Random Variable

Dr. Dinesh R. Helwade

Aegis
SCHOOL OF BUSINESS
SCHOOL OF DATA SCIENCE
SCHOOL OF TELECOMMUNICATION

Random Variable

A random variable *x* takes on a defined set of values with different probabilities.

- For example, if you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability one-sixth.

- For example, if you poll people about their voting preferences, the percentage of the sample that responds "Yes on Proposition 100" is a also a random variable (the percentage will be slightly differently every time you poll).

Random variables can be Discrete or Continuous

Discrete random variables have a countable number of outcomes

– Examples: Dead/alive, treatment effect, dice, counts, etc.

Continuous random variables have an infinite number of possible values in a given range.

– Examples: blood pressure, weight, the speed of a car, etc.

Probability functions

- A probability function maps the possible values of *x* against their respective probabilities of occurrence, *p(x)*

- *p(x)* is a number from 0 to 1.

- The area under a probability function is always 1.

Probability Mass Function

Definition:

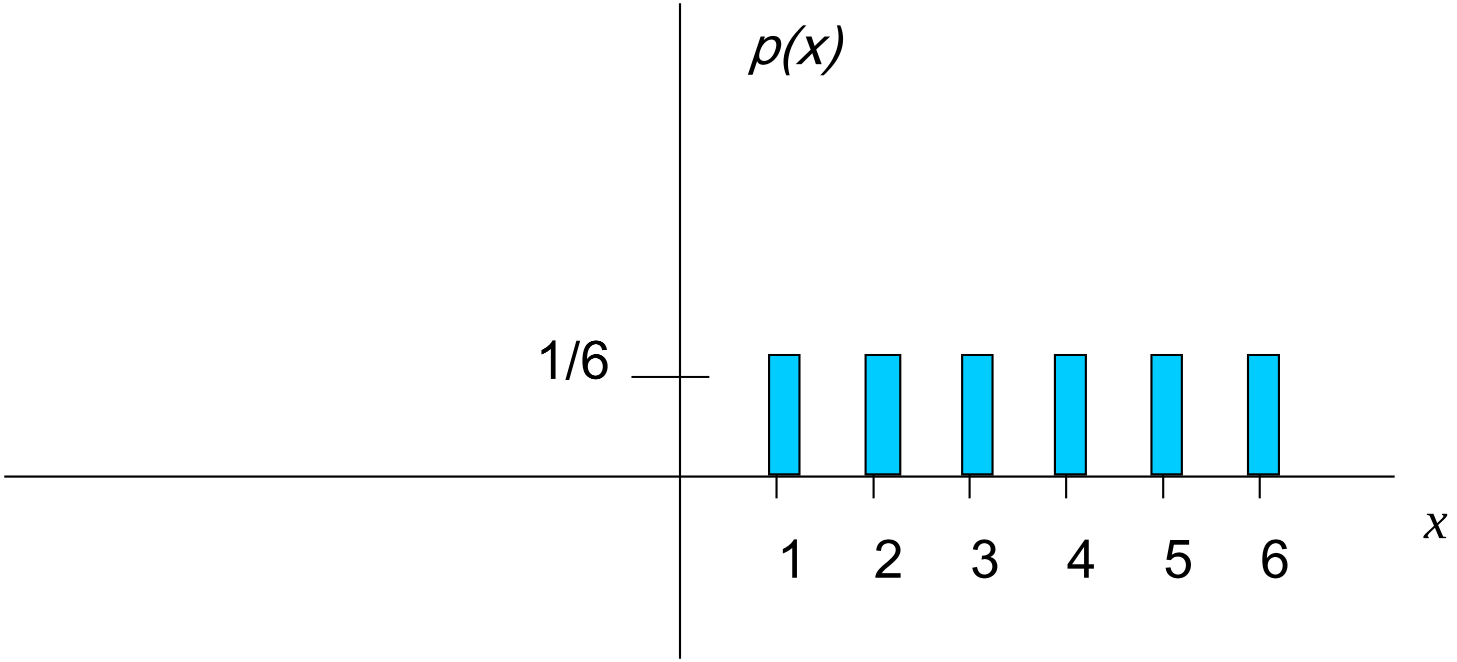
Let X be a discrete random variable with x_1, x_2 ,x_3 , ... Then the function

$$P_X(x) = P(X=x), \text{ for } x=1,2,3,...$$

is called the *probability mass function (PMF)* of X.

