Discrete Random Variables

Chapter 3: Discrete Random Variables

Anduo Wang
Temple University
Email: anduo.wang@gmail.com
https://cis-linux1.temple.edu/∼tug29203/25fall-2033/index.html

Random Variables

- To analyze random experiments, we focus on numerical aspects of the experiment.
- For example, if an entire soccer game is a random experiment, then numerical results like goals, shots, fouls, etc are *random variables*.
- **Random Variable** is a real valued variable whose value is determined by an underlying random experiment.

Random Variable- A Simple Example

- I toss a coin five times. This is a random experiment.
- The sample space is

$$S = \{TTTTT, TTTTH, ..., HHHHH\}.$$

- Say we are interested in the number of heads.
- We define random variable X whose value is the number of observed heads.
- X can take values 0,1,2,3,4 or 5 depending on the outcomes of the experiment.
- For example, $X = 0$ for the outcome $TTTTT$ and $X = 2$ for $THTHT$.
- Thus we see X is a function from the space S to real numbers.

Example S: Snakes and Ladders

- the board game "Snakes and Ladders," where the moves are determined by the sum of two independent throws with a die
- the obvious sample space

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, 2, \cdots, 6\}\} \\ &= \{(1,1), (1,2), \cdots, (1,6), (2,1), \cdots, (6,5), (6,6)\}.\end{aligned}$$

- players of the game only interested in the sum of the outcomes of the two throw

$$S(\omega_1, \omega_2) = \omega_1 + \omega_2, \text{for } (\omega_1, \omega_2) \in \Omega.$$

Example S: Snakes and Ladders

	ω_1					
ω_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure: Two throws with a die and the corresponding sum

- denote the event that the function S attains the value k by $S = k$

$$\{S = k\} = \{(\omega_1, \omega_2) \in \Omega : S(\omega_1, \omega_2) = k\}.$$

Example S: Snakes and Ladders

	ω_1					
ω_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure: Two throws with a die and the corresponding sum

- denote the event that the function S attains the value k by $S = k$

$$\{S = k\} = \{(\omega_1, \omega_2) \in \Omega : S(\omega_1, \omega_2) = k\}.$$

- list the outcomes in the event $\mathrm{S} = 8$.

Example S: Snakes and Ladders

	ω_1					
ω_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure: Two throws with a die and the corresponding sum

- denote the probability of the event $\{S = k\}$ by $P(S = k)$

Example S: Snakes and Ladders

ω_2 \ ω_1	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure: Two throws with a die and the corresponding sum

- denote the probability of the event $\{S = k\}$ by $P(S = k)$
- $P(S = 2) = P((1, 1)) = 1/36$

Example S: Snakes and Ladders

	ω_1					
ω_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure: Two throws with a die and the corresponding sum

- denote the probability of the event $\{S = k\}$ by $P(S = k)$
- $P(S = 2) = P((1, 1)) = 1/36$
- $P(S = 3) = P(\{(1, 2), (2, 1)\}) = 2/36$

Example S: Snakes and Ladders

	ω_1					
ω_2	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Figure: Two throws with a die and the corresponding sum

- denote the probability of the event $\{S = k\}$ by $P(S = k)$
- $P(S = 2) = P((1, 1)) = 1/36$
- $P(S = 3) = P(\{(1, 2), (2, 1)\}) = 2/36$
- determine $P(S = k) fork = 4, 5, \cdots, 12$

Example M

ω_2 \ ω_1	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

Figure: Two throws with a die and the corresponding maximum

- In this case we are interested in the value of the function $M : \Omega \to \mathbb{R}$, given by

$$M(\omega_1, \omega_2) = max\{\omega_1, \omega_2\}\text{for } (\omega_1, \omega_2) \in \Omega$$

Important Definitions

Random Variables: A random variable X is a function from the sample space to the real numbers.

$$X : \Omega \to \mathbb{R}$$

The range of a random variable X, shown by Range(X) or R_X, is the set of possible values of X.

Important Definitions

Random Variables: A random variable X is a function from the sample space to the real numbers.

$$X : \Omega \to \mathbb{R}$$

a discrete random variable X "transforms" a sample space Ω to a more "tangible" sample space $\tilde{\Omega}$

Discrete Random Variables

- Two important classes of random variables- *discrete* and *continuous*.
- A third class *mixed random variables* can be thought of as a mixture of discrete and continuous random variables.
- We define a discrete random variable as the following

  X is a discrete random variable if its range is countable.

- A set A is countable if either
  - A is a finite set such as $\{1, 2, 3, 4\}$, or
  - it can be put in one-to-one correspondence with natural numbers (in this case, the set is said to be countably infinite).

Probability Mass Function (PMF)

- A event $A = \{X = a\}$ is defined as the set of outcomes s in the sample space S for which the value of X is a.

$$A = \{s \in S | X(s) = a\}.$$

- The probabilities of events $\{X = a\}$ is given by the **probability mass function (PMF) of X**.

> Let X be a discrete random variable with range $R_X = \{a_1, a_2, a_3, ...\}$ (finite or countably infinite). The function
>
> $$P_X(a_k) = \mathrm{P}(X = a_k), \text{ for } k = 1, 2, 3, ...,$$
>
> is called the *probability mass function (PMF)* of X.

