Classical Probability and Discrete Probability Modeling

CURM Background Material, Fall 2014

Dr. Doreen De Leon

1 Introduction

A probability model is used to mathematically model situations that involve some level of uncertainty. For example, analysis of golf and other sports, population studies, and analysis of wildfire management decisions, among others, are modeled using probability and statistics. One other application of probability modeling with which you are no doubt familiar is the overbooking of airline flights, based on the analysis of the risk of a no-show versus the compensation that must be paid to a bumped passenger.

2 Introduction to Classical Probability

A **probability model** is a mathematical representation of a random phenomenom. It is defined by three characteristics: a **sample space**, **events** within the sample space, and the **probabilities** corresponding to each event.

The **sample space** for a model is the set of all possible outcomes. An event A is a subset of the sample space. Thus, it is a collection of results or outcomes.

For example, suppose that we have a bag containing three marbles, one black, one white, and one red. If someone were to choose the marbles one at a time from the bag, without putting the chosen marble back in (known as without replacement), the three marbles can be pulled out in the following orders: {(black, white, red), (black, red, white), (white, black, red), (white, red, black), (red, black, white), (red, white, black)}. This represents the set of all possible outcomes, or the sample space, for this situation.

A **probability** is a numerical value assigned to an event A. The probability of an event is denoted $P(A)$, and describes the relative frequency of the event in the long run. The first two basic rules of probability are

1. A probability must satisfy the relation $0 \leq P(A) \leq 1$.
2. The probability of the sample space, S must be 1: $P(S) = 1$.

If there are n possible outcomes in a sample space, and each outcome is equally likely, then

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } S} = \frac{\text{number of outcomes in } A}{n}.$$

So, in our example with three marbles, if we ask the probability of pulling a black marble out first when pulling the marbles out of the bag one at a time without replacement, the answer would be

$$\begin{aligned} P(\text{pulling a black marble out first}) &= \frac{\text{number of outcomes in which a black marble is pulled first}}{\text{total number of outcomes}} \\ &= \frac{2}{6} = \frac{1}{3}. \end{aligned}$$

If two events, A and B, have no outcomes in common, then they are called **disjoint**. For example, the possible outcomes of pulling out a single marble are disjoint, because there is only one marble of each color. If two events A and B are disjoint, then

$$P(A \text{ or } B) = P(A) + P(B).$$

$P(A \text{ or } B)$ is typically written $P(A \cup B)$.

The complement of an event A, denoted A^c, is the set of all other events in the sample space. The probability of A^c is given by

$$P(A^c) = 1 - P(A).$$

This follows from the facts that A and A^c are disjoint, so $P(A \cup A^c) = P(A) + P(A^c)$ and $A \cup A^c = S$.

Example: Consider a die that is being rolled once. Let X be the outcome. What is $P(X < 5)$? $X < 5$ is the complement of $X \geq 5$, so

$$\begin{aligned} P(X < 5) &= 1 - P(X \geq 5) \\ &= 1 - P(X = 5 \text{ or } X = 6) \\ &= 1 - (P(X = 5) + P(X = 6)) \text{ (since } X = 5 \text{ and } X = 6 \text{ are mutually exclusive)} \\ &= 1 - \left(\frac{1}{6} + \frac{1}{6}\right) \\ &= \frac{2}{3}. \end{aligned}$$

Consider two events that can occur in succession, such as two flips of a coin. If the outcome of the first event has no effect on the probability of the outcome of the second event, then the two events are called **independent**. If two events A and B are independent, then

$$P(A \text{ and } B) = P(A) \cdot P(B).$$

Note that $P(A \text{ and } B)$ is often written $P(A \cap B)$.

If two events A and B are not disjoint, then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B), \text{ or } P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

For example, consider tossing two coins. The probability of getting a head on either toss is $\frac{1}{2}$. Since the two tosses are independent, the probability of getting a head on both tosses is equal to

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

The probability of a head on *either* toss, i.e., $P(H$ on the first toss or H on the second toss$)$ is

$$\frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

Note that disjoint events are **not** independent. For example, consider a bag consisting of four marbles, one black, one white, one blue, and one red. Define event A as drawing a blue marble from the bag. The probability of drawing a blue marble from the bag is $\frac{1}{5}$. Suppose event B is drawing a red marble out of the bag. The events are disjoint, since B can occur even if A occurs. However, the events are not independent. Why? The outcome of event A affects the outcome of event B, because, when the second marble is removed from the bag the bag only contains four marbles. If, instead, two marbles were to be removed from the bag, with the first marble replaced before the second marble was removed, then the event of drawing the first marble would not affect the outcome of drawing the second marble. Therefore, the two events are independent and the probability of both events occurring would be the product of the probabilities of each event, $\frac{1}{5} \cdot \frac{1}{5} = \frac{1}{25}$.