The **probability mass function** of a discrete random variable is simply the collection of all these probabilities.

Discrete example: roll of a die



$$\sum_{\text{all } x} P(x) = 1$$

Probability mass function (pmf)

x	*p(x)*
1	$p(x=1)=1/6$
2	$p(x=2)=1/6$
3	$p(x=3)=1/6$
4	$p(x=4)=1/6$
5	$p(x=5)=1/6$
6	$p(x=6)=1/6$

Total : 1.0

Probability Density Function

Definition:

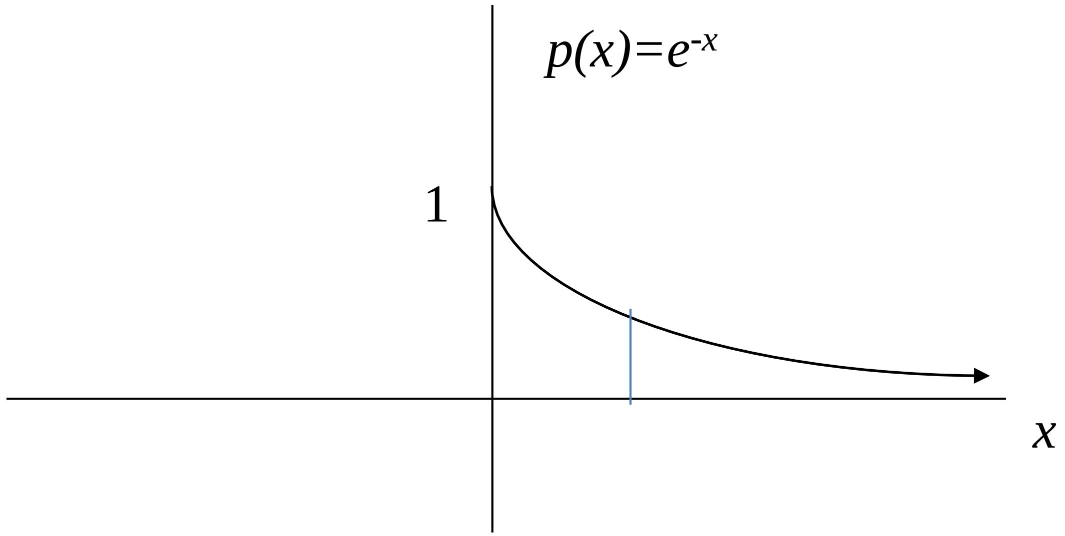
Let X be a continuous random variable, Then the function

$$P_X(a<x<b) = {}_a\int^b f_x\,(x)dx$$

is called the *probability density function (PDF)* of X.

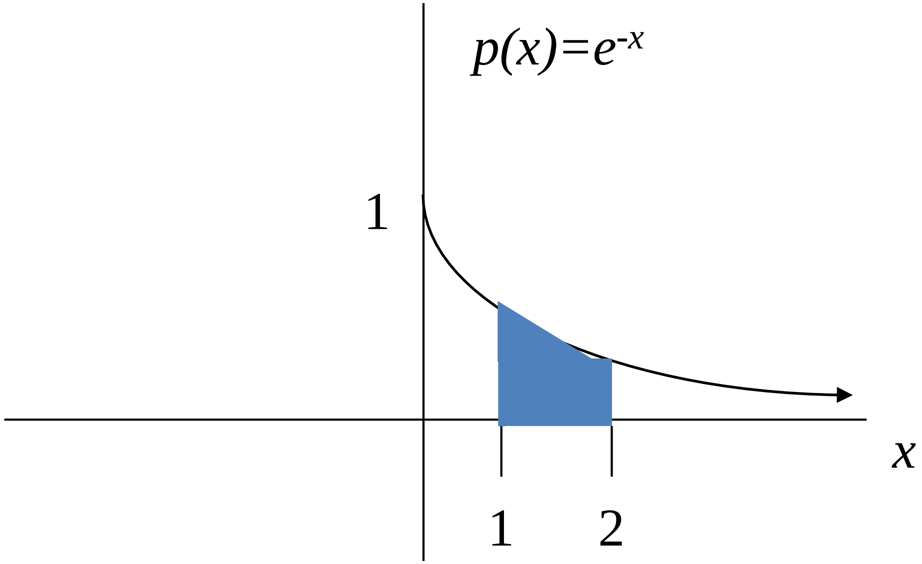
Most often, the equation used to describe a continuous probability distribution is called a **probability density function** .

