RV: CUMULATIVE DISTRIBUTION FUNCTION

Dr. Aleksandra Kaszubowska-Anandarajah

anandara@tcd.ie

Some material has been adopted from
the Pennsylvania State University online course STAT 414

Previous lecture

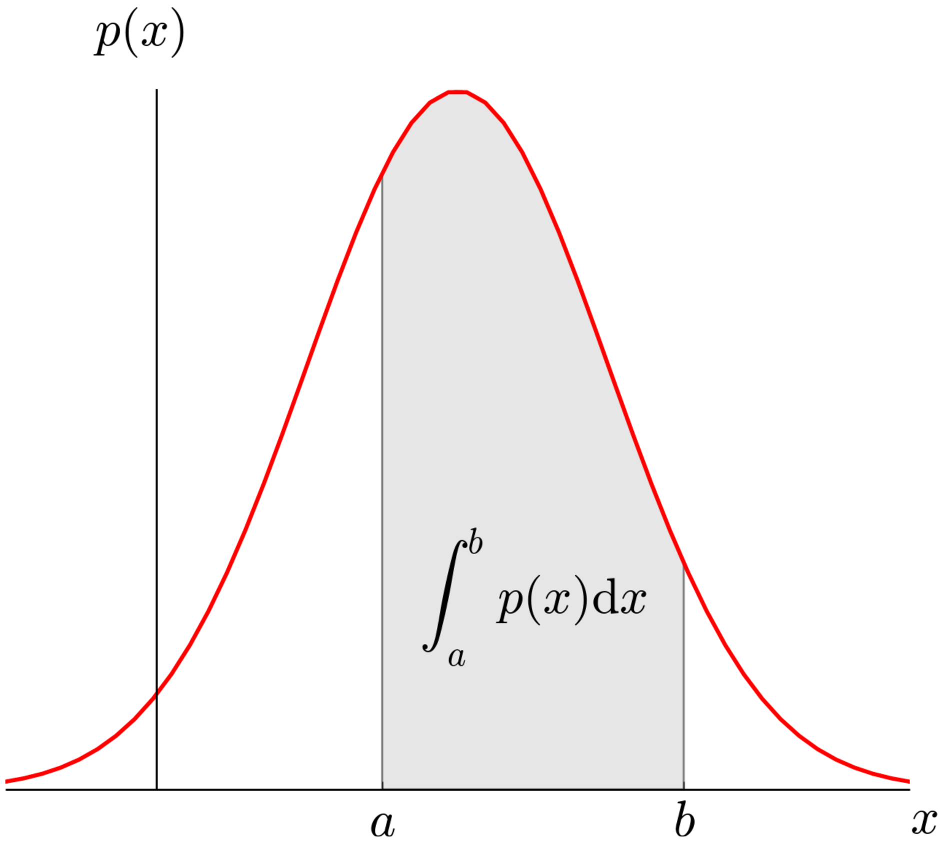
- Law of total probability
- Bayes theorem
- Statistical independence
- Random variable
- Probability mass function

Today's lecture

- Probability density function
- Cumulative distribution function
- Expected value and its properties

Probability Density Function (PDF)

- For <u>continues RV</u> the probability that X=x is zero.
- Instead we seek the probability that X falls within an interval a < X < b



- The probability that the continuous RV X, is between a and b corresponds to the area under the curve representing the probability density function between the limits a and b.
- For discrete RV it was a sum off $\sum p(X=x_i)$

PDF

Properties of PDF

PDF is an integrable function with the following properties:

1. f(x) > 0, for all x in S
2. The area under the curve f(x) in the support S is 1, that is: $\int_S f(x)dx = 1$
3. If f(x) is the PDF of x, then the probability that x belongs to some interval, is given by the integral of f(x) over that interval, that is:

$$\mathrm{P}(\mathrm{x} \in A) = \int_A f(x)dx$$

PDF - Example

Let $f(x) = 3x^2$ for $0 < x < 1$

1. Verify that $f(x)$ is a valid PDF

2. Calculate $P(1/2<X<1)$ & $P(X=1/2)$.

Note: $f(x)$ is not a probability => e.g.: for $x=0.7$ => $f(x)=1.47$ => $f(x)>1$!

