Lecture 4

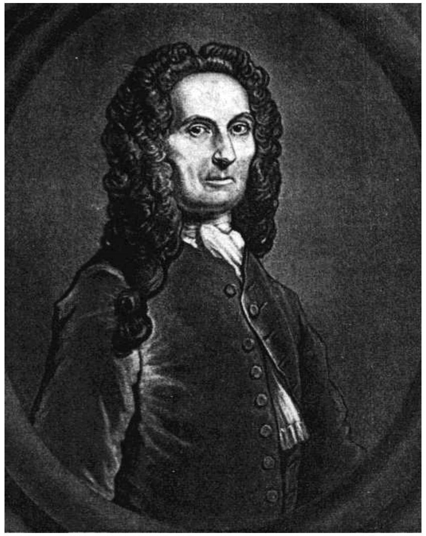
De Moivre-Laplace Limit Theorem, Random Variables, Notable Probability Distributions

De Moivre-Laplace Limit Theorem

Yet another result approximating binomial experiment is given by de Moivre-Laplace theorem (*local and integral*).

Pierre-Simon Laplace
1749-1827

Consider a binomial experiment (Bernoulli trials) consisting of n trials with p – the probability of success, and $q = 1 - p$ – the probability of failure.



Abraham de Moivre
1667-1754

De Moivre-Laplace Limit Theorem

The probability $b(k;n,p)$ of obtaining exactly k successes in the n trials is obtained by

$$b(k;n,p) = C_k^n p^k q^{n-k}.$$

De Moivre-Laplace limit theorem (local) provides the following approximation:

$$b(k;n,p) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}.$$

$\boldsymbol{npq \geq 20}$

De Moivre-Laplace Limit Theorem

The probability $P_n(m_1 \leq k \leq m_2)$ that number of successes would be $k \in [m_1; m_2]$ can be obtained by the exact formula

$$P_n(m_1 \leq k \leq m_2) = \sum_{i=m_1}^{m_2} C_i^n p^i q^{n-i}.$$

If $(m_2 - m_1)$ is large, the calculation might be rather tedious.

De Moivre-Laplace limit theorem (integral) :

$$P_n(m_1 \leq k \leq m_2) \approx \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{x^2}{2}} dx .$$

where

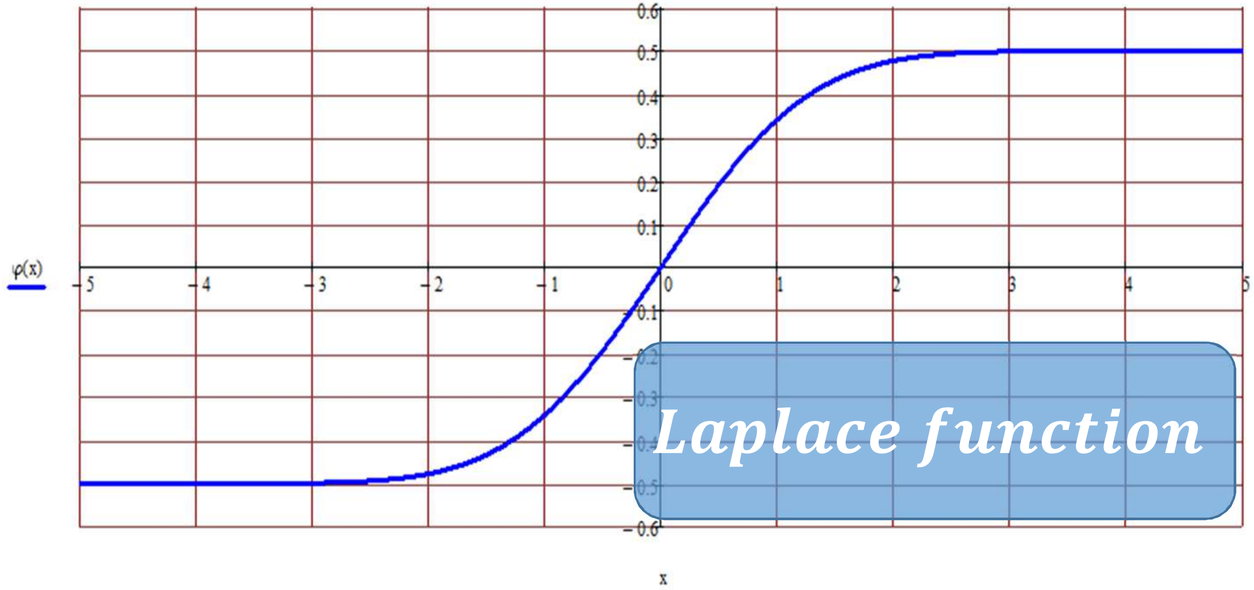
$$\alpha = \frac{m_1 - np}{\sqrt{npq}}, \qquad \beta = \frac{m_2 - np}{\sqrt{npq}}.$$

$\boldsymbol{npq \geq 20}$

De Moivre-Laplace Limit Theorem

Let’s introduce a function

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$



Algebraic properties:

$$\Phi_0(x) = -\Phi_0(-x)$$

$$\Phi_0(x) \to \pm 0{,}5 \quad as\ x \to \pm\infty$$

$$\Phi_0(x) \approx 0{,}5 \quad for\ x > 4$$

De Moivre-Laplace Limit Theorem

Then

$$P_n(m_1 \leq k \leq m_2) \approx \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{x^2}{2}} dx = \Phi_0(\beta) - \Phi_0(\alpha).$$

The values of $\Phi_0(x)$ are often tabulated.

