Discrete random variables

June 7, 2023

Objectives

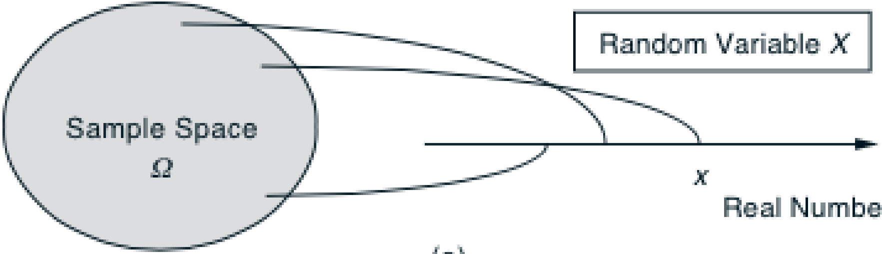
1. Understand random variables
2. For discrete random variables
   a. Determine probabilities from probability mass functions and the reverse
   b. Determine probabilities from cumulative distribution functions and cumulative distribution functions from probability mass functions, and the reverse

Intro

- Sample space and events are basic components of probability
- Similar to numbers in calculus
- Study the relations between numbers we use functions
- In probability we use *random variables*

Definition

A **random variable** X defined on a sample space Ω is a quantity that is calculated by the outcomes - a **function of outcomes**



(a)

Example

- Toss a coin three times
- X be the number of times that tails appear
- Sample space $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$
- $X(HHH) = 0, X(HHT) = 1 \dots$

w	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(w)$	$\boxed{0}$	1	1	2	1	2	2	3

All possible values of X are 0, 1, 2, 3. We say $Range(X) = \{0, 1, 2, 3\}$

$$P(X = 0) = P(\text{there is no tail}) = P(HHH) = \frac{1}{8}$$

$$P(X = 1) = P(\text{there is 1 tail}) = P(\{HHT, HTH, THH\}) = \frac{3}{8}$$

- Tossing a fair coin until a Head appear. The sample space is

$$\Omega = \{H, TH, TTH, TTTH, \ldots\}$$

- X: the number of tossing
- $Range(X) = \{1, 2, 3, \ldots \ldots\}$
- $X = 1$ if and only if the first coin turns head so $P(X = 1) = P(\text{first coin turns H}) = 0.5$

Question

Evaluate

$$P(X = 2)$$

and

$$P(1 < X \leq 4)$$

Types of RV

Based on **range of the random variable** X

- If the set of possible values of X is **finite or countable** like $\{0, 1, 2, 3\}, \{1, 2, \ldots\}$ then X is called *discrete RV*
- If the set of possible values of X is **uncountable** (like the interval $[a, b]$, $[a, \infty)$) then X is called *continuous RV*

Probability associated with a random variable X

$$P(X = a), P(X \geq a), P(X > a),$$
$$P(X \leq b), P(X < b),$$
$$P(a \leq X \leq b), P(a < X \leq b),$$
$$P(a \leq X < b), P(a < X < b)$$

Cumulative distribution function (cdf)

cdf of $F(.)$ of the random variable X is a function defined by

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

is the probability that X takes on values less than or equal to b

Use cdf to answer questions about a random variable

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

$$P(X < b) = \lim_{h \to 0^+} P(X \leq b - h) = \lim_{x \to b^-} F(x) = F(b^-)$$

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a)$$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

$$P(a \leq X \leq b) = F(b) - F(a^-)$$

$$P(a < X < b) = F(b^- - F(a))$$

Discrete Random Variables

Random variable can take on at most countable number of possible values

$$Range(X) = \{x_1, x_2, \dots\}$$

Probability mass function

The *probability mass function* (p.m.f) of the discrete random variable$X$ is defined as

$$p(x_i) = P(X = x_1) \text{ for all } x_i \in Range(X)$$

the probability that the value of X is equal to x_i

Example

Roll two fair dice then the sample space is

$$\Omega = \{(1,1,),\ldots,(6,6)\} = \{(i,j) : 1 \leq i,j \leq 6\}$$

Let X be the largest of numbers on two dice, i.e if the rolling result is (i,j) then

$$X(i,j) = \max(i,j)$$

Table values of X

		1st dice					
2nd dice		1	2	3	4	5	6
	1	1	2	3	4	5	6
	2	2	2	3	4	5	6
	3	3	3	3	4	5	6
	4	4	4	4	4	5	6
	5	5	5	5	5	5	6
	6	6	6	6	6	6	6

All possible values of X is

$$Range(X) = \{1, 2, 3, 4, 5, 6\}$$

so X is a discrete random variable.
In order to determine the pmf of X, we need to find all the probabilities

$$P(X = 1), P(X = 2), \ldots, P(X = 6)$$

$X = 1$ if and only if the outcome is $(1, 1)$. So

$$P(X = 1) = P((1, 1)) = 1/36$$

$X = 2$ if and only if the outcomes is one of $(1, 2), (2, 2), (2, 2)$. So

$$P(X = 2) = P(\{(1, 2), (2, 1), (2, 2)\})$$
$$= P((1, 2)) + P((2, 1)) + P((2, 2) = 3/36$$

Similar, we have $P(X = 3) = 5/36, P(X = 4) = 7/36, P(X = 5) = 9/36, P(X = 6) = 11/36$.

