Random Variables and Mathematical Expectations

(Discrete and Continuous)

Lesson-01

Dr. Tahseen A. Jilani
Postdoc (UoN - UK), PhD (UoK, HEC Indigenous Scholars Program),
MSc(Statistics), MA(Economics), BSc(CS)

Associate Professor, Department of Computer Science, University of Karachi

Brief introduction

- Teaching since March-2002 in CS/Stats in the UoK.
- Formally, the first PhD from DCS-UoK faculty, First HEC Indigenous faculty Scholar 2003-2007 (completed in 3.5 years).
- Post doctoral research funded by University of Nottingham (World QS ranking 74th position).
- Worked with more than 15 universities from UK, Canada, USA, China and top companies.
- Accepted by five universities based on academic and research performance.
- More than 65 research publications with total impact factor of 100+.

Motivation: Your hard work and positive attitude will always payoff.

CLASS RULES

- **DISCIPLINE.**
- **Give your 100% when you are in the class.**
- Classes are recorded for quality control and students' measures.
- Q&A is allowed during the lessons but for discussion, we can have time after the lesson hours.
- Any technical issues (electricity, internet, health, day to day) can be up to 20% of the total time.

Week-1: Fundamental Concepts

- Random variables (Discrete and continuous)
- Probability distributions (Discrete and continuous)
- Mean and variance
- Mathematical Moments

Random Variable

- An experiment with more than one possible outcome is termed as random experiment. Other names are **statistical**, **probabilistic** or **uncertain** experiments. They are **one-to-many** type experiments.
- A variable that expresses the results of a random experiment is called random variable. For example coin tossing, dice rolling
- A random variable *x* takes on a defined set of values with different probabilities e.g. P(head)= 0.5 and P(tail) = 0.5.
- A random variable is a numerical quantity with associated probability. For example say head: X= 1 and tail: X = 0 in a single toss of a coin.

Question: Think of other possible examples of random experiments? State any five such experiments and what will be their possible outcomes.

Examples

- 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" is also a random variable (the percentage will be slightly differently every time you poll).

- For example, Pakistan will win the Cricket World cup next time? Some experiments have equal probabilities for all possible outcomes and some don't have.

- Device battery charge decreasing over time (continuous).

Random variables can be discrete or continuous

- **Discrete** random variables have a countable number of outcomes.
  - Examples: survived/ not-survived, dice rolling, profit/loss, computer virus attack (yes/no), number of cars in vehicular network cloud.
  - We can list all possible values for discrete experiments.

- **Continuous** random variables have an infinite continuum of possible values.
  - Examples: blood pressure, weight, the speed of a car, real numbers from 1 to 6, daily share price, revenue/profit, claim-amounts in insurance, compound growth.
  - We need to express it with formula/ equation/ graph.

Probability functions

- The probability distribution provides a probability of occurrence for each possible value of a random variable.

Outcome of die roll	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6

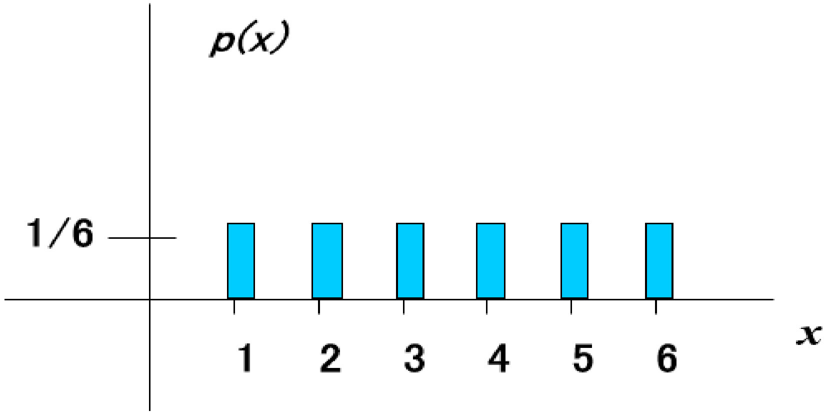
- A probability function maps the possible values of *x* against their respective probabilities of occurrence, P(X=x)

X (sum)	2	3	4	5	6	7	8	9	10	11	12
Outcomes	1	2	3	4	5	6	5	4	3	2	1
$P(X)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Discrete Probability Distribution (PMF)

- A table or formula listing all possible values that a discrete random variable can take on, together with the assumed probabilities is called a discrete probability distribution.

