STA255 Week-2

Shahriar Shams

13/01/2020

Review of Week-1

- Probability function
- Conditional Probability
  - probability calculated based on a restricted sample space
  - Probability of A conditioning on B, $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- Law of total probability
  - probability of event A calculated as a sum of bunch of conditional probabilities multiplied by the probability of those conditions.
  - $P(B) = P(B|A_1) * P(A_1) + P(B|A_2) * P(A_2) + ... + P(B|A_k) * P(A_k)$
- Bayes rule
  - Inverting a conditional probability
  - $P(A_j|B) = \frac{P(B|A_j)*P(A_j)}{P(B|A_1)*P(A_1)+P(B|A_2)*P(A_2)+...+P(B|A_k)*P(A_k)}$
- Independence
  - Event A and B are independent if any of the following is true

$$P(A|B) = P(A) \;\; or \;\; P(B|A) = P(B) \;\; or \;\; P(A \cap B) = P(A) * P(B)$$

Learning goals

- Random variable (Discrete random variable)
- Probability Mass Function
- Cummulative Distribution Function
- Some common discrete distributions (Bernoulli, Binomial, Geometric, Negative Binomial and Poisson)

Discrete Random Variable

A *random variable* is a function that maps the sample space (S) on to real numbers $\mathbb{R}$

Let revisit an example we used last week

- We are rolling a fair dice and we are only interested in the **sum of the two numbers**.
- Sample space S includes 36 outcomes: (1,1), (1,2),. . . ,(2,1), (2,2),. . . ,(6,6)
- Let us define A which represents the sum of the two numbers.
- A can take value 2, 3, 4,. . . , 12
- Each of the 36 elements of S will result in one element from the list of $\{2, 3, 4, ..., 12\}$

- A has "transformed" S into a new sample space $\tilde{S} = \{2, 3, 4, ..., 12\}$
- We already know $P(A = 11) = P[(6, 5) \text{ or } (5, 6)] = 2/36$ (you can do the rest)

Another example:

- Let M represents the maximum of the two numbers.
- M can take values $\{1, 2, 3, 4, 5, 6\}$
- M transforms S into $\tilde{S} = \{1, 2, 3, ..., 6\}$

Discrete random variable

- A random variable that takes **countable** number of values is called a *Discrete* random variable.
- In other words, if the transformed sample space $\tilde{S}$ is countable (finite/infinite) then the random variable is *discrete*.
- A and M both are examples of discrete random variable.

Probability Mass Function (pmf)

A *probability mass function*, denoted by p(), of a discrete random variable X is function from $\mathbb{R}$ to [0,1] which is defined as

$$p(x) = P(X = x) \text{ where, } -\infty < x < \infty$$

In plain language, a probability mass function, $p(x)$ of a discrete random variable X gives us the probability that X will take value x

Example:

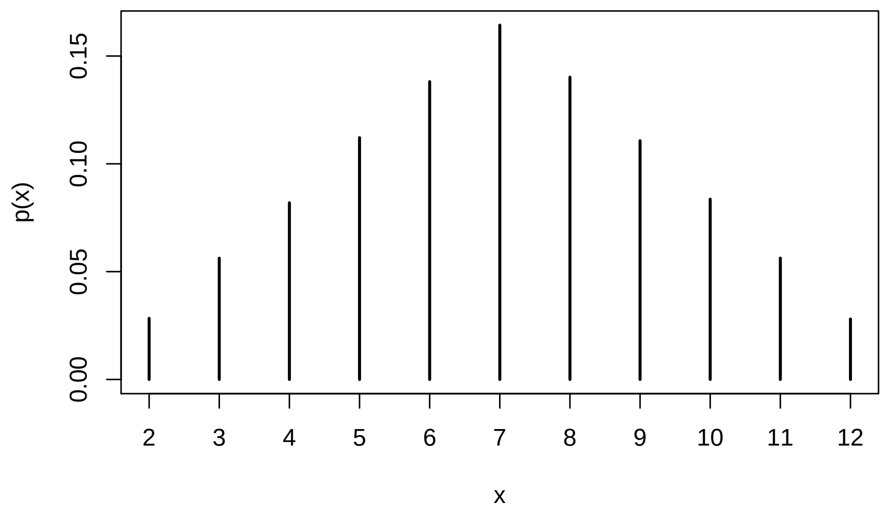
- Rolling a fair dice twice
- Random variable, A = sum of the two numbers
- probability mass function of A will look like the following table

x	2	3	4	...	11	12
p(x)=P(A=x)	1/36	2/36	3/36	...	2/36	1/36

Some intuitive properties of *pmf* :