A true die is tossed 12000 times.

Calculate

(a) the probability that there will be exactly 1800 rolls of 6;

(b) the probability that the number of 6’s lies in the interval

[1950, 2100].

We have a Bernoulli trials with $n = 12000, p = \frac{1}{6}, q = \frac{5}{6}$.

(a) $k = 1800$;

The exact probability is

$$b\left(1800; 12000, \frac{1}{6}\right) = C_{1800}^{12000} \left(\frac{1}{6}\right)^{1800} \left(\frac{5}{6}\right)^{10200}$$

Poisson Limit Theorem gives us $\lambda = np = 2000$ and

$$b\left(1800; 12000, \frac{1}{6}\right) \approx \frac{2000^{1800}}{1800!} e^{-2000}$$

De Moivre-Laplace Limit Theorem

According to de Moivre-Laplace Limit Theorem

$$b(k;n,p) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}.$$

$$b\left(1800;12000,\frac{1}{6}\right) \approx 9.772 \cdot 10^{-3} \cdot 6.144 \cdot 10^{-6} \approx 6 \cdot 10^{-8}$$

(b) $m_1 = 1950, m_2 = 2100$;

$$\alpha = \frac{m_1 - np}{\sqrt{npq}} = \frac{1950 - 2000}{\sqrt{12000 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \approx -1.225$$

$$\beta = \frac{m_2 - np}{\sqrt{npq}} = \frac{2100 - 2000}{\sqrt{12000 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \approx 2.449$$

Consulting a table of Laplace function $\Phi_0(x)$ or calculating its values for $x = \alpha$ and $x = \beta$ with Mathcad, we obtain

$\Phi_0(\alpha) = -\Phi_0(-\alpha) = -\Phi_0(-1{,}225) \approx -0{,}3897$;

$\Phi_0(\beta) = \Phi_0(2{,}449) \approx 0{,}4928$;

$P_n(1950 \leq k \leq 2100) \approx \Phi_0(\beta) - \Phi_0(\alpha) = 0{,}8825$.

In many applications, the outcomes of a probabilistic experiment are numbers, and we can use these numbers to obtain important information.

We can, for instance, describe in various ways how large or small these numbers are likely to be and compute likely averages and measures of spread.

For example, in three tosses of a coin, the number of heads obtained can range from 0 to 3, and there is one of these numbers associated with each possible outcome.

Informally, the quantity "number of heads" is called a random variable and the numbers 0 to 3 its possible values.

In general, such an association of numbers with each member of a set is called a function.

Random Variables

A random variable (abbreviated r.v.) is a real-valued function on a sample space Ω.

Random variables are usually denoted by capital letters, such as X, Y, Z, and sets like $\{s: X(s) = x\}$, $\{s: X(s) \leq x\}$, for any number x, are events in Ω.

All events in Ω have probabilities associated with them.

$$\Pr(X = x) \qquad \Pr(X \leq x)$$

The assignment of probabilities to all such events, for a given random variable X, is called the probability distribution of X.

A probability distribution is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment.

In more technical terms, the probability distribution is a description of a random phenomenon in terms of the probabilities of events.

Probability distributions are generally divided into two classes:

- discrete (discrete sample space);
- continuous (continuous sample space).

To define probability distributions, it is necessary to distinguish between discrete and continuous random variables.

In the discrete case, it suffices to list the possible values of X and their corresponding probabilities. This information is contained in the probability mass function of X.

The probability mass function (*pmf*) $p_X(x)$ of a discrete random variable X is the function $p_X\colon \mathbb{R} \to [0,1]$, defined by

$$p_X(x) = \Pr(X = x) \quad \text{for} \quad -\infty < x < \infty$$

A set $D_X = \{x \in \mathbb{R} : \Pr(X = x) > 0\}$ is called a domain of a random variable X.

At the same time, D_X is also called a support of a probability distribution: $D_X = \{x \in \mathbb{R} : p_X(x) > 0\}$.

If a domain of a discrete r.v. X contains but a few number of values, it is possible to define the probability distribution by simply listing values of X in a table together with their probabilities:

$\boldsymbol{x_i}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	…	…	$\boldsymbol{x_n}$
$Pr(x_i)$	$\boldsymbol{p_1}$	$\boldsymbol{p_2}$	…	…	$\boldsymbol{p_n}$

Obviously, $\sum_{i=1}^{n} p_i = 1$.