Outcome of die roll	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6



- A discrete probability distribution is called "Probability Mass Function" (PMF).
- A function is called pmf if it satisfies the following two conditions.

$$\mathbf{0} \leq \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}) \leq \mathbf{1} \text{ and } \sum_{i=1}^{n} \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}_i) = \mathbf{1}.$$

Exercise for Probability mass function (pmf)

1. $f(y) = \left(\frac{1}{2}\right)^y$; $y = 1,2,3$

2. $f(x) = \frac{6-|x-7|}{36}$; $for x = 2,3,4,\ldots,12$

3. $f(x) = \frac{x+2}{25}$; $for\ x = 1,2,3,4,5$

4. $f(x) = \frac{\binom{2}{x}\binom{4}{3-x}}{\binom{6}{3}}$; $for x = 0,1,2.$

5. $f(x) = \binom{5}{x}\left(\frac{1}{5}\right)^x\left(\frac{4}{5}\right)^{5-x}$; $for x = 0,1,\ldots,5$

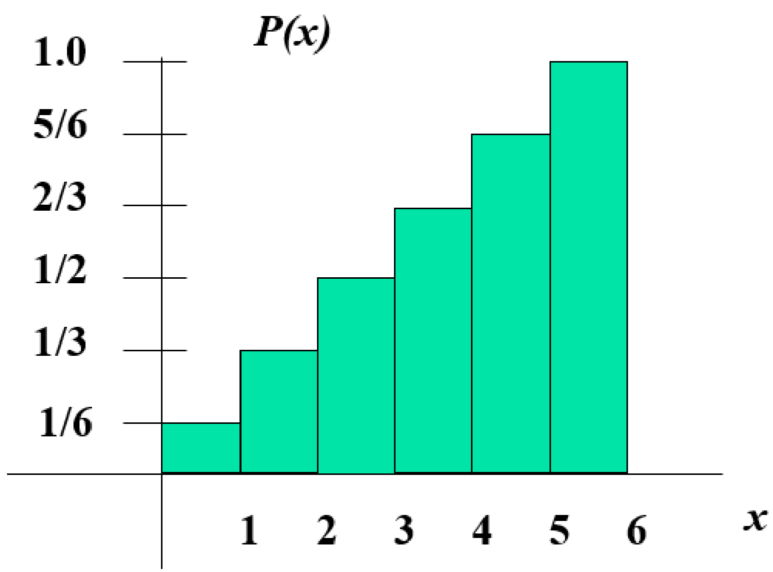
Dear Students:
For each of these, create a table and check that probabilities are between 0 and 1 and sum of probabilities =1. Also create a bar plot for visualisation.

Cumulative density function / Distribution function

- The main concept is same as cumulative frequency.
- Let X be a discrete random variable with density f(x). The cdf. denoted by F(x) is defined by

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

x	*P(x≤A)*
1	P(x≤1) = 1/6
2	P(x≤2) = 2/6
3	P(x≤3) = 3/6
4	P(x≤4) = 4/6
5	P(x≤5) = 5/6
6	P(x≤6) = 6/6



Example: Computing probabilities using pmf

- The number of patients seen in the Emergency department in any given hour is a random variable represented by *x*. The probability distribution (PMF) for *x* is:

x	**10**	**11**	**12**	**13**	**14**
P(x)	**0.4**	**0.2**	**0.2**	**0.1**	**0.1**

Find the probability that in a given hour:

a. **exactly** 14 patients arrive **$P(x = 14) = 0.1$**

b. **At least** 12 patients arrive **$P(x \geq 12) = (0.2 + 0.1 + 0.1) = 0.4$**

c. **At most** 11 patients arrive **$P(x \leq 11) = (0.4 + 0.2) = 0.6$**

Continuous probability distribution/pdf

- The probability function that accompanies a continuous random variable is a probability density function. For
- For f(x) to be pdf, have to check two conditions.
  - $f(x) \geq 0$, for all values of the domain set X.
  - $\int_L^U f(x)dx = 1$. Here L= lower limit and U = upper limit.

