STAT 270 - Chapter 4
Discrete Distributions

June 6, 2012

Distributions

- Discrete (chapter 4)
- Continuous (chapter 5)
- Other types that are not considered in this course

Random Variables, rv, RV

- A function of the sample space
- Denoted by capital letters X, Y, Z, etc.
- Don't confuse them with algebraic variables, e.g. $ax^2 + bx + c = 0$.

Example. 3 coin flips, X: number of heads.

Range of X

- Range of X: All the possible values that X can take.
- Range of X discrete $\rightarrow$ X discrete
- Bernoulli rv: range$= \{0, 1\}$. e.g.

Example. Maximum temprature in degrees Celisius in Vancouver on Feb. 15 2006.
sample space:

We can define a discrete rv on a continuous sample space:

Definition

Probability mass function (pmf) of a discrete rv X,

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

also denoted by $p(x)$.

- Interpretation: probability that $X = x$.
- p_X is induced by the original probability measure P defined on S, i.e., $p_X(x)$ is the probability of all s in S for which $X(s) = x$.

Example. (continued) 3 coin flips.

pmf

$p(x)$ satisfies,

1. $p(x) \geq 0$, for any x.
2. $\sum_x p(x) = 1$ where the sum is taken over the range of X.

- $p(x)$ is pmf **iff** $p(x)$ satisfies (1) and (2)
- We can use this definition for $p(x)$ and forget about the first definition!

Example. Roll two dice. X: sum of the two numbers.

Example

Consider a batter in baseball who gets a hit with probability 0.3. Let X be the number of "at bats" until getting a hit. Obtain $p_X(x)$.

Useful to know: when $|r| < 1$,

$$\sum_{x=0}^{\infty} r^x = \frac{1}{1-r}$$

$$\sum_{x=0}^{n} r^x = \frac{1-r^{n+1}}{1-r}$$

Definition

Cumulative distribution function (cdf) of a discrete rv X is given by,

$$F_X(x) = P(X \leq x) = \sum_{y \leq x} p_X(y)$$

i.e., the probability of realizations of X up to and including x.

Example. (continued) 3 coin flips. Obtain pmf and cdf of X (the number of heads).

Properties of $F(x)$

- $F(x)$ is normed, i.e.,

$$\lim_{x \to -\infty} F(x) = 0$$

$$\lim_{x \to \infty} F(x) = 1$$

- $F(x)$ is non-decreasing
- $F(x)$ is right continuous, i.e.,

$$\lim_{x \to x_0^+} F(x) = x_0$$

Examples

Example. (continued) for the baseball batter example obtain the cdf of X.

Example. At the end of an exam 4 books are left behind. At the begining of the next lecture the 4 books are randomly returned to the students who left their textbooks. Let X be the number of students who receive their own books. Obtain the pmf and cdf of X.

Examples

Example. A library subscribes to two weekly magazines that both are supposed to arrive on Wednesdays. But they actually arrive independently with probabilities $P(\text{Wed}) = 0.3$, $P(\text{Thu}) = 0.4$, $P(\text{Fri})$, $P(\text{Sat}) = 0.1$. Let Y be the number of days beyond Wednesday that it takes for both magazines to arrive. Obtain pmf and cdf of Y.

Expectation, μ_X, μ

One of the useful characteristics of a rv.
Expected value or **expectation** of a discrete rv X is

$$E(X) = \sum_x x p_X(x)$$

where the sum is over the range of X.

Interpretation of $E(X)$

Using the frequency definition of probability, average realized value of X in the long run is $E(X)$.

Repeat the experiment N times, if $x_1, x_2, \ldots$ are the values that X can take we expect to observe x_i approximately $Np(x_i)$,

$$E(X) = \frac{x_1 Np(x_1) + x_2 Np(x_2) + \ldots}{N} = x_1 p(x_1) + x_2 p(x_2) + \ldots$$

Also called **mean** or **population mean** of X.

Example

3 coin flips, X: number of heads. $E(X) =?$

μ doesn't have to be in the range of X. Interpretation: If you flip 3 coins many times the average number of head in a triple flip is 1.5.

$E(g(X))$

Let g be a function, then $g(X)$ is a rv and

$$E(g(X)) = \sum_x g(x) p_X(x)$$

Example. Roll a die, let X be the value that is observed. $E(X^2) = ?$

Interpretation:

Proposition $E(aX + b) = aE(X) + b$
proof.

Example

A store orders copies of a weekly magazine for its magazine rack. Let X be the weekly demand for the magazine with pmf

x	1	2	3	4	5	6
p(x)	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{4}{15}$	$\frac{3}{15}$	$\frac{2}{15}$

Suppose that the store owner pays \$1 for each copy of the magazine and the customer price is \$2. If leftover magazines at the end of the week have no salvage value, is it better for the owner to order three magazines or four magazines?