PDF - Example

Solution:

1.

a) Is f(x) nonnegative for all $x \in S$ i.e. $0<x<1$?

Answer: $3x^2$ is always positive

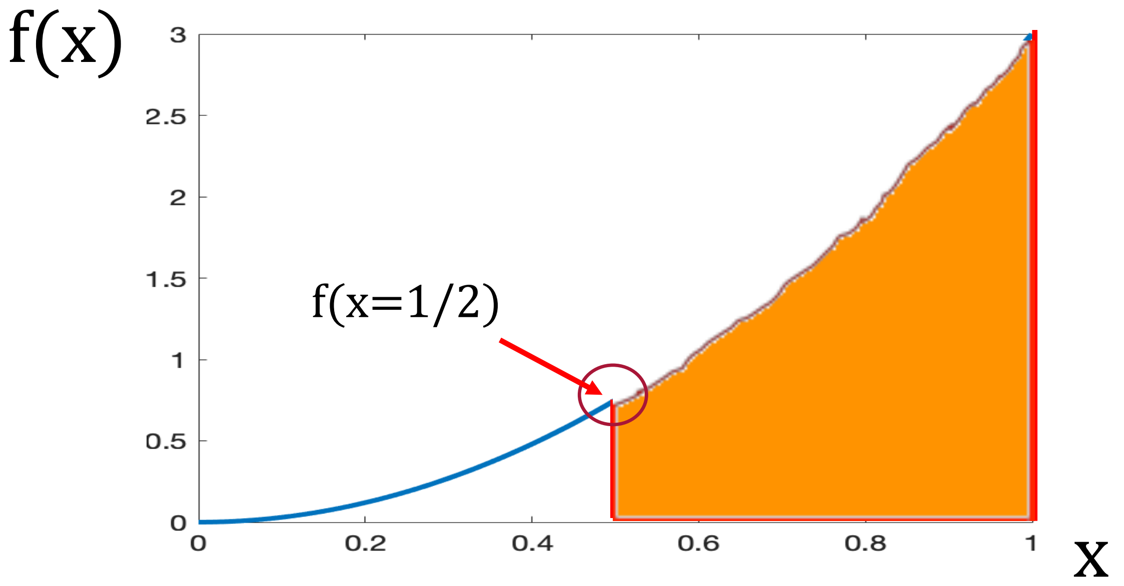
b) $\int_S f(x)dx = 1 \Rightarrow \int_0^1 3x^2 dx = [x^3]_0^1 = 1 - 0 = 1$

Answer: f(x) integrated over the entire support equals 1 i.e. f(x) is a valid PDF

PDF – example 1

2.

a) $P\left(\frac{1}{2} < x < 1\right)$



$$P\left(\frac{1}{2} < x < 1\right) = \int_{\frac{1}{2}}^{1} 3x^2 dx = [x^3]_{\frac{1}{2}}^{1} = 1 - \left(\frac{1}{2}\right)^3 = 1 - \frac{1}{8} = \frac{7}{8}$$

b) $P\left(x = \frac{1}{2}\right) = \int_{\frac{1}{2}}^{\frac{1}{2}} 3x^2 dx = [x^3]_{\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{8} - \frac{1}{8} = 0$