- For example, consider the following exponential function:

$$f(x) = e^{-x}, x > 0$$

Clearly, $f(x) \geq 0$, for all values of x and area under the curve is

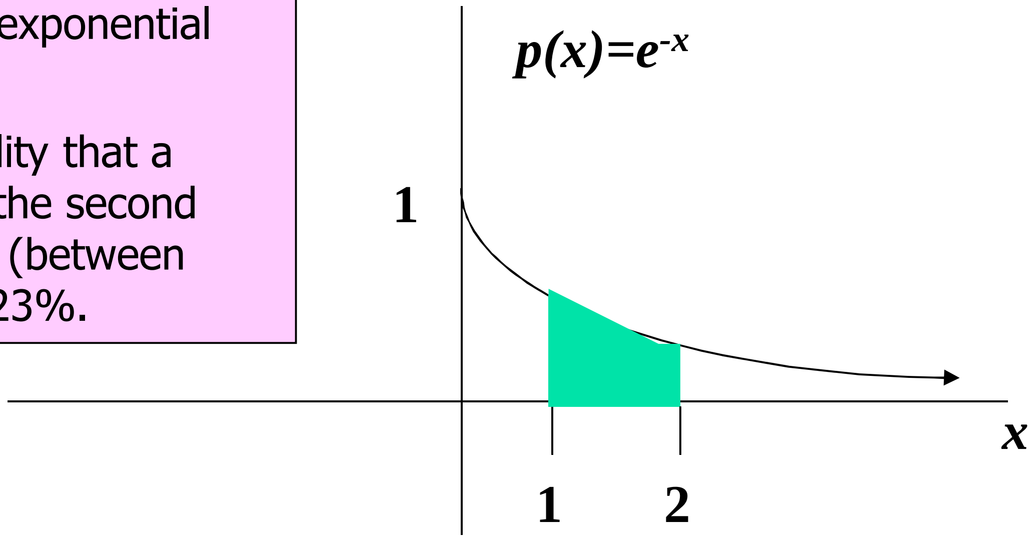
$$\int_0^{+\infty} e^{-x} = -e^{-x} \Big|_0^{+\infty} = 0+1=1$$

Example – calculating probability using a pdf

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%.



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

Example

- Consider the continuous random variable X with pdf.

$$f(x) = \frac{3}{x^4}\ , x > 1$$

- Clearly if we put different values of x>1, we have $f(x) > 0$. Alternatively, plot this function and the values should be in above x-axis.
- $\int_{x=1}^{\infty} \frac{3}{x^4}\, dx = 3 \int_{x=1}^{\infty} x^{-4}\, dx \ = 3 \left| \frac{x^{-4+1}}{-4+1} \right|_{x=1}^{\infty} = \frac{3}{-3}\left(\frac{1}{\infty^3} - \frac{1}{1^3}\right) = 1$

Median value of a pdf

- Median is the most middle value that have $P(X \leq M) = P(X \geq M) = 0.5$ where M is the median point/value.

Example

$$\int_{x=1}^{M} \frac{3}{x^4}\, dx = 0.5 \;\Rightarrow\; 3\left|\frac{x^{-4+1}}{-4+1}\right|_1^M = 0.5 \;\Rightarrow\; -(M^{-3} - 1) = 0.5$$

$$-(M^{-3} - 1) = 0.5 \;\Rightarrow\; 1 - \frac{1}{M^3} = 0.5 \;\Rightarrow\; M^3 = \frac{1}{0.5}$$

$$\boldsymbol{M = Median\ for\ f(x) = \sqrt[3]{2}}$$

Example – Piecewise function

Consider a piece-wise function. Is this function a pdf?

$$f(x) = \begin{cases} x & ; for\, 0 < x < 1 \\ 2 - x & ; for\, 1 \le x 2 \\ 0 & ; elsewhere \end{cases}$$

$$\int_0^2 f(x)\, dx = \int_0^1 f_1(x)\, dx + \int_1^2 f_2(x) dx = \int_0^1 (x)\, dx + \int_1^2 (2 - x) dx$$

$$= \left| \frac{x^2}{2} \right|_0^1 + \left| 2x - \frac{x^2}{2} \right|_1^2 = \frac{1}{2} + \left(2 - \frac{4-1}{2}\right) = \frac{1}{2} + \left(2 - \frac{3}{2}\right) = \frac{1}{2} + \frac{1}{2} = 1$$

Expected value (mean) of a random variable

- All probability distributions are characterized by an expected value (mean) and a variance.
- Examples:
  - Average time spent by each patient in an OPD.
  - Average number of vehicles entering UoK entrance gate.
  - Variance in daily expenses by an first year student during semester etc.
  - Weekly average sales by Food Panda in some specific area.