Continuous case: “probability density function” (pdf)



The probability that *x* is any exact particular value (such as 1.9976) is 0; we can only assign probabilities to possible ranges of x.

For example, the probability of *x* falling within 1 to 2:



Clinical example: Survival times after lung transplant may roughly follow an exponential function.

Then, the probability that a patient will die in the second year after surgery (between years 1 and 2) is 23%.

$$P(1 \le x \le 2) = \int_1^2 e^{-x} = -e^{-x} \Big|_1^2 = -e^{-2} - -e^{-1} = -.135 + .368 = .23$$

Distribution Function

Definition:

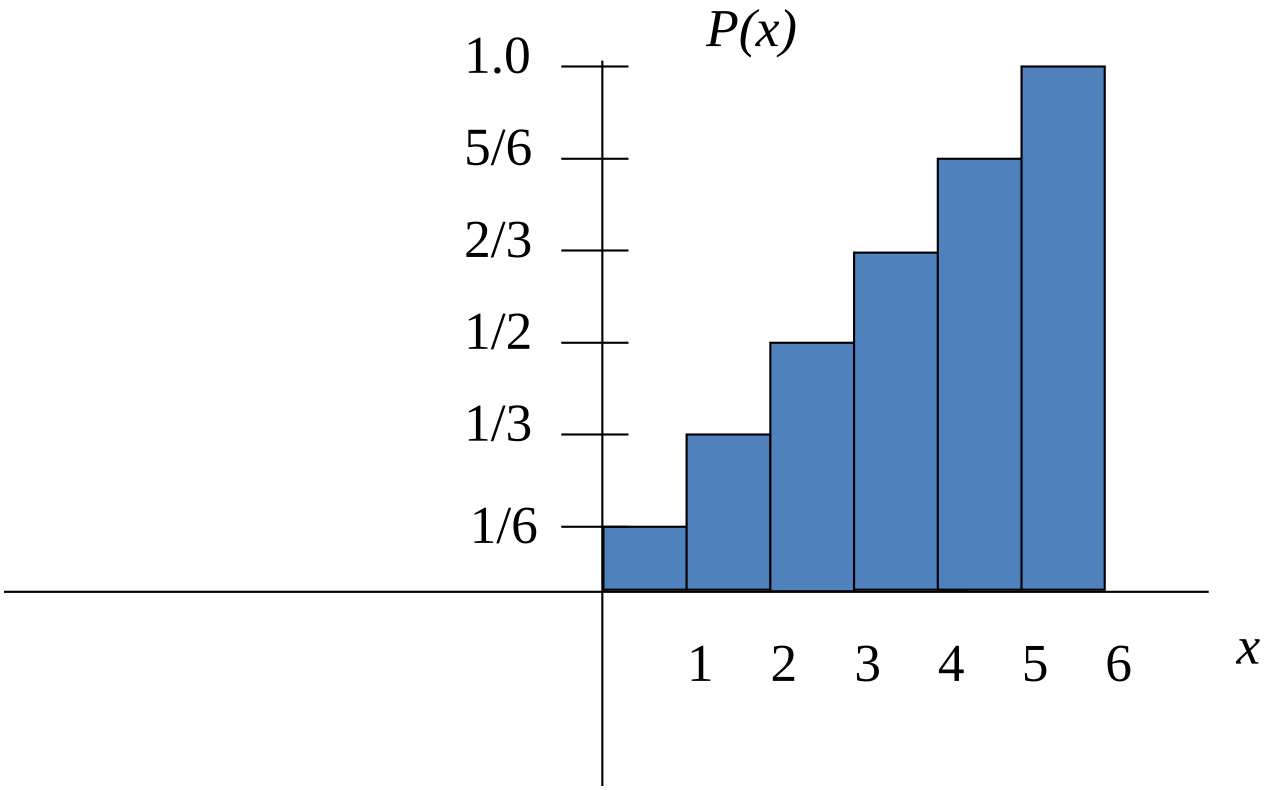
- The distribution function **F(x)**, also called the cumulative distribution function (CDF) or cumulative frequency function, describes the probability that a variate x takes on a value less than or equal to a number x.
- The distribution function is therefore related to a continuous probability density function by,

