, so we must create a model that reflects that uncertainty/risk.

Example 1.1.1

A company has 10 drugs in a development portfolio.

- Given a drug has been aproved, you have to predict its revenue.

- But, whether a drug is approved or not is an uncertain future event (a random variable). We must estimate the probability of approval.

- We wish to invest in the company if the company's total expected revenue for the portfolio is over \$10B in 5 years time.

We need to calculate the **probability distrubution of the total revenue** to **understand the investment risk**.

1.2   EXAMPLES OF PROBABILISTIC MODELS

Some probabilistic models commonly used in business and finance include:

- **Regression models**.
- **Probability trees**.
- **Monte Carlo simulation**.
- **Markov models**. We look at stages of a certain process.

1.3   REGRESSION MODEL

Definition 1.3.1 (**Regression Model**)

A **regression model** is based on a set of data. We use the data to *make a reverse engineering of the data to find a relationship between the variables in order to capture a realistic description of the process*.

The following is an example of a regression model.

Example 1.3.1

Lets suppose we have data from several diamond shops in a certain street in NY. In the picture, we draw the dots corresponding to the price vs number of carats of a diamond sold in that street.

We can see that there is a pattern which is linear, this is, there is a linear relation between the price of a diamond and its weight in carats. We use linear regression in order to find the best fitting line (we may use the best fitting curve, etc...) that predicts the price of a diamond given its weight in carats.

$$P = -259.2 + 3721 \times C$$

where P is the price in dollars and C is the weight in carats. In the graph, we have a band that is a **prediction inverval** which *captures the range of uncertainty of the model*.

If we check the points in the graph, we see that not all of them lie in a line, but some of them are outside of it, these points are called noise in the system. We must incorporate them in our prediction interval and forecast.
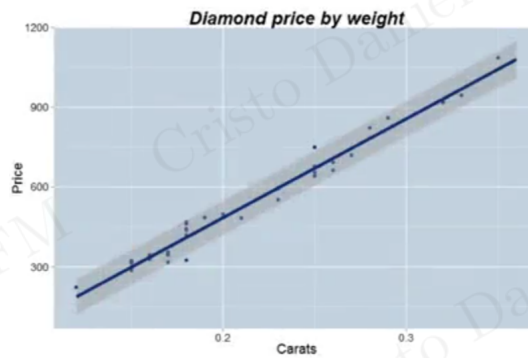
Figure 1.1: Linear Regression Model for Diamond Prices.

1.4 Probability Trees

> **Definition 1.4.1 (Probability Tree)**
> A **probability tree** (or **tree diagram**) is a *visual tool in math that maps out all possible outcomes of a series of events* (like coin flips or drawing cards) *and their likelihood*, using *branches for each outcome and labeling them with probabilities* (fractions/decimals).

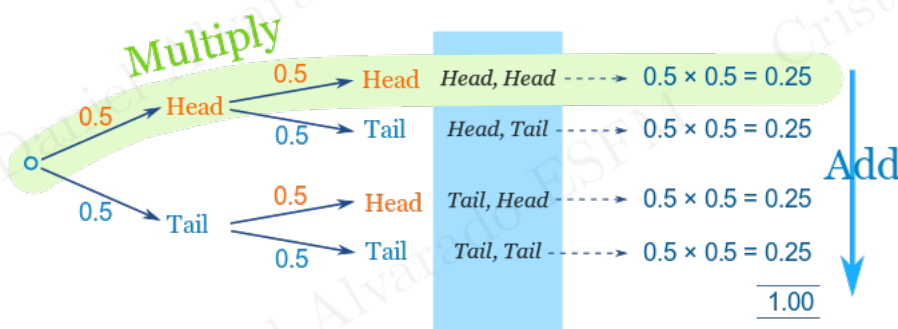Probability trees allow to propagate probabilities through a sequence of events.