Time

Space

- Expected value is just the average or mean (μ) of random variable *x*. You learnt it as arithmetic mean in BSCS-305.

- Expected value is an extremely useful concept for good decision-making! e.g. Portfolio analysis and portfolio selection.

Expected value (mean) and variance

Discrete case:

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

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

Continuous case:

$$E(X) = \int_{\text{all } x} x_i \, p(x_i) dx$$

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

Example-1: Mean and Variance for a pmf.

- Recall the following probability distribution of hospital emergency department arrivals:

X=x	10	11	12	13	14	Sum
P(X=x)	0.4	0.2	0.2	0.1	0.1	**1**
x . P(x)	4	2.2	2.4	1.3	1.4	**11.3**
x^2 . P(x)	40	24.2	28.8	16.9	19.6	**129.5**

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

- Here mean =11.3, and
- Variance = $E(X^2) - [E(X)]^2$ = 129.5 – 11.3^2 = 1.81

Exmaple-2: Mean and variance for a pmf

The expected value (mean) and <u>variance</u> of a coin toss (let head=H= 1, tail = T=0) are?

$$E(X) = 0 \times 0.5 + 1 \times 0.5 = 0.5$$

$$E(X^2) = 0^2 \times 0.5 + 1^2 \times 0.5 = 0.5$$

$$V(X) = E(X^2) - [E(X)]^2 = 0.5 - 0.5^2 = 0.25$$

Example- Mean and variance for a pdf

- Consider the pdf (we have proved it's pdf in earlier slides).

$$f(x) = \frac{3}{x^4}\ , x > 1$$

- Mean $= \mathrm{E}(x) = \int_{x=1}^{\infty} x.\frac{3}{x^4}\,dx = 3\int_{x=1}^{\infty} x^{-3}\,dx$

$$= 3.\left|\frac{x^{-3+1}}{-3+1}\right|_{x=1}^{\infty} = \frac{3}{-2}\left(\frac{1}{\infty^2} - \frac{1}{1^2}\right) = \frac{3}{2}$$

- Variance $=\mathrm{E}(x^2) - [E(x)]^2 = \int_{x=1}^{\infty} x^2.\frac{3}{x^4}\,dx - \left(\frac{3}{2}\right)^2$

$$= 3\int_{x=1}^{\infty} x^{-2}\,dx - \left(\frac{3}{2}\right)^2 = 3 - \frac{9}{4} = \frac{3}{4}$$

Thus mean = 3/2 and variance = 3/4

Mean/ Variance for a piecewise pdf

Consider the piecewise pdf:

$$f(x) = \begin{cases} x & ; for\, 0 < x < 1 \\ 2 - x & ; for\, 1 \leq x2 \\ 0 & ; elsewhere \end{cases}$$

$$Mean = \quad E(x) = \int_0^1 x.f_1(x)\,dx + \int_1^2 x.f_2(x)dx$$

$$= \int_0^1 x.x\,dx + \int_1^2 x.(2-x)dx = \left|\frac{x^3}{3}\right|_0^1 + \left|\frac{2x^2}{2} - \frac{x^3}{3}\right|_1^2 = \frac{1}{3} + 3 - \left(\frac{7}{3}\right) = 1$$

$$E(x^2) = \int_0^1 x^2.x\,dx + \int_1^2 x^2.(2-x)dx = \left|\frac{x^4}{4}\right|_0^1 + \left|\frac{2x^3}{3} - \frac{x^4}{4}\right|_1^2$$

$$= \frac{1}{4} + \left(\frac{14}{3}\right) - \left(\frac{15}{4}\right)$$

$$= 1.1667$$

Therefore, Variance = $E(x^2)$ - $[E(x)]^2 = 1.16667 - 1^2 = 0.1667$

Mathematical Moments

- Higher order averages are termed as mathematical models.
- The mean E(x) and variance V(x) are examples of mathematical moments. There are mainly two types of moments. Here f(x) is the pdf and P(X=x) represents the pmf.