$$F(x) = P(X \leq x) \qquad \text{(discrete)}$$

$$F(x) = {}_{(-\text{inf})}\int^{x} f_x (x)\, dx \qquad \text{(Continuous)}$$

Cumulative distribution function (CDF)

Ex. Throwing a die experiment



Cumulative distribution function

x	$P(x \leq A)$
1	$P(x \leq 1) = 1/6$
2	$P(x \leq 2) = 2/6$
3	$P(x \leq 3) = 3/6$
4	$P(x \leq 4) = 4/6$
5	$P(x \leq 5) = 5/6$
6	$P(x \leq 6) = 6/6$

Practice Problem:

- The number of patients seen in the ER in any given hour is a random variable represented by x. The probability distribution for x is:

x	10	11	12	13	14
$P(x)$	.4	.2	.2	.1	.1

Find the probability that in a given hour:

a. exactly 14 patients arrive $p(x=14) = .1$

b. At least 12 patients arrive $p(x \geq 12) = (.2 + .1 + .1) = .4$

c. At most 11 patients arrive $p(x \leq 11) = (.4 + .2) = .6$

Aegis
SCHOOL OF BUSINESS
SCHOOL OF DATA SCIENCE
SCHOOL OF TELECOMMUNICATION

Review Question 1

If you toss a die, what's the probability that you roll a 3 or less?

a. 1/6
b. 1/3
c. 1/2
d. 5/6
e. 1.0

Review Question 1

If you toss a die, what's the probability that you roll a 3 or less?

a. 1/6
b. 1/3
c. 1/2
d. 5/6
e. 1.0

Review Question 2

Two dice are rolled and the sum of the face values is six. What is the probability that at least one of the dice came up a 3?

a. 1/5
b. 2/3
c. 1/2
d. 5/6
e. 1.0

Review Question 2

Two dice are rolled and the sum of the face values is six. What is the probability that at least one of the dice came up a 3?

a. 1/5
b. 2/3
c. 1/2
d. 5/6
e. 1.0

How can you get a 6 on two dice?
1-5, 5-1, 2-4, 4-2, 3-3

One of these five has a 3.

$\therefore 1/5$

Mathematical Expectation

Expected Value

- Expected value is an extremely useful concept for good decision-making!

Expected Value and Variance

- All probability distributions are characterized by an expected value (mean) and a variance (standard deviation squared).

Expected value of a random variable

- Expected value is just the average or mean (μ) of random variable *x*.

- It's sometimes called a "weighted average" because more frequent values of X are weighted more highly in the average.

- It's also how we expect X to behave on-average over the long run.

Expected value, formally

$$\boxed{E(X) = \sum_{\text{all } x} x_i \, p(x_i)}$$

Symbol Interlude

- $E(X) = \mu$
  - these symbols are used interchangeably

Example: expected value

- Recall the following probability distribution of ER arrivals:

x	10	11	12	13	14
P(x)	.4	.2	.2	.1	.1

$$\sum_{i=1}^{5} x_i\, p(x) = 10(.4) + 11(.2) + 12(.2) + 13(.1) + 14(.1) = 11.3$$

Example: the lottery

- The Lottery

- A certain lottery works by picking 6 numbers from 1 to 49. It costs Rs. 1 to play the lottery, and if you win, you win Rs. 20,00,000.

- *If you play the lottery once, what are your expected winnings or losses?*

Ex:Lottery

Calculate the probability of winning in 1 try:

$$\frac{1}{\binom{49}{6}} = \frac{1}{\frac{49!}{43!6!}} = \frac{1}{13{,}983{,}816} = 7.2 \times 10^{-8}$$

"49 choose 6"

Out of 49 numbers, this is the number of distinct combinations of 6.