Grouped Frequencies

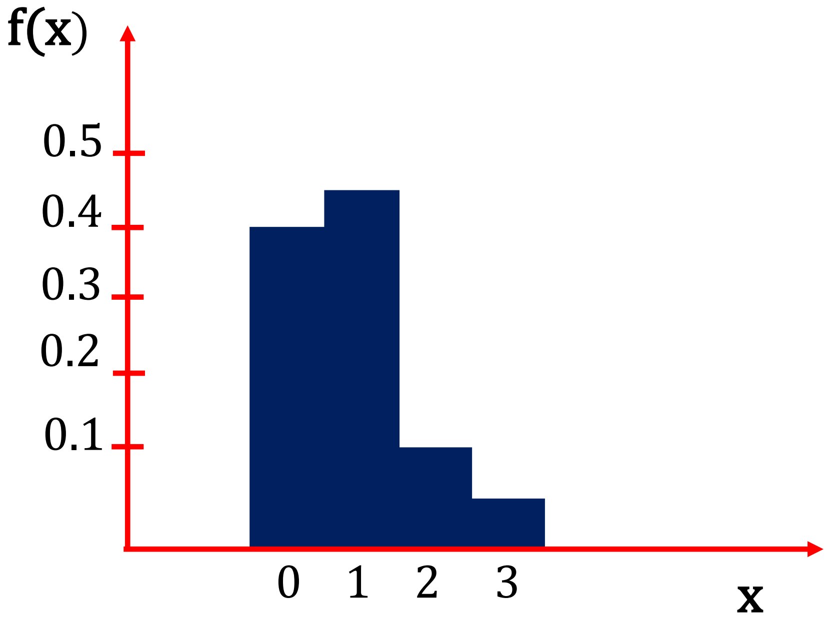
- For continues RV, within a range, there is an infinite number of possible values
- Thus, representing the continues data is easier by dividing the variate into intervals or classes and counting the frequency of occurrence for each class
- This is called the grouped frequency approach
- The width of the class is usually constant
- # of classes can be determine using empirical Sturges' rule:

$$\#\text{class intervals} \approx 1 + 3.3 \log_{10} N$$

N - total number of observations in the sample or population

Grouped Frequencies

- The class boundaries must be clear with no gaps and no overlaps
- The relative frequency: ratio of the class frequency and the total of all the class frequencies
- Often represented as a histogram



Histogram

Cumulative Distribution Function

CDF

$$F_X(s) = P(X \leq s)$$

- $F_X(s)$ is a nondecreasing function of s, for $-\infty < s < \infty$
- The CDF ranges from 0 to 1 ($F_X(s)$ is a probability!)
- Also called the distribution function.
- All probabilities concerning X can be stated in terms of $F_X(s)$.

CDF of a discrete RV

CDF of a discrete RV

$$F_X(s) = P(X \leq s) = \sum_{x \leq s} f(x)$$

- If X is a discrete random variable whose minimum value is z, then:

  $F_X(z)$ =P(X ≤ z)= P(X=z) = $f_X(z)$.

  If c is less than z, then $F_X(c) = 0$

- If the maximum value of X is b, then $F_X(b) = 1$
- CDF of a discrete RV is a non-decreasing step function

CDF – Example 1

PMF of a discrete random variable X is equal to:

$$f(x) = \frac{5-x}{10} \qquad \text{for } x=1,2,3,4.$$

Find the CDF of X.

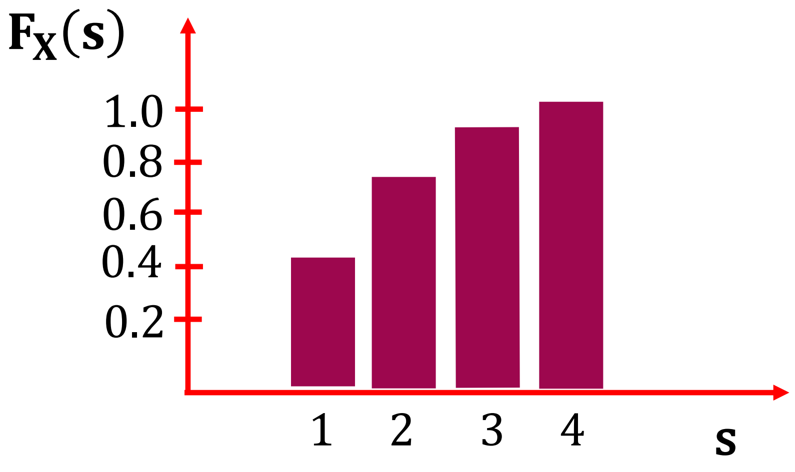
x	1	2	3	4
f(x)=P(X=x)				

