

MA-250 Probability and Statistics

Nazar Khan

PUCIT

Lecture 15

RANDOM VARIABLES

Random Variables

- Random variables come in 2 types
 1. Discrete – set of outputs is real valued, countable set
 2. Continuous – set of outputs is real valued, uncountable set

Today we study discrete random variables.

DISCRETE RANDOM VARIABLES

Assigning Values to Outcomes

- Consider the sample space, S , consisting of all possible outcomes of four tosses of a coin.

$$S = \left\{ \begin{array}{cccc} HHHH & HHHT & HHTH & HTHH \\ THHH & HHTT & HTHT & THHT \\ TTHH & THTH & HTTH & TTTH \\ TTHT & THTT & HTTT & TTTT \end{array} \right\}$$

- Which outcome of this sample space do you like the most?
- What if I give you a 1 free shirt for every Head?

Random Variable

- What if I give you a 1 free shirt for every Head?

$HHHH \rightarrow 4$ $HHHT \rightarrow 3$ $HHTH \rightarrow 3$ $HTHH \rightarrow 3$
 $THHH \rightarrow 3$ $HHTT \rightarrow 2$ $HTHT \rightarrow 2$ $THHT \rightarrow 2$
 $TTHH \rightarrow 2$ $THTH \rightarrow 2$ $HTTH \rightarrow 2$ $TTTH \rightarrow 1$
 $TTHT \rightarrow 1$ $THTT \rightarrow 1$ $HTTT \rightarrow 1$ $TTTT \rightarrow 0$.

- Now you have ranked every ω .
- When you assign a number to each outcome ω , you get a random variable.

Random Variable

- Let X = number of free shirts that you win
- Same as X = number of heads in 4 coin tosses

$HHHH \rightarrow 4$ $HHHT \rightarrow 3$ $HHTH \rightarrow 3$ $HTHH \rightarrow 3$
 $THHH \rightarrow 3$ $HHTT \rightarrow 2$ $HTHT \rightarrow 2$ $THHT \rightarrow 2$
 $TTHH \rightarrow 2$ $THTH \rightarrow 2$ $HTTH \rightarrow 2$ $TTTH \rightarrow 1$
 $TTHT \rightarrow 1$ $THTT \rightarrow 1$ $HTTT \rightarrow 1$ $TTTT \rightarrow 0$.

$$\{X = 2\} = \left\{ \begin{array}{ccc} HHTT, & HTHT, & THHT \\ TTHH, & THTH, & HTTH \end{array} \right\}$$

- $P(X=2) = 6/16$
- X is a variable that takes on random values between 0 and 4.

Density

- **Density** of a random variable X
 - Collection of random values that X takes
 - The corresponding probabilities

Values of X	0	1	2	3	4
Probabilities	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

**Density of the random variable
 X =Number of Heads in 4 tosses**

Random Variable

- **Experiment:** Roll a fair die twice and note the 2 faces that come up.

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

- Let X =sum of outcomes.
- $S_1 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Values of X	2	3	4	5	6	7	8	9	10	11	12
Probabilities	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Random Variable

- A random variable does 2 things
 1. partitions the sample space into disjoint events
 2. labels the elements of these events by appropriate numbers corresponding to their rankings
- If the labels are discrete/continuous, then the random variable is discrete/continuous.