The probability function (note, sums to 1.0):

X (in Rs.)	*p(x)*
-1	.999999928
20,00,000	7.2×10^{-8}

The probability function

X (in Rs.)	*p(x)*
-1	.999999928
20,00,000	7.2×10^{-8}

Expected Value

$$E(X) = 20{,}00{,}000 * 7.2 \times 10^{-8} + 0.999999928\,(-1)$$
$$= .144 - .999999928 = -0.855999928$$

Negative expected value is never good!
You shouldn't play if you expect to lose money!

Expected Value

If you play the lottery every week for 10 years, what are your expected winnings or losses?

520 x (- *0.855999928*) = - 445.12 Rs.

<u>Gambling</u> (how casinos can afford to give so many free drink !!!)

A roulette wheel has the numbers 1 through 36, as well as 0 and 00. If you bet \$1 that an odd number comes up, you win or lose \$1 according to whether or not that event occurs. If random variable X denotes your net gain, X=1 with probability 18/38 and X= -1 with probability 20/38.

$E(X) = 1(18/38) - 1\,(20/38) = -\$.053$

On average, the casino wins (and the player loses) 5 cents per game.

The casino rakes in even more if the stakes are higher:

$E(X) = 10(18/38) - 10\,(20/38) = -\$.53$

If the cost is \$10 per game, the casino wins an average of 53 cents per game. If 10,000 games are played in a night, that's a cool \$5300.

Variance/standard deviation

σ^2=Var(x) =E(x-μ)^2

"The expected (or average) squared distance (or deviation) from the mean"

$$\boxed{\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)}$$

Variance

Discrete case:

$$Var(X) = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

Continuous case?:

$$Var(X) = \int_{\text{all } x} (x_i - \mu)^2 p(x_i) dx$$

Symbol Interlude

- $Var(X) = \sigma^2$
- $SD(X) = \sigma$
  - these symbols are used interchangeably

Variance

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

$$\sigma^2 = \sum_{all\ x} (x_i - \mu)^2 p(x_i)$$

$$= (1 - 200{,}000)^2(.5) + (400{,}000 - 200{,}000)^2(.5)$$

$$= 200{,}000^2$$

$$\sigma = \sqrt{200{,}000^2} = 200{,}000$$

Now you examine your personal risk tolerance…

Review Question 3

The expected value and variance of a coin toss (H=1, T=0) are?

a. .50, .50
b. .50, .25
c. .25, .50
d. .25, .25

Review Question 3

The expected value and variance of a coin toss (H=1, T=0) are?

a. .50, .50
b. .50, .25
c. .25, .50
d. .25, .25

Theoretical Distribution

Discrete Distributions

1. Bernoulli Distribution
2. Binomial Distribution
3. Poisson Distribution

Continuous distribution

1. Normal Distribution

Criteria for a Binomial Probability Experiment

An experiment is said to be a **binomial experiment** provided

1. The experiment is performed a fixed number of times. Each repetition of the experiment is called a **trial**.

2. The trials are independent. This means the outcome of one trial will not affect the outcome of the other trials.

3. For each trial, there are two mutually exclusive outcomes, success or failure.

4. The probability of success is fixed for each trial of the experiment.

Notation Used in the Binomial Probability Distribution

- There are n independent trials of the experiment
- Let p denote the probability of success so that $1 - p$ is the probability of failure.
- Let x denote the number of successes in n independent trials of the experiment. So, $0 \leq x \leq n$.

EXAMPLE *Identifying Binomial Experiments*

Which of the following are binomial experiments?

(a) A player rolls a pair of fair die 10 times. The number *X* of 7's rolled is recorded.

(b) The 11 largest airlines had an on-time percentage of 84.7% in November, 2001 according to the Air Travel Consumer Report. In order to assess reasons for delays, an official with the FAA randomly selects flights until she finds 10 that were not on time. The number of flights *X* that need to be selected is recorded.

(c) In a class of 30 students, 55% are female. The instructor randomly selects 4 students. The number *X* of females selected is recorded.