CDF – solution

x	1	2	3	4
f(x)=P(X=x)	0.4	0.3	0.2	0.1

The CDF is: $F_X(s) = P(X \leq s)$

- For s=1, P(X≤1)=P(X=1)=f(1)= $\frac{5-s}{10} = \frac{4}{10}$
- For s=2, P(X≤2)=P(X=1 or X=2)=P(X=1)+P(X=2)= $\frac{5-1}{10} + \frac{5-2}{10} = \frac{7}{10}$
- For s=3, P(X≤3)= $\frac{5-1}{10} + \frac{5-2}{10} + \frac{5-3}{10} = \frac{9}{10}$
- For s=4, P(X≤4)= $\frac{5-1}{10} + \frac{5-2}{10} + \frac{5-3}{10} + \frac{5-4}{10} = 1$



s	1	2	3	4
$F_X(s) = P(X \leq s)$	0.4	0.7	0.9	1

CDF of a continues RV

CDF of a continues RV

$$F_X(s) = P(X \leq s) = \int_{-\infty}^{s} f(x)dx$$

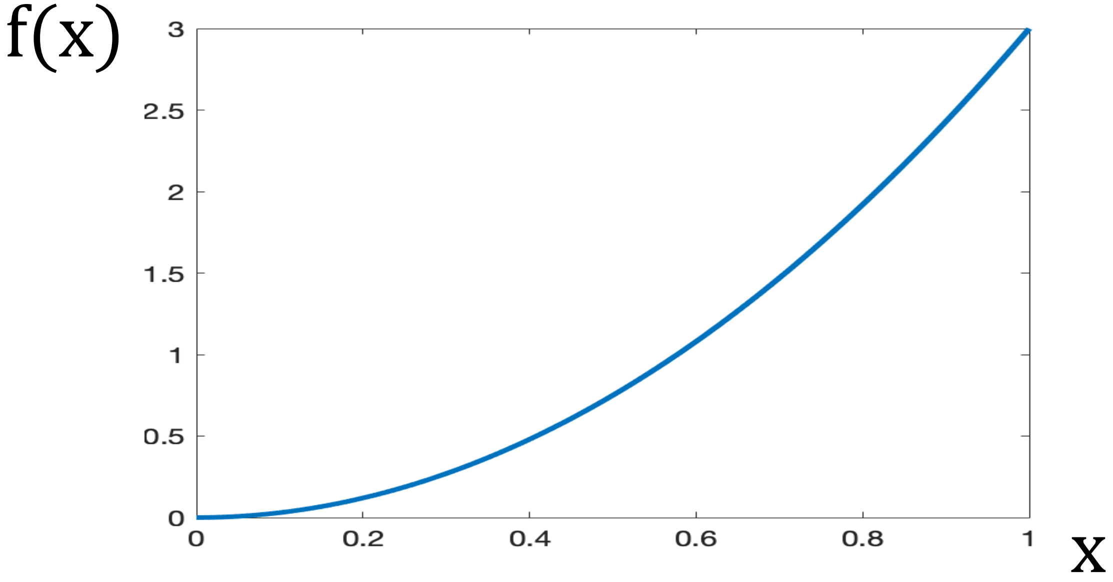
- ➢ CDF of a continues RV is a non-decreasing continues function

