STAT 305: Chapter 5

Part II

Amin Shirazi

ashirazist.github.io/stat305.github.io

Discrete Random Variables

Meaning, Use, and Common Distributions

General Info

Reminder: RVs

General Info About Discrete RVs

Reminder: What is a Random Variable?

Random Variables, we have already defined, take real-numbered ($\mathbb{R}$) values based on outcomes of a random experiment.

- If we know the outcome, we know the value of the random variable (so that isn't the random part).
- However, before we perform the experiment we do not know the outcome - we can only make statements about what the outcome is likely to be (i.e., we make "probabilistic" statements).
- In the same way, we do not know the value of the random variable before the experiment, but we can make probability statements about what value the RV might take.

Common Terms and Notation for Discrete RVs

Of course, we can't introduce a *sort of* new concept without introducing a whole lot of new terminology.

We use capital letters to refer to discrete random variables: $\boxed{X, Y, Z, \ldots}$

We use lower case letters to refer to values the discrete RVs can take: $\boxed{x, x_1, y, z,}\ldots$

While we can use $P(X = x)$ to refer to the probability that the discrete random variable takes the value x, we usually use what we call the **probability function**:

- For a discrete random variable X, the probability function $f(x)$ takes the value $P(X = x)$ 3
- In otherwords, we just write $\boxed{f(x)}$ instead of $\boxed{P(X = x)}$.

General Info

Reminder: RVs

Discrete?

Terms & Notation

Common Terms and Notation for Discrete RVs

We also have another function related to the probabilities, called the **cumulative probability function**.

For a discrete random variable X taking values $x_1, x_2, \ldots$ the CDF or **cumulative probability function** of X, $F(x)$, is defined as

(CDF)

$$F(x) = \sum_{\boxed{z \leq x}} f(z)$$

Which in other words means that for any value x,

- $f(x) = P(X = x)$ ✓

and

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

- Discrete : probability mass function. (PMF)
- Continuous : probability density function. (PDF)

Note: a probability mass function (pmf)

gives probabilities of occurance for individual values

e.g $P(X=a)$, $P(X \le 3) = P(X=1) + P(X=2) + P(X=3) =$

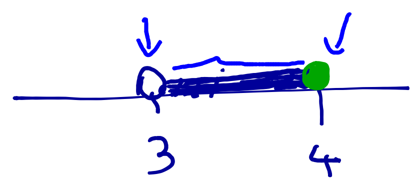
$X \in \{1, 2, \dots, 6\}$

$P(2 < X < 3)$ $\qquad P(X \in (2,3)) = 0$

(X takes individual (discrete) values. So, cannot take continuous values in an interval)

$P(3 < X \le 4)$

But, $P(X \in (3,4]) = P(X=4)$



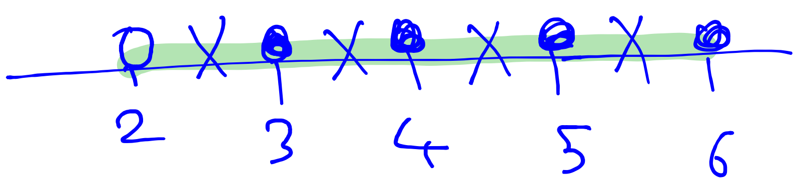
Note: in discrete r.v:

$$P(X=x_1, X=x_2) = P(X=x_1) + P(X=x_2)$$

$$= f(x_1) + f(x_2)$$

In rolling a die:

$$P(2 < X \le 6) = P(X=3) + P(X=4) + P(X=5) + P(X=6)$$



Important !

properties of a valid Probability function:

(No matter discrete or continuous probability function)

1. $f(x) \geq 0 \quad \forall x \in S$

2. $\boxed{\sum_{x} f(x) = 1}$ (probabilities sum to 1)

e.g: rolling a fair dice once:

x	1	2	3	4	5	6
$f(x): P(X=1)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

$f(x) \geq 0 \quad \forall x \in \{1, \dots, 6\}$

$+ \sum_{x=1}^{6} f(x) = 1$

e.g.: let $f(x)$ is a probability mass function

of the form:

$$f(x) := P(X=x) = \begin{cases} 1/8 & x=-1,1 \\ a & x=0 \\ 1/4 & x=-2,2 \end{cases}$$

what "a" must be to make $f(x)$ a valid pmf?