Raw moments (moments about origin):

- Discrete cases : $\mu^k(0) = E\left(x^k\right) = \sum_{x=l}^{x=u} x^k . f(x)$
- Continuous case: $\mu^k(0) = E\left(x^k\right) = \int_l^U x^k . f(x) dx$

Central moments (moments about mean):

- Discrete cases : $\mu^k(\mu) = \sum_{x=l}^{x=u} (x-\mu)^k . f(x)$
- Continuous case: $\mu^k(\mu) = \int_l^U (x-\mu)^k . f(x)\, dx$

Example- Mathematical Moments – Discrete

- Recall the emergency response (ER) department pmf.

X=x	10	11	12	13	14	Sum
P(X=x)	0.4	0.2	0.2	0.1	0.1	**1**
x.P(x)	4	2.2	2.4	1.3	1.4	**11.3**
x2.P(x)	40	24.2	28.8	16.9	19.6	**129.5**
x3.P(x)	400	266.2	345.6	219.7	274.4	**1505.9**
x4.P(x)	4000	2928.2	4147.2	2856.1	3841.6	**17773.1**

Results for First four raw moments are

- $E(X) = 11.3$
- $E(X^2) = 129.5$
- $E(X^3) = 1505.9$
- $E(X^4) = 17773.1$

Example- Mathematical Moments – Continuous

- $\mu^k(0) = E(x^k) = \int_{x=1}^{\infty} x^k . f(x)\ dx$

$$= \int_{x=1}^{\infty} x^k . \frac{3}{x^4} dx = 3. \int_{x=1}^{\infty} x^{k-4}\, dx$$

$$= 3. \left| \frac{x^{k-4+1}}{k-4+1} \right|_{x=1}^{\infty} = \frac{3}{k-3} \left(\frac{1}{\infty^{k-3}} - \frac{1}{1^{k-3}} \right)$$

$$\boldsymbol{E(x^k) = \frac{-3}{k-3}, \qquad for\ k = 1, 2, 3, 4}$$

$\boldsymbol{Therefore,}$

$\boldsymbol{Mean = E(x) = \frac{3}{2}, E(x^2) = 3;\ E(x^3) = undefined; E(x^4) = -3;}$

Assignment-01 (Please submit before 30th August-2021)

- Compute the mean and variance for the following p.m.f.

1. $f(y)=\left(\frac{1}{2}\right)^{y}\ ;\ y=1,2,3$

2. $f(x)=\frac{6-|x-7|}{36}\ ;\ for\,x=2,3,4,\ldots,12$

3. $f(x)=\frac{x+2}{25};\ for\,x=1,2,3,4,5$

4. $f(x)=\frac{\binom{2}{x}\binom{4}{3-x}}{\binom{6}{3}}\ ;\ for\,x=0,1,2.$

- Compute the mean and variance for the following p.d.f.

$$g(x)=\begin{cases}6x(1-x) & for\ \ 0<x<1\\ 0 & otherwise\end{cases}$$

$$f(x)=\begin{cases}x & ;\ for\,0<x<1\\ 2-x & ;\ for\,1\leq x2\\ 0 & ;elsewhere\end{cases}$$

Review Questions (not to submit)

Q1. If you toss a die, what's the probability that you (a). roll a 3 or less? (b). Roll at least 2 (c). Roll at most 5. (d). Between 2 and 5 inclusive. (e). Between 2 and 5.

Q2. 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?

Q3. 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?

HINT: 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.

Discuss with students (5-8min):

- **Recorded lessons**: You can watch at your own pace.
  - We will have weekly question-Answer session at the lecture time. Video will be played and you can ask question if you have difficulty.
  - You can ask question(s) at later hours. Please try to ask between 9am to 5pm. To avoid holidays and keep lesson on time.
- **Live lessons:**
  - With large class sized, it could take 10/15 min setup time. Save this time by watching lessons at your own leisure.
  - For attendance purpose we will have lesson on official time as well (see above for details).

ASSIGNMENT Deadlines will remain unchanged in both cases.

End of lesson-01

Thank you