Figure 1.2: Probability Three for the Event of Flipping a Coin Twice in a Row.

> **Example 1.4.1**
> Lets consider the following scenario. Imagine that we want to predict the probability of certain people of stop sharing copyright protected files online.
> A team made an estimation that:
>
> - When pirates get a notice from the goverment to stop infringing the law, 10% stop doing it.
> - If they continue infringing it and they get a second notice, 15% of them stop infringing the law.
> - If they continue infringing it and they get a third notice, 20% of them stop infringing the law.
>
> This can be visualized in the tree diagram in Figure 1.3.
> Lets suppose that we want to measure the probability of a pirate to stop infringing. We can calculate this by doing the multiplications in the diagram:
>
> $$P_{stop} = 0.1 + 0.9 \times 0.15 + 0.9 \times (0.85 \times 0.2) = 0.388$$
>
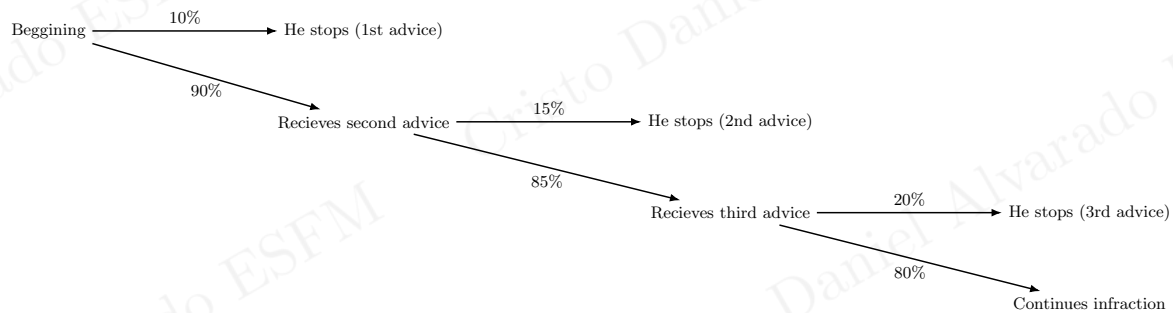> So, there is a 39% change for a pirate to stop infringing the law.

Figure 1.3: Probability Tree of the Example

1.5 MONTE CARLO SIMULATIONS

Monte Carlo simulations are very useful for modelating complicated scenarios. To give an example, lets go back to the demand model.

$$Q = 60,000P^{-2.5}$$

Here, Q is the quantity of the product and P is the price at which the product is sold.

This is a demand model, which describes, given the amount of a certain product, the price it must have.

> **Observation 1.5.1**
> In the latter chapter, we found out the optimal price given the cost of production of $2 per product is $3.33.

Here, we know the elasticity of demand, which is $b = -2.5$, but what if we dont know about it? In order to find this particular value, we can use Monte Carlo to find a suitable value of b.

> **Idea 1.5.1**
> The idea behind a Monte Carlo simulation is to generate a large number of random candidate values for b within a reasonable range (for example, from -3 to -2). For each candidate b_i, we can compute the corresponding prices and quantities predicted by the model and compare them with the observed market data (actual sales quantities at different prices).
>
> By defining a measure of error, such as the sum of squared differences between the predicted and observed quantities, we can evaluate how well each candidate b_i fits the data.

After repeating this process many times, the Monte Carlo simulation allows us to identify the value of b that minimizes the error. In this way, even without prior knowledge of the elasticity, we can estimate it empirically from the data using random sampling and statistical analysis.

> **Observation 1.5.2**
> The course gives the following idea. Lets suppose that b comes from a uniform distrubution between -3 to -2. In earlier sections we proved that:
>
> $$P_{opt} = \frac{cb}{1 + b}$$
>
> With this predicted price, we can find a distrubution for the optimal price of a product. In the course, the optimal price is between $3.1 and $3.7 (this is an interval that has 80% of all the sample results).