Two events A and B are **mutually exclusive** if the fact that A occurs means that B cannot occur, and vice versa. For example, if event A is pulling a blue marble out of the bag on the first draw and event B is pulling a blue marble out of the bag on the second draw, and the marbles are drawn without replacement, then the events are mutually exclusive. It makes sense, then, that $P(A \text{ and } B) = 0$.

Finally, we need to discuss **conditional probability**. The **conditional probability** of an event B is the probability that the event will occur given that event A has already occurred. This probability, $P(B$ given $A)$, is denoted $P(B|A)$. In the case where the two events are independent, then $P(B|A) = P(B)$. In general, if events A and B are not independent, then

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

Example 1: Suppose the probability that it will rain today is 60% and the probability that Sam will be wearing a hat if it is raining is 90%, the probability of Sam wearing a hat and it raining today is $P(\text{Sam wearing a hat and it's raining}) = P(\text{Sam wearing a hat}|\text{rain}) \cdot P(\text{rain}) = (0.90) \cdot (0.60) = 0.54$, or 54%.

Example 2: In a card game, suppose a player needs to draw two cards of the same suit to win. Of the 52 cards in a deck, there are 13 cards in each suit. Supopse first the player draws a heart. Then, in order to win, the player needs to draw another heart. Since one heart has already been drawn, there are now 12 hearts remaining in the deck of now 51 cards, so the probability of $P(\text{second draw a heart}|\text{first draw a heart}) = \frac{12}{51}$.

There is another idea, which is the **expected value**. Suppose that we have a random variable, which can take any of a discrete set of values

$$X \in \{x_1, x_2, \ldots, x_n\},$$

and suppose that $X = x_i$ occurs with probability p_i (i.e., $P(X = x_i) = p_i$). Then, $\sum p_i = 1$. Also, since X takes the value x_i with probability p_i, the average or expected value of X should be a weighted average of the possible values x_i, where the weight is the likelihood of x_i occuring, which is p_i. We write $E(X)$ to denote the expected value of X, and define

$$E(X) = \sum x_i p_i.$$

The probabilities p_i represent the probability distribution of X.

Example 3: Dice Game (adapted from *Mathematical Modeling*, 3rd edition, by Mark Meerschaert)

In a simple game of chance, two dice are rolled and the bank pays the player the number of dollars shown on the dice. The person running the game is charging $10 to play. Should you play the game?

Let X denote the number shown on the dice. There are $6 \times 6 = 36$ possible outcomes, and each is equally likely. There is only one way to roll a 2, so

$$P(X = 2) = \frac{1}{36}.$$

There are two ways to roll a 3 (1 and 2, or 2 and 1), so

$$P(X = 3) = \frac{2}{36} = \frac{1}{18}.$$

All of the possibilities appear in the table below.

i	2	3	4	5	6	7	8	9	10	11	12
$P(X = i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Note that the probabilities sum to 1. Now, we seek the expected value of X:

$$\begin{aligned} E(X) &= \sum_{i=1}^{11} x_i p_i \\ &= 2\left(\frac{1}{36}\right) + 3\left(\frac{2}{36}\right) + 4\left(\frac{3}{36}\right) + 5\left(\frac{4}{36}\right) + 6\left(\frac{5}{36}\right) + 7\left(\frac{6}{36}\right) \\ &\quad + 8\left(\frac{5}{36}\right) + 9\left(\frac{4}{36}\right) + 10\left(\frac{3}{36}\right) + 11\left(\frac{2}{36}\right) + 12\left(\frac{1}{36}\right) \\ &= 7. \end{aligned}$$

Therefore, you should not pay more than $7 to play this game.

There is a theorem called the "law of large numbers" which says that for any sequence of independent, identically distributed random variables $X_1, X_2, \ldots$, with $E(X)$ finite, the following holds true:

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to E(X)$$

as $n \to \infty$ with probability 1. In other words, if you play the game for a long time you are virtually certain to win about $7 per roll.

3 A Quality Control Example

This example was taken from *Mathematical Modeling*, 3rd edition, by Mark Meerschaert.

An electronics manufacturer produces a variety of diodes. Quality control engineers attempt to insure that faulty diodes will be detected in the factory before they are shipped. It is estimated that 0.3% of the diodes produced will be faulty. It is possible to test each diode individually. It is also possible to place a number of diodes in series and test the entire group. If this test fails, it means that one or more of the diodes in that group are faulty. The estimated testing cost is 5 cents for a single diode, and $4 + n$ cents for a group of $n > 1$ diodes. If a group test fails, then each diode in the group must be retested individually to find the bad one(s). Find the most cost-effective quality control procedure for detecting bad diodes.

We will use the five-step method to solve this problem.

Step 1: Identify the Problem.

Determine the size of a group of diodes that should be tested for faulty diodes to minimize the cost of testing.

