Discrete Random Variables[1]
(Sections 2.1-2.3 and parts of 2.5)

STA 256: Fall 2019

[1] This slide show is an open-source document. See last slide for copyright information.

Overview

Random Variable: The idea

The idea of a random variable is a *measurement* conducted on the elements of the sample space.

- S could be the set of Canadian households, all equally likely to be sampled. $X(s)$ is the number of people in household s.
- Toss a coin with $P(\text{Head}) = p$, three times. $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. $X(s)$ is the number of Heads for outcome s.
- $X(s)$ could be one if s is employed, and zero if s is unemployed.

Formal Definition of a random variable

A random variable is a function from S to the set of real numbers.

- This is consistent with the idea of measurement.
- It takes an element s, and assigns a numerical value to it.
- This is why we were writing $X(s)$.
- Often, a random variable is denoted by X,
- But it's really the function $X(s)$.

Probability statements about a random variable

The probability that $X(s)$ will take on various numerical values is *determined* by the probability measure on the subsets of S.

$$P(X = 2) = P\{s \in S : X(s) = 2\}$$

$$P(X = x) = P\{s \in S : X(s) = x\}$$

There is a critical difference between capital X and little x.

$$P(X \in B) = P\{s \in S : X(s) \in B\}$$

Example

Toss a fair coin twice.

- $P\{HH\} = P\{HT\} = P\{TH\} = P\{TT\} = \frac{1}{4}$.
- Let X equal the number of heads.
- $P(X = 0) = P\{TT\} = \frac{1}{4}$.
- $P(X = 1) = P\{HT, TH\} = \frac{1}{2}$.
- $P(X = 2) = P\{HH\} = \frac{1}{4}$.

Distribution of a random variable

Leaves out some technicalities

The *distribution* of a random variable X is the collection of probabilities $P(X \in B)$ for all $B \subseteq \mathbb{R}$.

Discrete Random Variables

The random variable X is said to be *discrete* if there exist distinct $x_1, x_2, \ldots$ (perhaps only finitely many) with

- $P(X = x_j) > 0$ for all j, and
- $\sum_{j=0}^{\infty} P(X = x_j) = 1$.

A better definition (for some people): The random variable $X : S \to \mathbb{R}$ is said to be discrete if its range is countable.

Not all random variables are discrete

The random variable X is said to be *discrete* if there exist distinct $x_1, x_2, \ldots$ (perhaps only finitely many) with

- $P(X = x_j) > 0$ for all j, and
- $\sum_{j=0}^{\infty} P(X = x_j) = 1$.

Let the random variable X take values in $(0, 1)$, with $P(X \in B) =$ length of $B \cap (0, 1)$. Because $P(X = x) = 0$ for all real x, the random variable X is not discrete.

Probability Function of a discrete random variable

Also called the **probability mass function**

Suppose the random variable X takes on the values $x_1, \ldots, x_n$ or $x_1, x_2, \ldots$ with non-zero probability. The *probability function* of X is written

$$p_X(x) = P(X = x)$$

for all real x.

For the 2 fair coins example, $p_X(0) = \frac{1}{4}$, $p_X(1) = \frac{1}{2}$ and $p_X(2) = \frac{1}{4}$.

$p_X(14) = 0$.

Cumulative Distribution Function

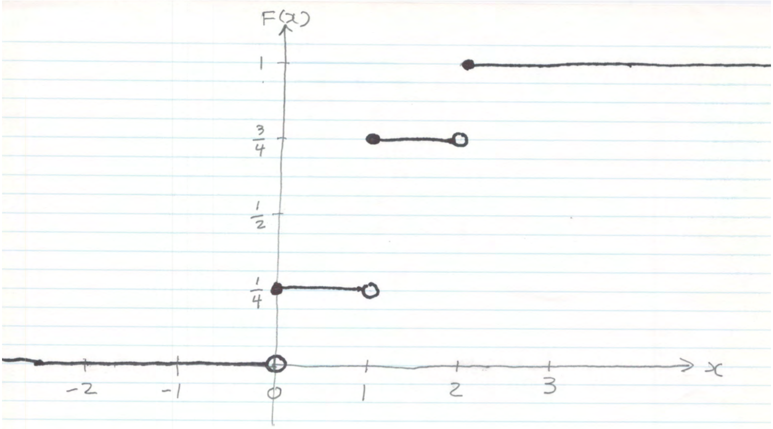
The *cumulative distribution function* of a random variable X is defined by

$$F_X(x) = P(X \leq x)$$

- Note that X is the random variable, and x is a particular numerical value.
- $F(x)$ is defined for all real x.
- $F(x)$ is non-decreasing. This is because
- If $x_1 < x_2$, $\{s : X(s) \leq x_1\} \subseteq \{s : X(s) \leq x_2\}$.
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Cumulative distribution function for the coin toss example

Fig. 2.5.1 on page 65 gets only part marks. CDFs are right continuous.



Common Discrete Distributions

- The Degenerate distribution
- The Bernoulli distribution
- The Binomial distribution
- The Geometric distribution
- The Negative Binomial distribution
- The Poisson distribution
- The Hypergeometric distribution

Degenerate distribution

Example 2.3.1

$$p_X(x) = \begin{cases} 1 & \text{for } x = c \\ 0 & \text{for } x \neq c \end{cases}$$

The Bernoulli Distribution

Example 2.3.2

- Simple probability model: Toss a coin with $P(\text{Head}) = \theta$, one time. Let X equal the number of heads.
- Probability (mass) function of X:

$$p_X(x) = \begin{cases} \theta^x(1-\theta)^{1-x} & \text{for } x = 0 \text{ or } 1 \\ 0 & \text{Otherwise} \end{cases}$$

- An *indicator random variable* equals one if some event happens, and zero if it does not happen.
  - 1=Female, 0=Male
  - 1=Lived, 0=Died
  - 1=Passed, 0=Failed
- Indicators are usually assumed to have a Bernoulli distribution.