- Let X be the random variable taking values $x_1, x_2, x_3, ..., x_m$ then
  - $P(x) \geq 0$
  - $p(x_1) + p(x_2) + ... + p(x_m) = 1$
  - $p(x) = 0$ for any value x that does not belong to $\{x_1, x_2, x_3, ..., x_m\}$

Graph of pmf



Cummulative Distribution Function (cdf)

Cummulative Distribution function, denoted by $F()$, of random variable X is a function from $\mathbb{R}$ to [0,1] which is defined as

$$F(x) = P(X \leq x) \text{ where, } -\infty < x < \infty$$

In plain language, distribution function, $F(x)$ adds all the probabilities of X starting from $-\infty$ to the value x.

For our example,

$$\begin{aligned} F(3) &= P(A \leq 3) \\ &= P(A = 2) + P(A = 3) = 1/36 + 2/36 = 3/36 \end{aligned}$$

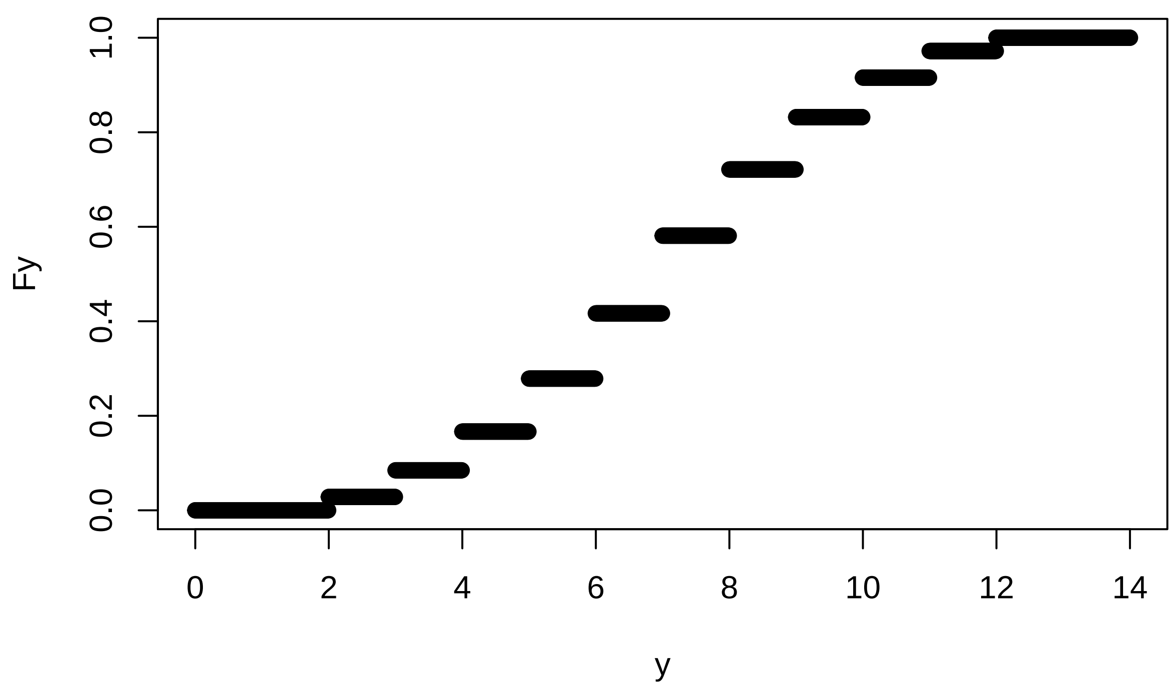
Similarly, we can calculate $F(x)$ for all the values of x

$$F(x) = \begin{cases} 0 & when, x \in (-\infty, 2) \\ 1/36 & when, x \in [2, 3) \\ 3/36 & when, x \in [3, 4) \\ ... & ... \\ 35/36 & when, x \in [11, 12) \\ 1 & when, x \in [12, \infty) \end{cases}$$

Some properties of distribution function:

- if $a \leq b$ then $F(a) \leq F(b)$.
- $F(-\infty) = 0$ and $F(\infty) = 1$
- F(x) is **right continuous**.