① $\underbrace{f(x)}_{=f(0)=a} \geq 0 \quad \forall x \in \{-2,-1,0,1,2\} \Rightarrow a \geq 0$

② $\sum_{x\in\{-2,-1,0,1,2\}} f(x) = 1 \rightarrow f(-1)+f(1)+f(0)+f(2)+f(-2)$

$= 1/8 + 1/8 + a + 1/4 + 1/4$ must be $= 1$

$\rightarrow \quad a + 2/8 + 1/2 = 1 \Rightarrow \boxed{a = 1/4}$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Common Terms and Notation for Discrete RVs (cont)

The values that X can take and the probabilities attached to those values are called the **probability distribution** of X (since we are talking about how the total probability 1 gets spread out on (or distributed to) the values that X can take).

Example

Suppose that the we roll a die and let T be the number of dots facing up. Define the probability distribution of T. Find $f(3)$ and $F(6)$.

$P(T=3)$

CDF $P(T \le 6)$

T	1	2	3	4	5	6
$f(T)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

$f(3) = P(T=3) = 1/6$ | $F(6) = P(T \le 6) = \sum_{t=1}^{6} f(t) = 1$

$$F(3) \overset{\text{def.}}{=} P(T \le 3) = P(T=1) + P(T=2)$$
$$+ P(T=3)$$
$$= 1/6 + 1/6 + 1/6$$
$$= 3/6 = 1/2$$

General Info

Reminder: RVs

Discrete?

Terms & Notation

Example: [Torque]

Let $Z =$ the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

Calculate the following probabilities:

- $P(Z \leq 14)$ -
- $P(Z > 16)$
- $P(Z \text{ is even})$
- $P(Z \in \{15, 16, 18\})$

Example: [Torque]

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

- $P(Z \leq 14)$ $= P(Z=11 \text{ or } Z=12 \text{ or } Z=13 \text{ or } Z=14)$

$= P(Z=11) + P(Z=12) + P(Z=13)$

$+ P(Z=14)$

$= 0.03 + 0.03 + 0.03 + 0.06 = 0.15$

- $P(Z > 16)$

① $P(Z > 16) = P(Z=17 \text{ or } Z=18 \text{ or } Z=19 \text{ or } Z=20)$

$= P(Z=17) + P(Z=18) + P(Z=19) + P(Z=20)$

$= 0.12 + 0.2 + 0.15 + 0.03 = 0.5$ ✓

② $P(Z > 16) = 1 - P(\{Z > 16\}^c) = 1 - P(Z \leq 16)$

$= 1 - P(Z=11 \text{ or } Z=12 \text{ or } Z=13 \text{ or } Z=14 \text{ or }$

$Z=15 \text{ or } Z=16)$

$= 1 - 0.5 = 0.5$

Example: [Torque]

Z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03

- $P(Z \text{ is even})$ $= P(Z=12 \text{ or } Z=14 \text{ or } Z=16 \text{ or } Z=18 \text{ or } Z=20)$

  $= P(Z=12) + P(Z=14) + P(Z=16) + P(Z=18) + P(Z=20)$

  $= 0.03 + 0.06 + 0.09 + 0.2 + 0.03$

  $= 0.41$

- $P(Z \in \{15, 16, 18\})$

  $= P(Z=15 \text{ or } Z=16 \text{ or } Z=18)$

  $= P(Z=15) + P(Z=16) + P(Z=18)$

  $= 0.26 + 0.09 + 0.2 = 0.55$

General Info

Reminder: RVs

Discrete?

Terms & Notation

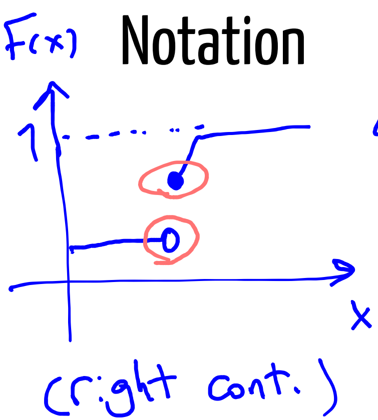
More on CDF

cumulative dist. function

> The *cumulative probability distribution (cdf)* for a random variable X is a function $F(x)$ that for each number x gives the probability that X takes that value or a smaller one, $F(x) = P[X \leq x]$.

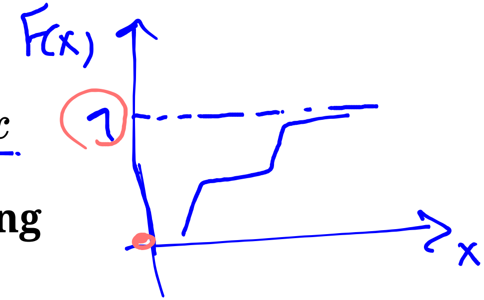
Since (for discrete distributions) probabilities are calculated by summing values of $f(x)$,

$$F(x) = P[X \leq x] = \sum_{y \leq x} f(y)$$



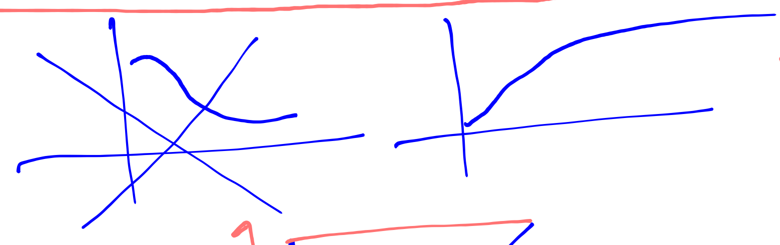
More on CDF

Properties of a mathematically valid cumulative distribution function:



- $F(x) \geq 0$ for all real numbers x
- $F(x)$ is monotonically **increasing**
- $F(x)$ is right continuous
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$
  - This means that $0 \leq F(x) \leq 1$ for **any CDF**

> In the discrete cases, the graph of $F(x)$ will be n stair-step graph with jumps at possible values of our random variable and height equal to the probabilities associated with those values

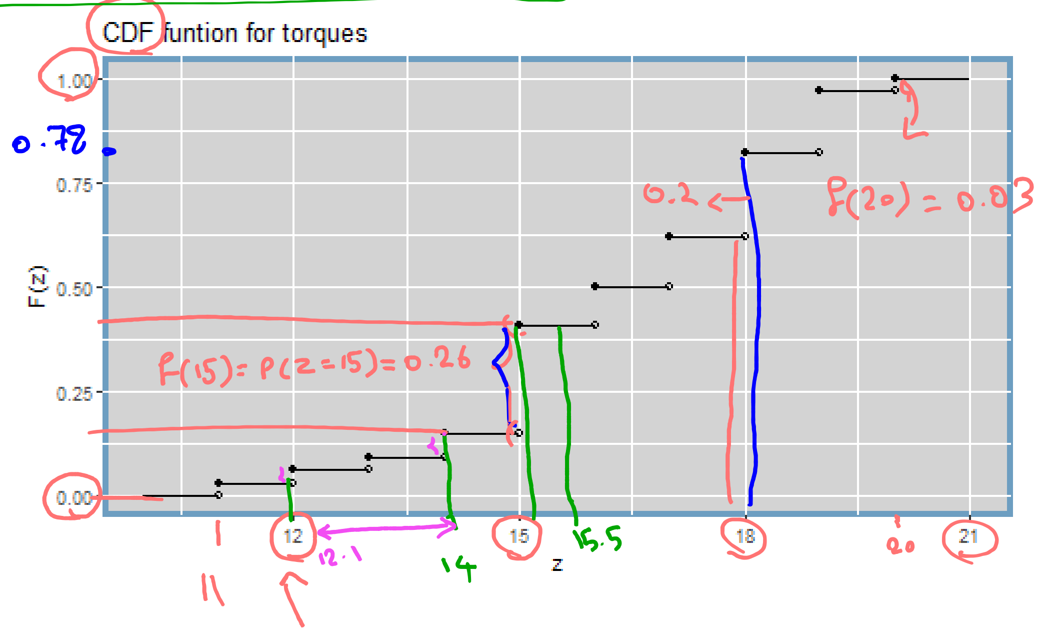


More on CDF

Example: [Torque] Let $Z =$ the torque, rounded to the nearest integer, required to loosen the next bolt on an apparatus.

pmf =

z	11	12	13	14	15	16	17	18	19	20
f(z)	0.03	0.03	0.03	0.06	0.26	0.09	0.12	0.20	0.15	0.03



General Info

Reminder: RVs

Discrete?

Terms & Notation

More on CDF

Calculate the following probabilities using the **cdf only**:

- $F(10.7)$ $= P(Z \leq 10.6) = 0$
- $P(Z \leq 15.5)$ $= P(Z \leq 15.5) = F(15)$

$= P(Z \leq 15)$

$= \sum_{z=11}^{15} f(z) = 0.41$

- $P(12.1 < Z \leq 14)$

$= P(13 \leq Z \leq 14) = P(Z=13 \text{ or } Z=14)$

$= f(13) + f(14) = 0.09$

- $P(15 \leq Z < 18)$

$= P(15 \leq Z \leq 17) = P(Z=15 \text{ or } Z=16 \text{ or } Z=17)$

$= P(Z=15) + P(Z=16) + P(Z=17) = 0.47$

More on CDF

One more example

Say we have a random variable Q with pmf:

q	f(q)
1	0.34
2	0.10
3	0.22
7	0.34

$F(q)$

$P(q \leq 1) = P(q = 1) = 0.34$

$P(q \leq 2) = 0.44$

$P(q \leq 3) = 0.66$

$P(q \leq 4) = 1$

Draw the $\boxed{\text{CDF}}$