Discrete Random Variables

Instructor: Shandian Zhe

Janurary 28, 2025

Random Variables

A **random variable** is a function from a sample space to real numbers. The mathematical notation for a random variable X on a sample space Ω is:

$$X : \Omega \to \mathbb{R}$$

Example: Sum of dice

- Sample space: $\Omega = \{(i, j) : i, j \in \{1, \ldots, 6\}\}$
- Random variable: $S(i, j) = i + j$

Why Introduce Random Variables

- Raw sample spaces vary much across scenarios; random variables provide a way to look into all kinds of events in a unified space (e.g., real space)
- Allow us to find out common properties and rules from various random events of different nature.
- A random variable can reveal some feature of the sample space that may be more interesting than the raw sample space outcomes.

Defining Events Using Random Variables

The notation $\{X = a\}$ defines the event of all elements in our sample space for which the random variable X evaluates to a. In set notation:

$$\{X = a\} = \{\omega \in \Omega : X(\omega) = a\}$$

The probability of this event is denoted $P(X = a)$.

Example: Sum of dice

- What is $\{S = 5\}$? What is $P(S = 5)$?
- How about $\{S = 7\}$?

In-class Exercise

For the two-dice experiment, define the random variable:

$$X(i,j) = i \times j$$

- For $a = 3, 4, 12, 14$, what are the events $\{X = a\}$?
- What are the probabilities $P(X = a)$?

Probability Mass Function (PMF)

The **probability mass function (PMF)** for a random variable X is defined as:

$$f(a) = P(X = a)$$

- This function is zero for values of a that are not possible outcomes.
- Sometimes, a pmf is also called a **probability density function (PDF)** or just a **density**.

Cumulative Distribution Function (CDF)

The **cumulative distribution function (cdf)** for a random variable X is defined as:

$$F(a) = P(X \leq a)$$

- CDF is the probability of a particular type of events.

CDF example

- One-dice experiment: $X(i) = i$ $(i = 1, \ldots, 6)$. What is PMF and CDF?
- Two-dice experiment: $X(i, j) = i + j$ $(i, j = 1, \ldots, 6)$. What is PDF and CDF?

Bernoulli Distribution

Defined by the following PMF:

$$f_X(1) = p, \quad \text{and} \quad f_X(0) = 1 - p$$

- p is a single number between 0 and 1 (not a probability function).
- If X is a random variable with this PMF (or PDF), we say X follows a Bernoulli distribution, or X is a Bounoulli random variable with parameter p. This can be denoted as $X \sim \text{Ber}(p)$.

Example: A Bernoulli trial is like flipping a coin with p as the probability of success (heads); in other words, the chance of failure (tails) is $1 - p$.

Binomial Distribution

- The binomial distribution describes the probabilities for repeated Bernoulli trials – such as flipping a coin 10 times in a row.
- Each trial is assumed to be independent of the others (e.g., flipping a coin once does not affect any of the outcomes for future flips).
- It is characterized by two parameters n and p, where n is the number of trials, and p is the probability of success in each trial.
- Random variable X is defined as the total number of successes from the n trials.

Binomial Distribution

- We say X a binomial random variable with parameters n and p or use notation $X \sim \text{Bin}(n, p)$.
- How to compute the PMF?

Binomial Distribution

- Consider a special case $n = 5$ and $p = 0.5$.
- Let us consider $p(X = 1)$
- How to represent the event $X = 1$?

Binomial Distribution

For each sequence in the event, what is the probability?

$$
\begin{aligned}
p((H,T,T,T,T)) &= p^1(1-p)^{5-1} \\
p((T,H,T,T,T)) &= p^1(1-p)^{5-1} \\
&\dots \\
p((T,T,T,T,H)) &= p^1(1-p)^{5-1}
\end{aligned}
$$

Adding them together, we have

$$p(X = 1) = 5p^1(1-p)^4$$

Binomial Distribution

Let us consider $X = 2$. The English language is we flip the coin for 5 times, we observe 2 heads.

- How to represent the event $X = 2$?

Binomial Distribution

Let us consider $X = 2$. The English language is we flip the coin for 5 times, we observe 2 heads.

- How to represent the event $X = 2$?

Binomial Distribution

For each sequence in the event, what is the probability?

$$p((H, H, T, T, T)) = p^2(1-p)^{5-2}$$
$$p((H, T, H, T, T)) = p^2(1-p)^{5-2}$$
$$\ldots$$

How many sequences are in the event? *5 choose 2*

Binomial Distribution

In general, how to compute *n choose k*?

Binomial Distribution

Remember the definition for factorial:

$$n! = n \times (n-1) \times \cdots \times 2 \times 1$$

This is the number of ways to put n objects into distinct orders.

Binomial Distribution

The definition for "n choose k":

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

This is the number of ways to select k objects out of n objectives, where the order of the selected objects does not matter

Binomial Distribution

The **binomial distribution** describes probabilities for repeated Bernoulli trials (independent trials). The PMF is given by:

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $\binom{n}{k} = \frac{n!}{(n-k)!k!}$.

- n: Number of trials
- k: Number of successes
- p: Probability of success

Geometric Distribution

The **geometric distribution** gives the probability that the first $k - 1$ trials are failures, and the kth trial is the first success. Its pmf is:

$$f_X(k) = (1 - p)^{k-1} p$$

- Denoted $X \sim \text{Geo}(p)$.
- Example: What is the probability of losing the first 3 times and winning on the 4th try?

In-class Problem: Monty Hall

Monty Hall problem: If we switch doors, we have a $\frac{2}{3}$ chance of winning and $\frac{1}{3}$ chance of losing.

- If we play the game 4 times, what is the probability of winning exactly once?
- How about exactly 0, 2, 3, or 4 times?
- What is the probability of losing the first three times and winning on the 4th try?

Key to Variable Names

- n: Number of trials
- k: Number of successes (Binomial) or first success (Geometric)
- p: Probability of success