CDF – Example 2

For the RV X with the probability density function:

$$f(x) = 3x^2 \quad \text{for } 0 < x < 1.$$

calculate the CDF

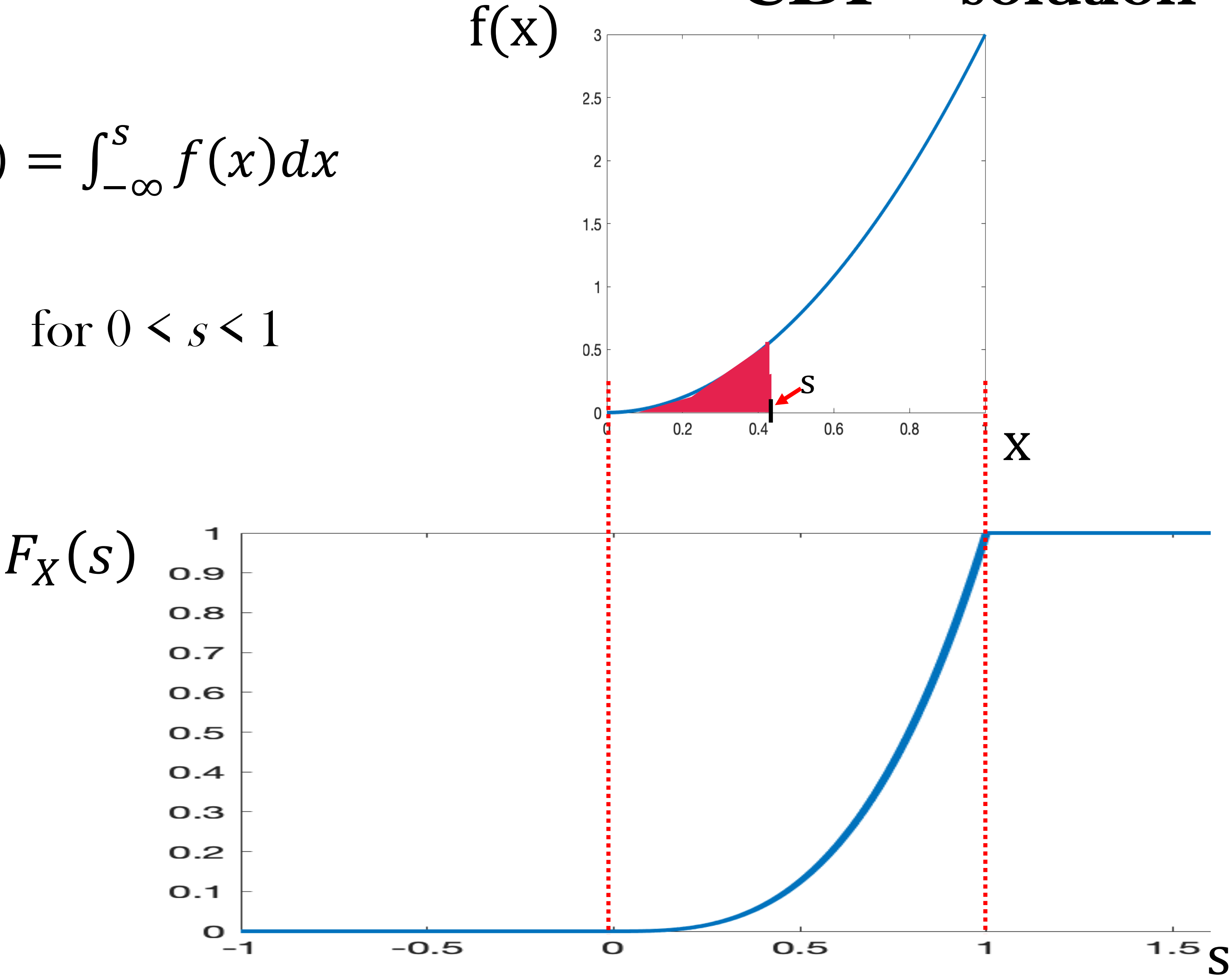


CDF – solution

The CDF is: $F_X(s) = P(X \leq s) = \int_{-\infty}^{s} f(x)dx$

$$F_X(s) = \int_0^s 3x^2 dx = [x^3]_0^s = s^3 \quad \text{for } 0 < s < 1$$

$$F_X(s) = \begin{cases} 0, & \text{for } s \leq 0 \\ s^3, & \text{for } 0 < s < 1 \\ 1, & \text{for } s \geq 1 \end{cases}$$