$$\{X = 2\} = \{(1,1)\}$$

$$\{X = 3\} = \{(1,2), (2,1)\}$$

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

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

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

$$\{X = 7\} = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$$

$$\{X = 8\} = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$$

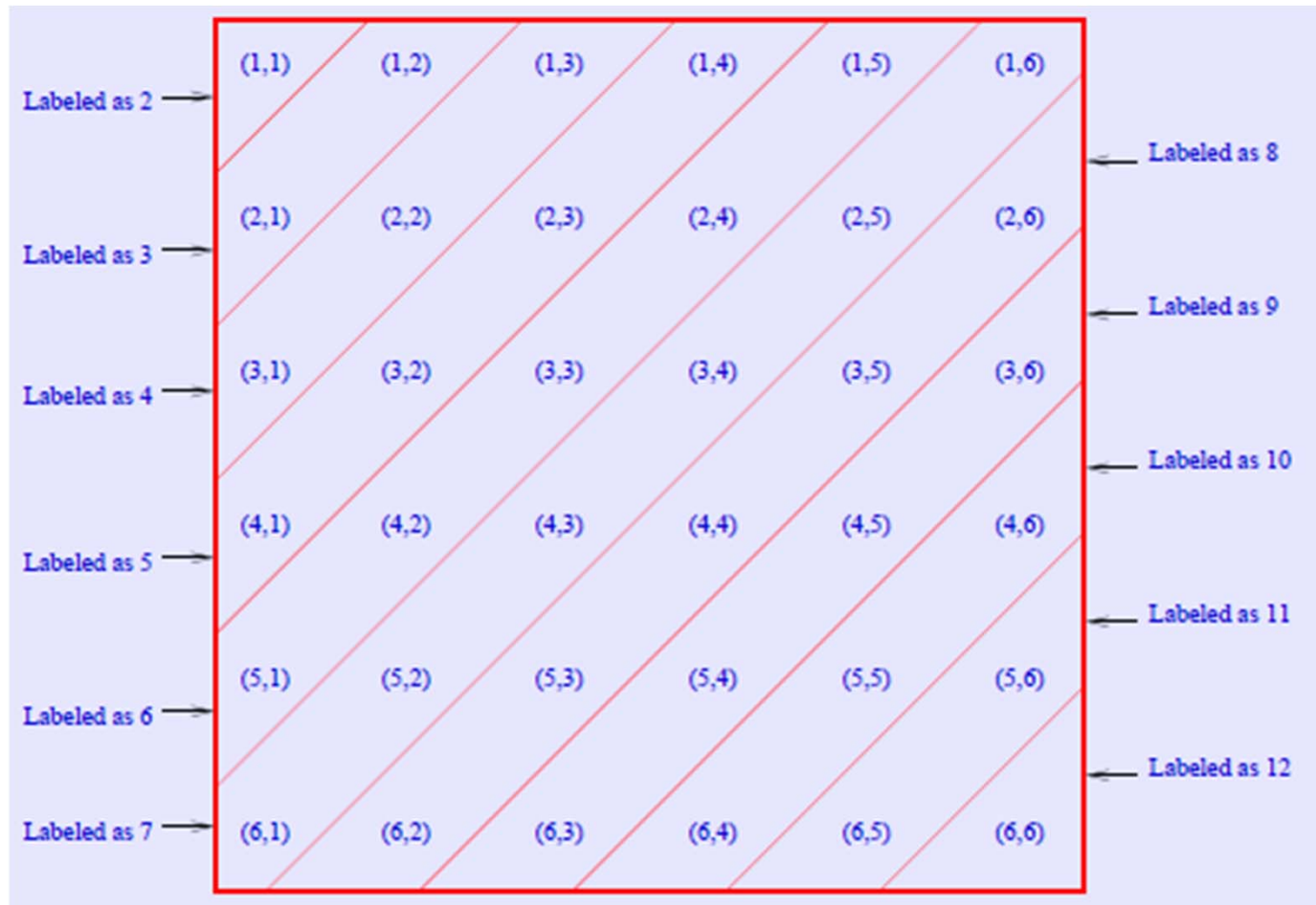
$$\{X = 9\} = \{(3,6), (4,5), (5,4), (6,3)\}$$

$$\{X = 10\} = \{(4,6), (5,5), (6,4)\}$$

$$\{X = 11\} = \{(5,6), (6,5)\}$$

$$\{X = 12\} = \{(6,6)\}$$

Random Variable



Partition of S induced by the random variable X

Random Variable

Values of X	2	3	4	5	6	7	8	9	10	11	12
Probabilities	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- Find $P(X \geq 9)$?
- Find $P(4 \leq X < 8)$?