Graph of *cdf*



Some common discrete distributions

Bernoulli Distribution:

A random variable X has a *Bernoulli* distribution with parameter p (it's a number between 0 and 1) if its probability mass function is

$$p_X(x) = \begin{cases} p & when \quad x = 1 \\ 1-p & when \quad x = 0 \end{cases}$$

In notation, we say $X \sim Bern(p)$

In some books this same function is also written in the following form (which is exactly the same thing),

$$p_X(x) = p^x(1-p)^{(1-x)} \text{ where, x =0, 1}$$

Note:

- this is the simplest type of distribution.
- Any experiment with two outcomes will result in this distribution.
  - tossing a coin, let $X = 1$ if it's a head and $X = 0$ if it's a tail
- Any random variable can be transformed into a Bernoulli random variable.
  - for the A=sum of two faces example, let $X = 1$ if $A \leq 6$ and $X = 0$ otherwise.
  - Let Y be the height of students. Let $X = 1$ if $Y < 160cm$ and $X = 0$ otherwise.

Binomial Distribution:

- Binomial is the sum of n independent Bernoulli random variables.
- Suppose $X_1, X_2, ..., X_n$ are all independent of each other.
- Individually they all have a $Bern(p)$ distribution.
- Let Y be another random variable where, $Y = \sum_{i=1}^{n} X_i$
- Y will have a Binomial distribution with parameter n and p
- In notation, $Y \sim Bin(n, p)$

then *pmf* of Y is given by

$$p_Y(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ where, } x = 0, 1, 2, ..., n$$

Here,

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

Example: (we have learned this last week)

- Tossing a coin $n = 3$ times (or tossing three independent coins)
- Let X_1 be the outcome of the first toss with $X_1 = 1$ if it's a head , 0 otherwise
- Similarly X_2 and X_3 represent the second and the third toss.
- Each time the probability of head is p
- we can write $X_1 \sim Bern(p)$, $X_2 \sim Bern(p)$ and $X_3 \sim Bern(p)$
- Let Y = the number of heads in $n = 3$ tosses $\implies Y = X_1 + X_2 + X_3$

$$\begin{aligned}
P(Y = 1) &= P(X_1 + X_2 + X_3 = 1) \\
&= P\left(Y = 1 + 0 + 0 \;\; or \;\; Y = 0 + 1 + 0 \;\; or \;\; Y = 0 + 0 + 1\right) \\
&= p(1 - p^2) + p(1-p)^2 + p(1-p)^2 \\
&= 3.p(1-p)^2 \\
&= \binom{3}{1} p^1 (1-p)^{3-1}
\end{aligned}$$

$\binom{n}{k}$ gives us the number of ways we can have a sum of k out of n Bernoulli variables.

Example: Exercise 62 (page 136)

$n = \#$ of drivers $= 20$

$P[\text{stopping}] = 25\% = 0.25$

Let. $Y = \#$ of drivers who stopped.

- $P[\text{exactly one driver stopped}]$

$= P[Y = 1] = \binom{20}{1}(0.25)^1(0.75)^{19}$

- $P[\text{atmost 6 will stop}]$

$= P[Y \leq 6]$

$= P[Y=0] + P[Y=1] + \dots + P[Y=6]$

$= \binom{20}{0}(0.25)^0(0.75)^{20} + \binom{20}{1}(0.25)^1(0.75)^{19} + \dots$

- $P[\text{at least 6 will stop}]$

$= P[Y \geq 6]$

$= 1 - P[Y \leq 5]$

$= 1 - \left(P[Y=0] + P[Y=1] + \dots + P[Y=5]\right)$

$=$

Geometric distribution

- Suppose we have a Bernoulli trail (with two outcomes: success and failure).
- Geometric distribution gives us the probability that the first success will come after the x^{th} failure, $x = 0, 1, 2, ...$
  - Example: tossing a coin until a head appears
- For each trial, let the probability of success be p.
- First success on the $(x + 1)^{th}$ trial means *failure on the first x trial* and *success on the $(x + 1)^{th}$ one*
- the probability is then $(1 - p)(1 - p)(1 - p)...(1 - p)p = (1 - p)^x p$

A random variable X has a geometric distribution with parameter p is it has a pmf

$$p_X(x) = (1-p)^x p \quad \text{where} \quad x = 0, 1, 2, ...$$

In notation we write, $X \sim Geom(p)$

Example: Exercise 35 (page 73)

total # of songs = 100

# of Beatles song = 10

# of non-Beatles song = 90

- Assumption: Same song can be played over & over ⇒ prob. of any song being played does not change.