Example

- Calculate average of the result of 36 dice throws

1	3	2	4	5	6	5	4	4	1	2	3	
2	1	3	6	4	5	5	4	3	1	6	6	.....
2	5	3	1	2	4	6	2	5	6	3	1	

$$\text{Avg(x)} = \frac{(1+1+1..+1)+(2+2..+2)+\cdots(6+6+\cdots+6)}{36} =$$

$$= \frac{\cancel{6}(1)+\cancel{6}(2)+\cdots\cancel{6}(6)}{\cancel{3}6}$$

Probabilities of throwing a given no on the dice for large no of observations

Remember that $P(A) = \lim_{n\to\infty}\left(\frac{N_n(A)}{n}\right)$

Expected Value of a RV

- Expected value - the mean of all possible results for an infinite number of trials
- Expected value is an average of the values weighted by their probabilities

Expected value or a MEAN of a RV

$\mu_X = E(X) = \sum_{x \in S} x\, f(x)$ for discrete RV

$\mu_X = E(X) = \int_{-\infty}^{\infty} x\, f(x) dx$ for continuous RV

- E(X) exists only if sum exists/integral converges absolutely ($\neq \infty$)

Expected Value – Example 1

The probability that a 30-year-old man will survive a fixed length of time is 0.995. The probability that he will die during this time is thus 1– 0.995 = 0.005. An insurance company will sell him a $20,000 life insurance policy for this length of time for a premium of $200.00. What is the expected gain for the insurance company?

Answer:

If the man lives through the fixed length of time, the company's gain will be $200.00. The probability of this is 0.995.

If the man dies during this time, the company's gain will be

+$200.00 – $20,000.00 = – $19,800.00. The probability of this is 0.005.

Thus the expected gain for the company is

$$E(X) = (\$200.00)(0.995) + (-\$19{,}800.00)(0.005) = \$199.00 - \$99.00 = \$100.00$$

Expected Value of a Function of RV

Expected value of a function of a RV

$E(g(X)) = \sum_{x \in S} g(X) f(x)$ for discrete RV

$E(g(X)) = \int_{-\infty}^{\infty} g(X) f(x) dx$ for continuous RV

Expected Value – Example 2

Calculate E(g(X)) for a function g(x) = cos(X),

where X is a RV with following uniform PDF:

$$\mathrm{f}(x) = \begin{cases} \dfrac{1}{2\pi} & for\ -\pi < x < \pi \\ 0 & otherwise \end{cases}$$

Solution:

$$E(X) = \int_{-\pi}^{\pi} \cos(x) \frac{1}{2\pi} dx$$

Properties of Expected Value

Properties of E(X)

1. $E(c) = c$ where c is a constant
2. $E(cg(X)) = cE(g(X))$
3. $E(c_1 g_1(X) + c_2 g_2(X)) = c_1 E(g_1(X)) + c_2 E(g_2(X))$

Property 3 can be extended to:

$$E\left(\sum_i c_i g_i(X)\right) = \sum_i c_i E(g_i(X))$$

Properties of Expected Value

Proof 1:

$$E(c) = \sum_{x \in S} cf(x) = c \sum_{x \in S} f(x) = c$$

=1 from the property 2 of PMF

Proof 2:

$$E(cg(X)) = \sum_{x \in S} cg(x)f(x) = c \sum_{x \in S} g(x)f(x) = cE(X)$$

=E(X) from the definition

Properties of Expected value

Proof 3:

$$E(c_1g_1(X) + c_2g_2(X)) = \sum_{x \in S}(c_1g_1(x) + c_2g_2(x))f(x) =$$

$$= \sum_{x \in S} c_1g_1(x)f(x) + \sum_{x \in S} c_2g_2(x)f(x) = c_1\sum_{x \in S} g_1(x)f(x) + c_2\sum_{x \in S} g_2(x)f(x)$$

E(X) from the definition

$$= c_1E(g_1(X)) + c_2E(g_2(X))$$

What have we learnt?

- Probability density function
- Cumulative distribution function
- Expected value and its properties