Binomial Probability Distribution Function

The probability of obtaining *x* successes in *n* independent trials of a binomial experiment where the probability of success is *p* is given by

$$P(X = x) = {}_nC_x \; p^x (1-p)^{n-x}$$

$x = 0, 1, \ldots, n$

where *p* = probability of success

Binomial distribution, generally

if you have only two possible outcomes (call them 1/0 or yes/no or success/failure) in *n* independent trials, then the probability of exactly *X* "successes" is given by,

n = number of trials

$$P(X = x) = \binom{n}{X} p^X (1-p)^{n-X}$$

X = # successes out of *n* trials

p = probability of success

1-p = probability of failure

Aegis
SCHOOL OF BUSINESS
SCHOOL OF DATA SCIENCE
SCHOOL OF TELECOMMUNICATION

Mean and Standard Deviation of a Binomial Random Variable

A binomial experiment with *n* independent trials and probability of success *p* will have a mean and standard deviation given by the formulas

$$\mu = np \qquad \text{and} \qquad \sigma = \sqrt{np(1-p)}$$

Phrase	***Math Symbol***
“at least”	$\geq$
“more than” or “greater than”	$>$
“fewer than” or “less than”	$<$
“no more than”	$\leq$
“exactly”	$=$

Review Question 1

In your case-control study of smoking and lung-cancer, 60% of cases are smokers versus only 10% of controls. What is the odds ratio between smoking and lung cancer?

a. 2.5
b. 13.5
c. 15.0
d. 6.0
e. .05

Review Question 1

In your case-control study of smoking and lung-cancer, 60% of cases are smokers versus only 10% of controls. What is the odds ratio between smoking and lung cancer?

a. 2.5
b. 13.5
c. 15.0
d. 6.0
e. .05

$$\frac{\frac{.6}{.4}}{\frac{.1}{.9}} = \frac{3}{2} x \frac{9}{1} = \frac{27}{2} = 13.5$$

Review Question 2

What's the probability of getting exactly 5 heads in 10 coin tosses?

a. $\binom{10}{0}(.50)^5(.50)^5$

b. $\binom{10}{5}(.50)^5(.50)^5$

c. $\binom{10}{5}(.50)^{10}(.50)^5$

d. $\binom{10}{10}(.50)^{10}(.50)^0$

Review Question 2

What's the probability of getting exactly 5 heads in 10 coin tosses?

a. $\binom{10}{0}(.50)^5(.50)^5$

b. $\binom{10}{5}(.50)^5(.50)^5$

c. $\binom{10}{5}(.50)^{10}(.50)^5$

d. $\binom{10}{10}(.50)^{10}(.50)^0$

Review Question 3

A coin toss can be thought of as an example of a binomial distribution with N=1 and p=.5. What are the expected value and variance of a coin toss?

a. .5, .25
b. 1.0, 1.0
c. 1.5, .5
d. .25, .5
e. .5, .5

Review Question 3

A coin toss can be thought of as an example of a binomial distribution with N=1 and p=.5. What are the expected value and variance of a coin toss?

a. .5, .25
b. 1.0, 1.0
c. 1.5, .5
d. .25, .5
e. .5, .5

Review Question 4

If I toss a coin 10 times, what is the expected value and variance of the number of heads?

a. 5, 5
b. 10, 5
c. 2.5, 5
d. 5, 2.5
e. 2.5, 10

Review Question 4

If I toss a coin 10 times, what is the expected value and variance of the number of heads?

a. 5, 5
b. 10, 5
c. 2.5, 5
d. 5, 2.5
e. 2.5, 10

Poisson Distribution

A random variable *X*, the number of successes in a fixed interval, follows a **Poisson process** provided the following conditions are met

1. The probability of two or more successes in any sufficiently small subinterval is 0.

2. The probability of success is the same for any two intervals of equal length.

3. The number of successes in any interval is independent of the number of successes in any other interval provided the intervals are not overlapping.

Poisson Probability Distribution Function

If X is the number of successes in an interval of fixed length t, then the probability formula for X is

$$P(X = x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} \qquad x = 0, 1, 2, 3, \ldots$$

where λ represents the average number of occurrences of the event in some interval of length 1.