We can summary p.m.f of X in the p.m.f table

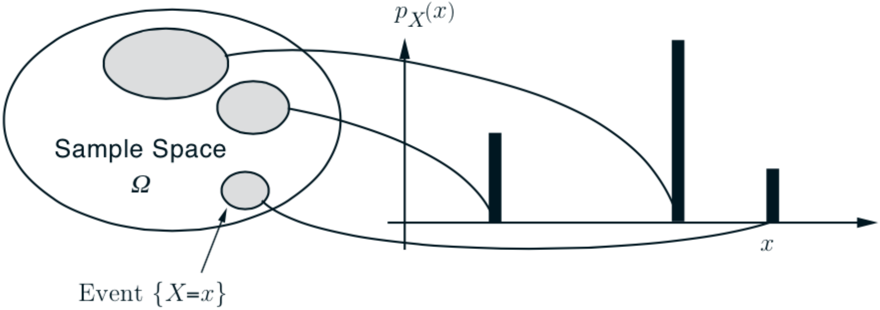
x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Similar, we have $P(X = 3) = 5/36, P(X = 4) = 7/36, P(X = 5) = 9/36, P(X = 6) = 11/36.$

We can summary p.m.f of X in the p.m.f table

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Illustration to calculate pmf



For each possible value x, we collect all the outcomes that give rise to $X = x$ and add their probabilities to obtain $p_X(x) = P(X = x)$.

One can use p.m.f of the discrete random variable X to answer any question of X such as

$$P(1 < X < 4) = P(X = 2 \text{ or } X = 3)$$
$$= P(X = 2) + P(X = 3) = 3/36 + 5/36$$

using additive rule for disjoin set

$$P(A \cup B) = P(A) + P(B) \text{ if } A \cap B = \emptyset$$

for $A = \{X = 2\}, B = \{X = 3\}$

Practice

Probability mass function of discrete random variable X is

x	-2	-1	0	1	2
$p(x) = P(X = x)$	$1/8$	$2/8$	$2/8$	$2/8$	$1/8$

Determine

1. $P(X \leq -1 \text{ or } X = 2)$
2. $P(-1 \leq X \leq 1)$

Practice

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, **find the probability mass function (p.m.f)** for the number of defectives.

Practice

An urn contains 11 balls, 3 white, 3 red, and 5 blue balls. Take out 3 balls at random, without replacement. You win \$1 for each red ball you select and lose a \$1 for each white ball you select. Determine the p.m.f of your loss/profit X.

Properties of pmf

- X is discrete
  $\rightarrow Range(X) = \{x_1, \ldots, x_n \ldots\}$
- $p(x_i) = P(X = x_i) \geq 0$
- $P(X \in A) = \sum_{x_i \in A} p(x_i)$
- **Normalization**

$$P(-\infty < X < \infty) = 1 \Rightarrow \sum_{i=1}^{\infty} p(x_i) = 1$$

Properties of pmf

- X is discrete
  $\to Range(X) = \{x_1, \ldots, x_n \ldots\}$
- $p(x_i) = P(X = x_i) \geq 0$
- $P(X \in A) = \sum_{x_i \in A} p(x_i)$
- **Normalization**

$$P(-\infty < X < \infty) = 1 \Rightarrow \sum_{i=1}^{\infty} p(x_i) = 1$$

Properties of pmf

- X is discrete
  $\rightarrow Range(X) = \{x_1, \ldots, x_n \ldots\}$
- $p(x_i) = P(X = x_i) \geq 0$
- $P(X \in A) = \sum_{x_i \in A} p(x_i)$
- **Normalization**

$$P(-\infty < X < \infty) = 1 \Rightarrow \sum_{i=1}^{\infty} p(x_i) = 1$$

Example

Suppose X has 3 values $1, 2, 3$ and

$$p(1) = \frac{1}{2},\ p(2) = \frac{1}{3}$$

then what is $p(3)$?