The Binomial Distribution

Example 2.3.3

- Simple probability model: Toss a coin with $P(\text{Head}) = \theta$. Toss it n times. Let X equal the number of heads.
- Probability (mass) function of X:

$$p_X(x) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x} & \text{for } x = 0, 1, \ldots, n \\ 0 & \text{Otherwise} \end{cases}$$

- The Bernoulli is a special case of the Binomial, with $n = 1$.

Why does $p_X(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$

For the Binomial Distribution?

Toss a coin n times with $P(\text{Head}) = \theta$, and let X equal the number of heads. Why does $P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$?

- The sample space is the set of all strings of n letters composed of H and T.
- By the Multiplication Principle, there are 2^n elements.
- If two different strings have x heads (and $n - x$ tails), they have the same probability.
- For example, $P\{HHTH\} = P\{THHH\} = \theta^3(1-\theta)$ by independence.
- Count the number of ways that x positions out of n can be chosen to have the symbol H.
- n choose x is $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.
- So $P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ ■

Geometric Distribution

Example 2.3.4

- Simple probability model: Toss a coin with $P(\text{Head}) = \theta$ until the first head appears, and then stop. Let X equal the number of times the coin comes up tails, *before* the head occurs.
- Probability (mass) function of X:

$$p_X(x) = \begin{cases} (1-\theta)^x \theta & \text{for } x = 0, 1, \dots \\ 0 & \text{Otherwise} \end{cases}$$

Negative Binomial Distribution

Example 2.3.5

- Simple probability model: Toss a coin with $P(\text{Head}) = \theta$ until r heads appear, and then stop. Let X equal the number of tails before observing the rth head.
- First we observe $r-1$ heads and x tails, in no particular order.
- Then we observe another head.
- For $x = 0, 1, \ldots$, the probability (mass) function is

$$
\begin{aligned}
p_X(x) &= \binom{x+r-1}{x} \theta^{r-1}(1-\theta)^x \theta \\
&= \binom{x+r-1}{x} \theta^{r}(1-\theta)^x
\end{aligned}
$$

The Geometric distribution is a special case of the negative binomial, with $r = 1$.

Poisson distribution

Example 2.3.6

Useful for count data. For example,

- Number of rasins in a loaf of rasin bread.
- Number of alpha particles emitted from a radioactive substance in a given time interval.
- Number of calls per minute coming in to a customer service line.
- Bomb craters in London during WWII.
- Number of rat hairs in a jar of peanut butter.
- Number of deaths per year from horse kicks in the Prussian army, 1878-1898.

Conditions for the Poisson distribution

We are usually counting events that happen in an interval, or in a region of time or space (or both).
The following are rough translations for the technical conditions for the number of events to have a Poisson distribution.

- Independent increments: The occurrence of events in separate intervals (regions) are independent.
- The probability of observing at least one event in an interval or region is roughly proportional to the size of the interval or region.
- As the size of the region or interval approaches zero, the probability of more than one event in the region or interval goes to zero.

If these conditions are approximately satisfied, the probability distribution of the number of events will be approximately Poisson.

Poisson Probability Function, with parameter $\lambda > 0$

$$p_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \text{for } x = 0, 1, \ldots \\ 0 & \text{Otherwise} \end{cases}$$

Where the parameter $\lambda > 0$.

Note $\sum_{x=0}^{\infty} p_X(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$

Hypergeometric Distribution
Example 2.3.7

- Simple probability model: Jar with N balls, of which M are white and $N - M$ are black. Randomly sample $n \leq N$ balls without replacement. Let X denote the number of white balls in the sample.
- Probability function of X:

$$p_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

- But for what values of x is this correct?

$p_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$ For some values of x. For all other values, $p_X(x) = 0$

Jar with N balls, M white and $N - M$ black. Sample $n \leq N$ balls. X = number of white balls selected.

- Definitely $0 \leq x \leq n$.
- Look at the binomial coefficients. You cannot sample more objects than you have.
- So $x \leq M$. And we have $x \leq \min(n, M)$.
- Also $n - x \leq N - M \Leftrightarrow x \geq n - (N - M)$
- This last restriction makes sense. Suppose the size of the sample is greater than the number of black balls (possible). Like there are 7 white and 3 black, and you choose 5 balls. You will get at least 2 white balls. $x \geq n - (N - M)$
- $x \geq 0$ and $x \geq n - (N - M)$, so $x \geq \max[n - (N - M), 0]$.

Putting it all together

The Hypergeometric Distribution

Jar with N balls, M white and $N-M$ black. Sample $n \leq N$ balls. X = number of white balls selected.

$$p_X(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \text{for } x = \max[n-(N-M),0],\ldots,\min[n,M] \\ 0 & \text{Otherwise} \end{cases}$$

The set of values where a random variable has positive probability is called its *support*.

The big Three

The most useful discrete distributions in applications are

- Bernoulli
- Binomial
- Poisson

Copyright Information