Proposition

For a discrete rv X with pmf $p(x)$

$$E[\sum_{i=1}^{k} g_i(X)] = \sum_{i=1}^{k} E[g_i(X)]$$

proof:

Variance

Let $g(X) = [X - E(X)]^2$

$$var(X) = E(g(X)) = E[X - E(X)]^2$$

denoted by σ_X^2 or σ^2.
Describes spread or dispersion of the rv X.
More convenient formula:

$$var(X) = E(X^2) - [E(X)]^2$$

standard deviation of X,

$$\sigma_X = \sqrt{var(X)}$$

Proposition

$$var(aX + b) = a^2 var(X)$$

proof:

Intuition:

sample quantities $(\bar{x}, s^2)$ describe a finite data set
population quantities (μ, σ^2) describe a hypothetical population

Example

Consider a rv X with pmf

x	4	8	10
p(x)	0.2	0.7	0.1

Compute $E(X)$, $E(X^2)$, σ, and $E(3X + 4X^2)$.

Example

Let X be the average temprature in January in degrees Celcius where $E(X) = 5C$ and $var(X) = 3C^2$. Let Y be the average temprature in January in degrees Farenheit compute $E(Y)$ and $var(Y)$.

Example 4.11.

Consider a game where you bet x dollars and you win y dollars with probability p. What is the value of x to have a fair game?
(losing x dollars is the same as winning $-x$ dollars)

Distribution

- Not easy to define
- Probabilities assiciated with a rv
- pmf determines the distribution

The Binomial Distribution

- Discrete distribution

Definition: $X \sim Binomial(n, p)$ if the pmf of X is given by

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer, and $0 < p < 1$.
Satisfies the properties of pmf:

- A **family** of distributions indexed by **parameters** n and p.
- Fixed n and $p \rightarrow$ a member of the family.
- Bernoulli distribution is a special case arising from $n = 1$.

Motivation

In n independent trials with $P(\text{success}) = p$ what is $P(\text{obtaining } x \text{ successes})$?

S: success, F: Failure

Consider the sequence of x successes and $n - x$ failures

$$P(SS \dots SFF \dots F) = p^x (1 - p)^{n-x}$$

number of sequences: $\binom{n}{x}$

Example

Roll two dice four times. What is the probability that x pairs are obtained?

If $X \sim Binomial(n, p)$ then

$$E(X) = np$$
$$var(x) = np(1-p)$$

Proof:

When is the Binomial distribution appropriate for modelling?

- n trials, each resulting in either success ot failure.
- Trials are independent.
- Trials have the same probability of succes p.

Example

Suppose 35% of the employees of a company are statisticians. If the employees leave their offices independently at the end of the day, of the next 10 employees who walk out of the building, how many people do you expect to be statisticians?

Example

A typesetter, on average, makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be more than 50 errors in 5 pages?

The Poisson distribution

- The second most applicable distribution

Definition: $X \sim \mathit{Poisson}(\lambda)$ if the pmf of X is given by

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

- Family of distributions indexed by λ
- Defined on the infinite set of non-negative integers
- Used for modelling rare events

Statisfies the properties of pmf:

Mean and variance

If $X \sim Poisson(\lambda)$ then

$$E(X) = var(X) = \lambda$$

Proof:

The relationship between the binomial and Poisson distributions

When n is large in compare to np ($n \gg np$, rare event)

$$Binomial(n, p) \approx Poisson(np)$$

Proof:
$\lim_{n \to \infty} p_X(x) =$

Example 4.15

A rare type of blood occurs in a population with frequency .001. If n people are tested, what is the probability that at least two people have this rare blood type?

Example

A box containing 5000 eggs arrives at a grocery store where the owner randomly selects 25 eggs to examine and returns the box if there are more than three broken eggs. If 0.5% of the eggs are broken, what is the probability that the box is accepted?

The Poisson Process

- $X_{(t_1,t_2)}$: the number of successes in the (t_1, t_2) interval.
  e.g. the number of cars passing a section of a highway between times t_1 and t_2.
- $p(x; t_1, t_2)$: the probability that x successes occur in the time interval (t_1, t_2).

Assumptions of the Poisson process

1. $X_{(t_1,t_2)}$ and $X_{(t_1',t_2')}$ are independent if $t_1' > t_2$ or $t_1 > t_2'$
   (non-overlapping time intervals)
2. $p(x; t_1, t_2) = p(x; t_2 - t_1)$: stationary for all time intervals (t_1, t_2)
3. $p(1, \delta t) = \lambda(\delta t)$, where δt is a small time interval.
   Also $p(0, \delta t) = 1 - p(1, \delta t) = 1 - \lambda(\delta t)$

Example

A switchboard receives calls at a rate of three per minute during a busy period. What is the probability of receiving more than two calls in a two minute interval during the busy period?

Example (review)

A limousine can accommodate up to four passengers. The company accepts up to six reservations and passengers must have a reservation to travel.From records 20% of passengers with reservations do not show up. (a) If six reservations are made, what is the probability that at least one passenger cannot be accommodated? (b) If six reservations are made, what is the expected number of available places when the limo departs? (c) Suppose that the pmf of the number of reservations R is

r	3	4	5	6
p(r)	0.1	0.2	0.3	0.4

Obtain the pmf of the number of passengers X who show up.