$P[\text{first Beatles song is the 5th song played}]$

$= P[B'B'B'B'B]$

$= \frac{90}{100} * \frac{90}{100} * \frac{90}{100} * \frac{90}{100} * \frac{10}{100}$

↳ Geometric distn

- Assumption: The song that just has been played can not be played again. ⇒ prob. changes from one song to the next one.

$P[B'B'B'B'B]$

$= \frac{90}{100} * \frac{89}{99} * \frac{88}{98} * \frac{87}{97} * \frac{10}{96}$

↳ this is NOT a geometric distn problem

Negative Binomial distribution

- It's a generalization of Geometric distribution.
- In Geometric distribution we count the number of failures until we get the first success.
- In Negative Binomial distribution, we count the number of failures **until the 2nd, 3rd, 4th or** r^{th} **success**.

For example,

- Suppose we are counting number of failures until the 2nd success.
- In all the possibilities, the last trial is a success

- Probability of success in any trial is p
- If the total trial is is 10,
  - this means we have 2 success and 8 failures
  - the last one is a success and the other success might have come in any of the previous 9 trials
  - the number of combinations is then $\binom{9}{1}$
  - for each combination the probability is $(1-p)^8p^2$
  - Hence, the probability of the count of failure being 8 is $\binom{9}{1}(1-p)^8p^2$
- Let's generalize, if the count of failure is x and we are looking for r success
  - r success means x failure $\implies$ the probability of each combination is $(1-p)^xp^r$
  - Given the last trial is a success, the previous $(r-1)$ successes might have come from any of the previous $(x+r-1)$ trial
  - the number of combination is then $\binom{x+r-1}{r-1}$

A random variable X has a Negative Binomial distribution with parameter r and p (r is the number of success we are looking for, p is the probability of success in each trial) if it has a pmf

$$p_X(x) = \binom{x+r-1}{r-1}(1-p)^xp^r \quad \text{where,} \quad x = 0, 1, 2, ...$$

In notation we write, $X \sim NB(r, p)$

Note: Geometric distribution is a special case of Negative Binomial. If we put $r = 1$ in Negative binomial distribution, we end up with Geometric distribution.

Poisson distribution

- Binomial distribution is characterized using two parameters: number of trials (n) and probability of success in each trial(p).
- When the value of n approaches infinity (∞) and p approaches 0, the functional form of a binomial pmf takes a different form.
- This limiting distribution ($n \to \infty$, $p \to 0$) is called Poisson distribution.
- Typically a Poisson distribution is characterized using the parameter λ.
- The relationship between λ and the two parameters of Binomial distribution is

$$\lambda = n.p$$

A random variable X has a Poisson Binomial distribution with parameter λ if it has a pmf

$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{where,} \quad x = 0, 1, 2, ...$$

In notation we write $X \sim Pois(\lambda)$

- λ often represents "rate".
- For example: In a certain coffee store, on an average 10 customers come into the store in any given 15 minute window. What is the probability that 5 customers will come in the next 15 minutes?
  - the rate here is 10 per 15-min. So $\lambda = 10$
  - Using the Poisson pmf

$$p(5) = \frac{e^{-10}10^5}{5!}$$

Exercise: Show $Bin(n, p) \to Pois(\lambda)$

Given in details on page 147 of the text book.

Permutation vs Combination

A B C D E

- Combination: How many $\boxed{\text{teams}}$ can be formed taking two letters from this list?

AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

10 in total $\equiv \binom{5}{2} \equiv {}^5C_2 = \frac{5!}{2!\,3!} = 10$

Order DOES NOT matter.

- Permutation: How many different $\boxed{\text{words}}$ can be made taking two letters from the list?

AB, AC, AD, ... 20 in total.
BA, CA, DA

$${}^5P_2 = \frac{5!}{(5-2)!} = \frac{5!}{3!} = 20$$

Order matters.

Homework

From the exercise

2, 5, 6, 7,

12, 13, 14, 15, 19, 23,

31, 35, 36, 43,

44, 45, 46, 49, 50, 51, 52,

60, 65, 69, 70, 72, 76, 77,

84, 90, 91, 92,

93, 94, 95, 104, 108,

113, 117, 127, 133

Note: Leave the parts of these problems that require you to calculate mean/variance/Moment generating funciton which you will learn in Week-4.