Step 2: Identify Relevant Facts about the Problem

- It is estimated that 0.3% of the diodes produced are faulty.
- Each diode may be tested individually, costing 5 cents per diode.
- A group of n diodes may be tested simultaneously, costing $4 + n$ cents.
- If a group test fails, then all of the diodes in the group must be tested individually.

Step 3: Choose the Type of Modeling Method

We will solve this as an optimization problem, using probability.

Step 4: Make Simplifying Assumptions.

- Variables
  - n = number of diodes per test group
  - C = testing cost for one group (cents)
  - A = average testing cost (cents/diode)
- Assumptions
  - If $n = 1$, then $A = 5$ cents.
  - If $n > 1$, then
    - $C = 4 + n$ if the group test indicates that all diodes are good, or
    - $C = (4 + n) + 5n$ if the group test indicates a failure.

– If $n > 1$, then $A = \dfrac{\text{average value of } C}{n}$.

Step 5: Construct the Model.

The random variable C takes on one of two possible values for any fixed $n > 1$. If all of the diodes are good, then $C = 4 + n$. Otherwise, $C = (4 + n) + 5n$, since we have to retest each diode. Let p denote the probability that all of the diodes are good. Then, the probability that one or more diodes is bad is $1 - p$. Thus, the average or expected value of C is

$$E(C) = (4 + n)p + [(4 + n) + 5n](1 - p).$$

Step 6: Solve and Interpret the Model.

There are n diodes, and the probability that one individual diode is bad is 0.003 (so, the probability that one individual diode is good is 0.997). Assuming independence, the probability that in a group of n diodes, all diodes are good is $p = (0.997)^n$. The expected value of the random variable C is

$$\begin{aligned} E(C) &= (4 + n)(0.997)^n + [(4 + n) + 5n](1 - (0.997)^n) \\ &= (4 + n) + 5n(1 - (0.997)^n) \\ &= 4 + 6n - 5n(0.997)^n. \end{aligned}$$

Therefore, the average testing cost per diode is

$$A = \frac{4}{n} + 6 - 5(0.997)^n.$$

The strong law of large numbers tells us that this formula represents the long-run average cost we will incur if we test groups of size n. Now, we need to minimize A as a function of n.

Let

$$A(x) = \frac{4}{x} + 6 - 5(0.997)^x$$

denote the average cost function as a function of x. Then, we need to solve $A'(x) = 0$ over the set $x > 0$.

First,

$$A'(x) = -\frac{4}{x^2} - 5\ln(0.997)(0.997)^x$$

and

$$A''(x) = \frac{8}{x^3} - 10\ln(0.997)(0.997)^x.$$

For $x > 0$, $A''(x) > 0$. Therefore, solving $A'(x) = 0$ does, in fact, give us the minimum value of A.

Using Newton's method to solve $A'(x) = 0$ gives (see Maple worksheet) $x = 17$. So, the minimum of $A = 1.48$ occurs at $n = 17$.

Therefore, the manufacturer should test groups of 17 diodes, for an average cost of 1.48 cents per diode

Step 7: Validate the Model.

Quality control procedures for detecting faulty diodes can be made much more economical by group testing methods. Individual testing costs approximately 5 cents per unit, and bad diodes occur only rarely, at a rate of 3 per 1,000. By testing groups of 17 diodes each, in series, we can reduce testing costs by a factor of 3 (to approximately 1.5 cents per unit) without sacrificing quality.

We will do a sensitivity analysis here. Note that the implementation of a quality control procedure will depend on several factors outside the scope of our model. For example, it may be easier to test diodes in batches of 10 or 20, or perhaps the testing group should be a multiple of 4 or 5, depending on the details of the manufacturing process. The average cost A does not vary significantly between $n = 10$ and $n = 35$, so this is not a big issue. The parameter $q = 0.003$, which represents the failure rate in the manufacturing process, must also be considered. For example, this value may vary with the environmental conditions inside the plant. Generalizing the original model gives

$$A = \frac{4}{n} + 6 - 5(1-q)^n.$$

At $n = 17$, we have that

$$S(A, q) = \frac{dA}{dq} \cdot \frac{q}{A} = 0.16,$$

so small variations in q are not likely to affect the cost very much.

A more general robustness analysis would consider the assumption of independence. We have assumed that there is no correlation between the times of successive failures in the manufacturing process. It may be, in fact, that bad diodes tend to be produced in batches, perhaps due to a passing anomaly in the manufacturing environment, such as a power surge or vibration. The mathematical analysis of dependent random variable models cannot be dealt with completely at this time. Stochastic process models can represent some types of dependence, but other types of dependence do not have tractable analytic formulations. In practice, simulation results tend to indicate that expected value models based on independent random variables are quite robust. More importantly, experience has shown that such models provide useful, accurate approximations of real-life behavior in many cases.