SOME SPECIAL DISCRETE RANDOM VARIABLES

Binomial Random Variable

- We have already seen this one.
- When we
 1. count the number of successes
 2. in n independent tries
 3. where probability of success in any try is p
- Then $P(X=k \text{ successes in } n \text{ tries}) = C(n,k)p^k(1-p)^{n-k}$ where $k \in \Delta = \{0,1,2,\dots,n\}$.
- We say that X follows a Binomial distribution with parameters n and p.
 - $X \sim B(n,p)$

Geometric Random Variable

- When we
 1. count the number of failures **before** the first success
 2. where probability of success in any try is p
- Then $P(X=k \text{ failures before } 1^{\text{st}} \text{ success}) = p(1-p)^k$ where $k \in \Delta = \{0, 1, 2, \dots\}$.
- We say that X follows a Geometric distribution with parameter p .
 - $X \sim G(p)$

Shifted Geometric Random Variable

- When we
 1. count the number of trials to get k successes
 2. where probability of success in any try is p
- Then $P(X=k \text{ trials before } k^{\text{th}} \text{ success}) = p(1-p)^{k-1}$
where $k \in \Delta = \{1, 2, \dots\}$.
- We say that X follows a Shifted Geometric distribution with parameter p.

Poisson Random Variable

- For modeling
 1. Radio-active decay
 2. Telephone traffic, etc.
- $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ where $k \in \Delta = \{0, 1, 2, \dots\}$.
- We say that X follows a Poisson distribution with parameter λ .
 - $X \sim \text{Poisson}(\lambda)$
- Example: $P(X=2)$ can represent the probability of waiting 2 seconds before the next phone call.

Distribution Function

- The distribution function (or cumulative distribution function, cdf) of any random variable is $F(k) = P(X \leq k)$, $k \in \mathbb{R}$.
- For some random variables one can write $F(k)$ in a closed form, but for many we cannot do so.
- In general, the cdf of a discrete random variable has the following properties:
 1. $F(k)$ is a non-decreasing function of k ,
 2. $F(-\infty) = 0$, $F(\infty) = 1$,
 3. $F(k)$ is a right continuous function,
 4. $P(X = a) = F(a) - F(a^-)$, $a \in \Delta$.

Distribution Function vs. Density Function

- **Density** function gives probabilities of individual values $P(X=a)$.
- **Distribution** function gives cumulative probabilities $P(X \leq a)$.
- The terms density and distribution occur often in literature. You should know what means what.

Density	Distribution
$P(X=a)$	$P(X \leq a)$

Example

Three dice are rolled.

You select a number from the set $\{1, 2, 3, 4, 5, 6\}$, (say 4).

If no 4 comes on the three dice you loose 100 Rupees.

Otherwise you win $100 \times (\text{number of 4s})$

Suppose you decide to play this game once. What are your chances of (i) loosing 100 Rupees, (ii) winning 100 Rupees, (iii) winning 200 Rupees, (iv) winning 300 Rupees?

Let X be the random variable representing your net gain (or loss) after one game. What is the density of X ?

Example

$$S = \{(i, j, k) \mid i, j, k = 1, 2, \dots, 6\}.$$

$$|S| = 6 * 6 * 6 = 216$$

$$P(X=-100) = P(\text{no 4s}) = (5/6)^3,$$

$$P(X=100) = P(\text{one 4s}) = C(3,1)(1/6)(5/6)^2,$$

$$P(X=200) = P(\text{two 4s}) = C(3,2)(1/6)^2(5/6),$$

$$P(X=300) = P(\text{three 4s}) = (1/6)^3,$$

Practice Questions

- Toss a fair coin 4 times. Let X be the number of heads that you get. Write down the density of X ? Write down the distribution of X ?

