Discrete Random Variables[1]

STA 256: Fall 2018

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

Random Variable: The idea

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

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

Formal Definition of a random variable

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

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

Probability statements about a random variable

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

$$P(X = 2) = P\{\omega \in \Omega : X(\omega) = 2\}$$

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

Example

Toss a fair coin twice.

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

Probability Mass Function

Also called the **frequency function**

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

$$p(x_i) = P(X = x_i)$$

Note $\sum_i p(x_i) = 1$

For the 2 fair coins example, $p(0) = \frac{1}{4}$, $p(1) = \frac{1}{2}$ and $p(2) = \frac{1}{4}$. $p(14) = 0$.

Cumulative Distribution Function

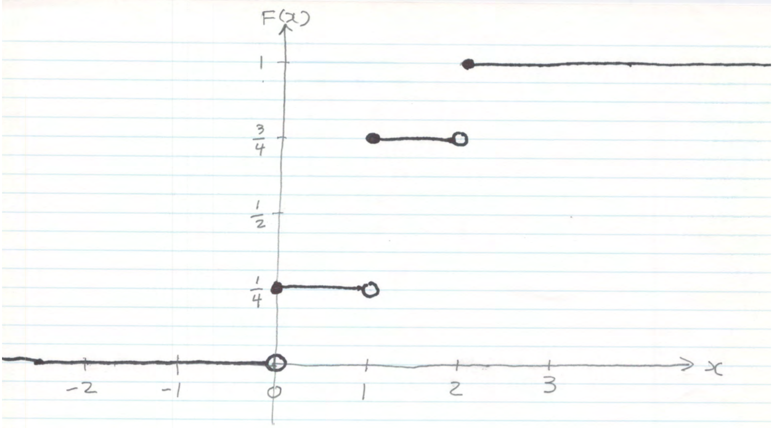
The *cumulative distribution function* of a random variable X is defined by

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

- Note that X is the random variable, and x is a particular numerical value.
- You will frequently see things like $P(X = x)$. There is a critical difference between capital X and little x.
- $F(x)$ is defined for all real x.
- $F(x)$ is non-decreasing. This is because
- If $x_1 < x_2$, $\{\omega : X(\omega) \leq x_1\} \subseteq \{\omega : X(\omega) \leq x_2\}$.
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Cumulative distribution function for the coin toss example

Fig. 2.2 on page 37 is incorrect. CDFs are right continuous.



The Bernoulli Distribution

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

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

- 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

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

$$p(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, \ldots, n \\ 0 & \text{Otherwise} \end{cases}$$

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

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

For the Binomial Distribution?

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

- 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 strings have k heads (and $n-k$ tails), they have the same probability.
- For example, $P\{HHTH\} = P\{THHH\} = p^3(1-p)$ by independence.
- Count the number of ways that k positions out of n can be chosen to have the symbol H.
- n choose k is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
- So $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ ■

Geometric Distribution

- Simple probability model: Toss a coin with $P(\text{Head}) = p$ until the first head appears, and then stop. Let X equal the number of times the coin is tossed.
- Probability mass function (frequency function) of X:

$$p(k) = P(X = k) = \begin{cases} (1-p)^{k-1}\,p & \text{for } k = 0, 1, \ldots \\ 0 & \text{Otherwise} \end{cases}$$

Negative Binomial Distribution

- Simple probability model: Toss a coin with $P(\text{Head}) = p$ until r heads appear, and then stop. Let X equal the number of times the coin is tossed.
- Probability mass function (frequency function) of X:

$$p(k) = P(X = k) = \begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r} & \text{for } k = r, r+1, \dots \\ 0 & \text{Otherwise} \end{cases}$$

- In the first $k-1$ trials there are $r-1$ successes. So the frequency function could be written

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

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

Hypergeometric Distribution

- Simple probability model: Jar with n marbles, of which r are black and $n-r$ are white. Randomly sample m without replacement. Let X denote the number of black balls in the sample.
- Probability mass function (frequency function) of X:

$$p(k) = P(X = k) = \begin{cases} \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} & \text{for } k = 0, \ldots r \\ 0 & \text{Otherwise} \end{cases}$$

- This just summarizes what we have done earlier.

Poisson distribution

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 Mass Function with parameter λ

Frequency Function

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

Where the parameter $\lambda > 0$.

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

The big Three

The most useful discrete distributions in applications are

- Bernoulli
- Binomial
- Poisson

Copyright Information