$$p(3) = 1 - p(1) - p(2) = 1/6.$$

Example

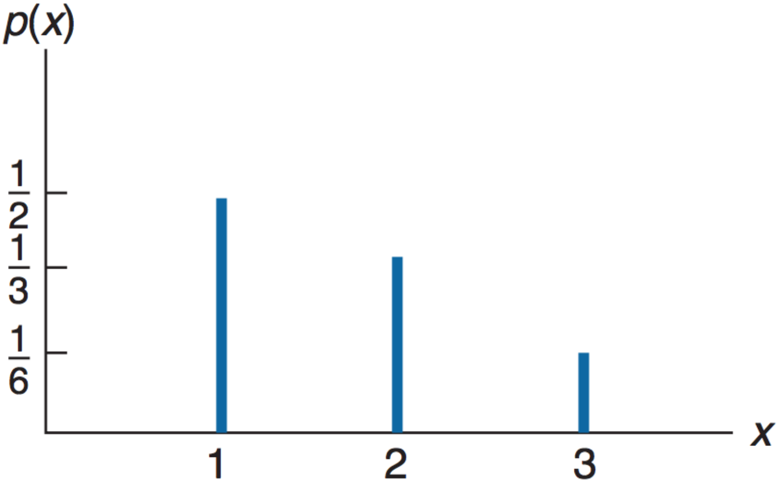
Suppose X has 3 values $1, 2, 3$ and

$$p(1) = \frac{1}{2},\ p(2) = \frac{1}{3}$$

then what is $p(3)$?

$$p(3) = 1 - p(1) - p(2) = 1/6.$$

Graph of $p(x)$



Suppose that the pmf of random variable X is given by

$$p(x) = c(x+5), \;\; x = 0, 1, 2, 3, 4$$

Find c and $P(0 < X < 2.5)$.

Cumulative distribution function (cdf)

Probability that X does not exceed a given value

$$F(b) = P(X \leq b) = \sum_{x_i \leq b} P(X = x_i)$$

Example

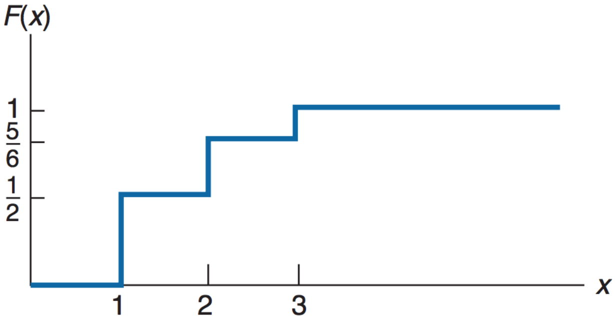
Suppose that pmf of X is given by $p(1) = \frac{1}{2}, p(2) = \frac{1}{3}, p(3) = \frac{1}{6}$ then

$$F(0.5) = P(X \leq 0.5) = 0$$

$$F(2.4) = P(X \leq 2.4) = P(X = 1) + P(X = 2) = \frac{5}{6}$$

The formula of the cdf of X is

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{2}, & 1 \leq x < 2 \\ \frac{5}{6}, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$



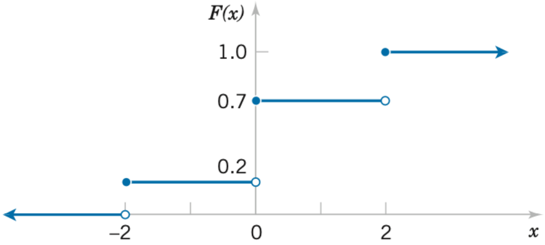
Remark

Jump size at 1 is $P(X = 1)$, ...

Determine the p.m.f of X from the c.d.f

$$F(x) = \begin{cases} 0 & \text{if } x < -2 \\ 0.2 & \text{if } -2 \leq x < 0 \\ 0.7 & \text{if } 0 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

Graph of $F(x)$ is



$P(X=a)=F(a)-F(a^-)$ is nonzero at the points $-2, 0, 2$.
The p.m.f at each point is the change (jump size) of c.d.f at the point

$$
\begin{aligned}
p(-2) &= P(X=-2) = F(-2) - F(-2^-) \\
&= 0.2 - 0 = 0.2 \\
p(0) &= P(X=0) = F(0) - F(0^-) \\
&= 0.7 - 0.2 = 0.5 \\
p(2) &= P(X=2) = F(2) - F(2^-) \\
&= 1 - 0.7 = 0.3
\end{aligned}
$$

Properties of cdf of a discrete RV

- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
- F_X has a piecewise constant and staircase-like form.

c.d.f of discrete random variable X is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ 0.7 & \text{if } 1 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases}$$

Compute

1. $P(X \leq 2)$ and $P(X > 2)$
2. $P(1 \leq X \leq 2)$

Keywords

- pmf of a discrete RV with range $\{x_1, \ldots, x_n, \ldots\}$

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

  - $0 \leq p(x_i) \leq 1$
  - $\sum p(x_i) = 1$
  - $P(X \in A) = \sum_{x_i \in A} p(x_i)$

- cdf

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$