Practice Questions

- Let S be the sample space representing the outcomes of three tosses of a weighted coin for which “H” is three times as likely as a “T”. Let X be a random variable that equals the largest number of successive heads in the outcome. For instance, $X(\text{HHT}) = 2$ and $X(\text{HTH}) = 1$. Find the density of X .

Practice Questions

- Which of the following tables represent a density of some random variable X . Explain your answers?

Values of X	-1	0	1
Probabilities	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{3}$

Values of X	0	1	2
Probabilities	$-\frac{1}{2}$	$\frac{3}{4}$	$\frac{3}{4}$

Values of X	-3	-2	-1
Probabilities	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Practice Questions

- What values of p will make the following table a density function of X ?

Values of X	-2	-1	0	1	2
Probabilities	0.10	0.20	$1 - p$	p	0.10

Practice Questions

- A random variable X has the following density,
 $P(X = k) = Ck^2$ for $k = 1, 2, 3, 4$.
 1. Find the constant C .
 2. Find the distribution of X .

Practice Questions

- A basket contains 10 marbles, four of which are red and the rest are black. Three marbles are pulled out all at once. Let X be the number of marbles selected that are red. If the basket is shaken fully before drawing the three marbles, find the density of X . (It is called a **hypergeometric density**.)

Practice Questions

- Five cards are dealt from a well shuffled standard deck of 52 cards. Let X be the random variable that counts the number of spade cards in the hand.
 1. Find the density of X .
 2. Now compute $P(X \geq 1)$.

♣=club (13 in total)

♦=diamond (13 in total)

♥=heart (13 in total)

♠=spade (13 in total)

Practice Questions

- A fair die is rolled 10 times. Let X be the number of times an even face occurred over those ten rolls. Find the density of X and evaluate the probability that X does not exceed 3.

Practice Questions

- An internet router receives data packets in a stream. Each packet can be either of two types — time sensitive (F) or time insensitive (S). About 20% of the packets are of time sensitive type. The stream forms a Bernoulli sequence of trials.
 - i. If X is the number of packets of type F in a segment of n packets, find $P(X = k)$.
 - ii. If Y is the number of packets of S type before the first packet of F type, find $P(Y = k)$.
 - iii. If Z is the number of packets observed to see the first packet of F type, find $P(Z = k)$.
 - iv. If W is the number of packets of S type before observing the 30-th packet of F type, find $P(W = k)$.
 - v. If T is the number packets to see the 30-th packet of F type, find $P(T = k)$.

Practice Questions

- Roll two fair dice (a red die and a green die)
Define a random variable, X , by $X((i, j)) = i^2 - j^3$,
where (i, j) is the outcome of the experiment,
 $i, j = 1, 2, \dots, 6$.
 - Find the density of X .
 - Find the distribution of X .

Practice Questions

- A box contains 20 mangos, six of which are rotten. A mango is selected at random, examined and then returned to the box. This is repeated three times.
 - i. What is the chance that exactly one rotten mango will be picked in the three draws?
 - ii. If X is the number of rotten mangoes drawn in the three draws, write the density of X .
 - iii. Obtain the cdf of X .

Practice Questions

- A box contains 20 mangos, six of which are rotten. Three mangos are selected in one draw.
 - i. What is the chance that there is exactly one rotten mango in the draw?
 - ii. If X is the number of rotten mangoes drawn, write the density of X .
 - iii. Obtain the cdf of X .

Summary

- Probability
 - Random Experiment, Outcome, Sample space, Event, $P(\text{event}) \rightarrow [0,1]$
- Independence
 - Joint=product of marginals ($P(AB)=P(A)P(B)$)
- Random Variables
 - Assign numbers to outcomes
 - Can be discrete (countable set) or continuous
- Discrete Random Variables
 - Binomial
 - Geometric
 - Poisson
 - Hyper-geometric
 - ...