Probability Mass Function (PMF)

- The PMF is a probability measure that gives us probabilities of the possible values of a random variable.
- We use P_X as the standard notation where the subscript indicates that this is the PMF of the random variable X.

Probability Mass Function (PMF)

Properties of PMF:

- $0 \leq P_X(a_i) \leq 1$ for all a_i;
- $\sum_{a \in R_X} P_X(a) = 1$;
- for any set $A \subset R_X$, $P(X \in A) = \sum_{a \in A} P_X(a)$.

Example

- the probability mass function p of M

a	1	2	3	4	5	6
$p(a)$	$1/36$	$3/36$	$5/36$	$7/36$	$9/36$	$11/36$

The probability mass function & the distribution function

Definitions

The probability mass function p of a discrete random variable X is the function $p : \mathbb{R} \to [0, 1]$, defined by

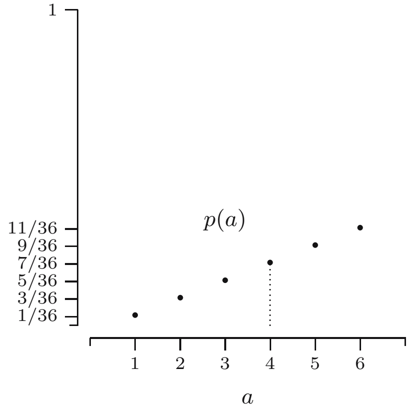
$$p(a) = P(X = a) \quad \text{for } -\infty < a < \infty.$$

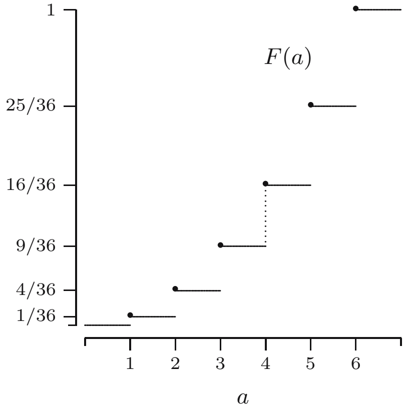
Definitions

The distribution function F of a random variable X is the function $F : \mathbb{R} \to [0, 1]$, defined by

$$F(a) = P(X \leq a) \quad \text{for } -\infty < a < \infty.$$

Example: Probability mass function and distribution function of M





Cumulative Distribution Function (CDF)

- Cumulative distribution function (CDF) of a random variable is a method to describe the distribution of random variables.
- The advantage of CDF is that it can be defined for any kind of random variable; while PMF cannot be defined for a continuous random variable.
- We have

Definition: The cumulative distribution function (CDF) of random variable X is defined as

$$F_X(x) = P(X \leq x), \text{ for all } x \in \mathbb{R}.$$

Cumulative Distribution Function (CDF)

- In general, let X be a discrete random variable with range $R_X = \{x_1, x_2, x_3, \ldots\}$, such that $x_1 < x_2 < x_3 < \ldots$.
- $F_X(x) = 0$ for $x < x_1$. Note that the CDF starts at 0, i.e., $F_X(-\infty) = 0$.
- CDF is in the form of a staircase. It jumps at each point in the range.
- CDF stays flat between x_k and x_{k+1}, so we can write

$$F_X(x) = F_X(x_k), \text{ for } x_k \leq x < x_{k+1}.$$

- CDF jumps at each x_k. We can write

$$F_X(x_k) - F_X(x_k - \epsilon) = P_X(x_k), \text{ for } \epsilon > 0 \text{ small enough.}$$

Cumulative Distribution Function (CDF)

- CDF is always a non-decreasing function, i.e., if $y \geq x$ then $F_Y(y) \geq F_X(x)$.
- It approaches 1 as x becomes large. We can write

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

- If we have the PMF, we can calculate the CDF. If $R_X = \{x_1, x_2, x_3, \ldots\}$, then

$$F_X(x) = \sum_{x_k \leq x} P_X(x_k).$$

- We have a useful formula that

For all $a \leq b$, we have

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

Cumulative Distribution Function (CDF)

- The CDF of a discrete random variable.

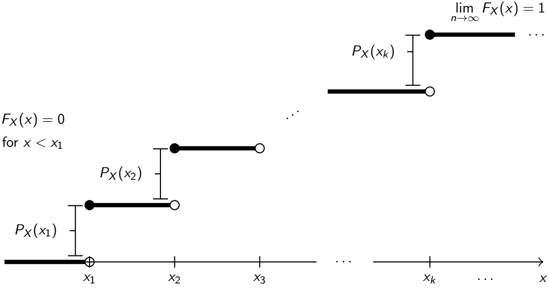


Figure: CDF of a discrete random variable

Cumulative Distribution Function

- Let X be a discrete random variable with range $R_X = \{1, 2, 3, ...\}$. Suppose the PMF of X is given by

$$P_X(k) = \frac{1}{2^k}$$

  - Find and plot the CDF of X, $F_X(x)$?
  - Find $P(2 < X \leq 5)$?
  - Find $P(X > 4)$?