Notes: Discrete Random Variables

CS 3130/ECE 3530: Probability and Statistics for Engineers

Jan 25, 2024

Random Variables:
A **dsicrete random variable** (RV) is a function from a sample space to the real numbers. The mathematical notation for a random variable X on a sample space Ω looks like this:

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

A random variable defines some *feature* of the sample space that may be more interesting than the raw sample space outcomes.

There can be a finite number of discrete RVs like this: $a_1, a_2, \ldots, a_n$
Or there can be an (countably) infinite number, like this: $a_1, a_2, \ldots$
Example: Sum of dice
Sample space: $\Omega = \{(i, j) : i, j \in \{1, \ldots, 6\}\}$, Random variable: $S(i, j) = i + j$

We can define 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 for $\{S = 7\}$?

In-class Exercise: Also for the two dice experiment, define the random variable $X(i, j) = i \times j$, i.e., X is the product of the two dice values. For $a = 3, 4, 12, 14$, what are the events $\{X = a\}$ and the probabilities $P(X = a)$?

Probability mass function:
The **probability mass function (pmf)** for a random variable X is a function $f : \mathbb{R} \to [0, 1]$ defined by

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

Notice this function is zero for values of a that are not possible outcomes.
Sometimes denote $f_X(a)$
DRAW DIAGRAM

Cumulative distribution function:
The **cumulative distribution function (cdf)** for a random variable X is a function $F : \mathbb{R} \to [0, 1]$ defined by $F(a) = P(X \leq a)$.
DRAW DIAGRAM
Properties to discuss:

- $f(a) \geq 0$, for $-\infty \leq$ a $\leq \infty$, $p(a_1) + p(a_2) + \ldots = 1$
- $a \leq b \Leftrightarrow F(a) \leq F(b)$
- Limits of $F(a)$ at infty and -infty
- Right continuous.

Discuss how to derive a PMF form a CDF (and vise versa).
Bernoulli distribution:
Defined by the following pmf:

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

Don't let the p confuse you, it is a single number between 0 and 1, not a probability function. If X is a random variable with this pdf, we say "X is a Bernoulli random variable with parameter p", or we use the notation $X \sim Ber(p)$. You can think of a Bernoulli trial as flipping a coin where the chance of heads is p and the chance of tails is $1 - p$. Often we call 0 a "failure" and 1 a "success", so p is the probability of success.

Example. Two soccer teams compete in a match and one team is better, with a probablity 0.65 of winning. What is the Bernoulli distribution for this scenario?
Binomial distribution:
The binomial distribution describes the probabilities for repeated Bernoulli trials – such as flipping a coin ten times in a row. Each trial is assumed to be independent of the others (for example, flipping a coin once does not affect any of the outcomes for future flips). First, we need some definitions.

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.

And 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 a possible n, where the order does *not* matter.

The **binomial distribution** with parameters n and p is given by the pmf:

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

This is denoted $X \sim Bin(n, p)$. This distribution is for repeated Bernoulli trials, and it gives the probability that you get k successes out of n trials.

Examples:

- Two teams playing 5 games (without stopping early).
- Student taking a test with random answers.
- Two teams playing a best of 7 series, what is the probability that they need to play the 7th game.

Geometric distribution:
The **geometric distribution** is also for repeated Bernoulli trials, and it gives the probability that the first $k-1$ trials are failures, while the kth trial is the first success. Its pmf is

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

This is denoted $X \sim Geo(p)$.

In-class Problem: Remember the Monty Hall problem – if we switch doors, we have a 2/3 chance of winning and 1/3 chance to lose. If we play the game 4 times, what is the probability that we win exactly once? How about exactly 0, 2, 3, or 4 times? What is the chance that we loose the first three times and finally win on the 4th try?

Key to variable names

It's important to keep straight what all the variables mean in the above equations. Here is a summary:

n : Number of trials
k : Number of successes in Binomial, OR first success that occurs in Geometric
p : Probability of success