Let $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ describe three tosses of a coin, and let X denotes the number of "heads" obtained. Then the values of X, for each outcome ω in Ω, are given in the following table:

ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(\omega)$	3	2	2	1	2	1	1	0

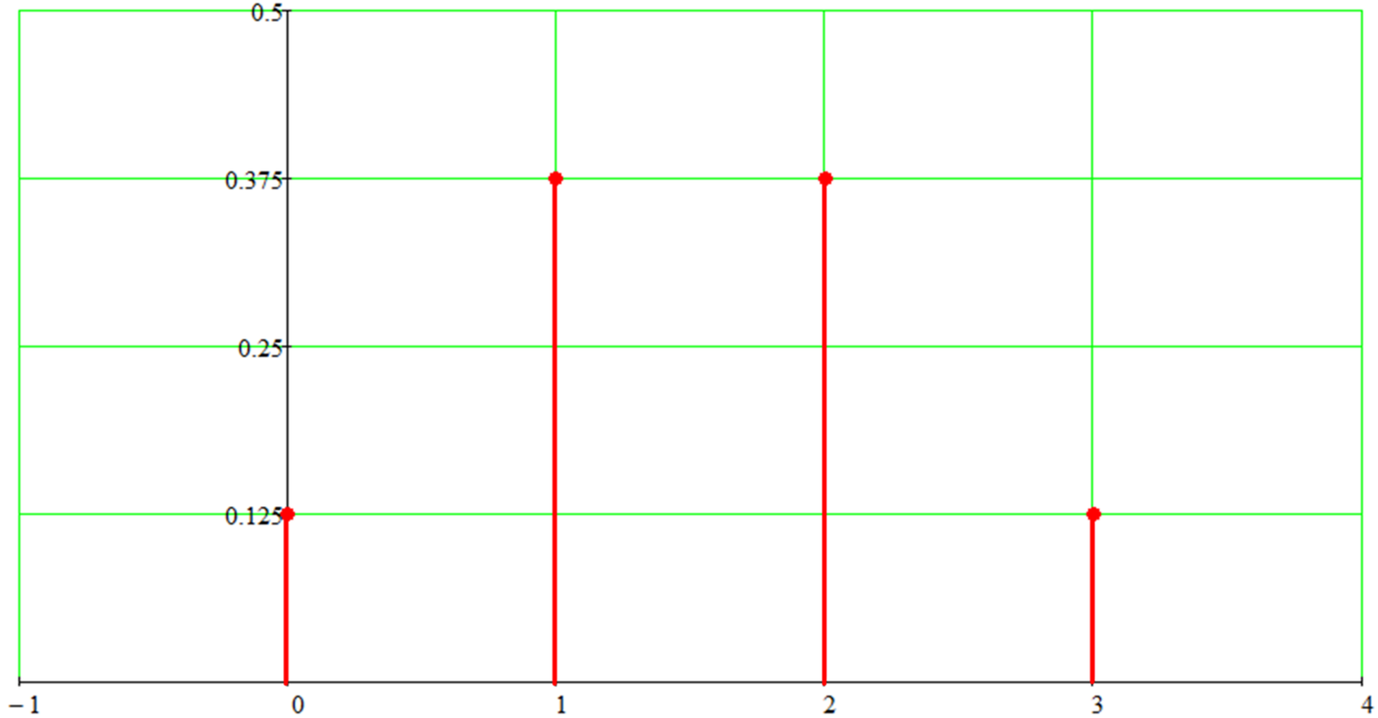
ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(\omega)$	3	2	2	1	2	1	1	0

Since Ω contains eight outcomes and X takes on values $0,1,2,3$, the probability distribution of X can be defined as:

x_i	**0**	**1**	**2**	**3**
$\Pr(x_i)$	$^1/_8$	$^3/_8$	$^3/_8$	$^1/_8$

Random Variables

x_i	0	1	2	3
$\Pr(x_i)$	$^1/_8$	$^3/_8$	$^3/_8$	$^1/_8$



Describing a probability distribution of a discrete r.v. X as a table becomes impractical, if there are large number of values x_i, and impossible, if X assumes infinite number of values.

In this case, the distribution should be defined by the probability mass function $p_X(x)$ in analytical form.

However, *pmf* is not suitable for defining continuous random variables, since the probability of X taking on a certain value is zero.

The distribution function of a random variable X (also known as the cumulative distribution function, *cdf*) allows us to treat discrete and continuous random variables in the same way.

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

$$F_X(x) = \Pr(X \leq x) \quad \text{for } -\infty < x < \infty$$

Both the *pmf* and the *cdf* of a discrete random variable X contain all the probabilistic information of X; the probability distribution of X is determined by either of them.

In fact, the distribution function F_X of a discrete random variable X can be expressed in terms of the probability mass function p_X of X and vice versa.

If X attains values $x_1, x_2, \ldots$, such that

$$p_X(x_i) > 0, \qquad p_X(x_1) + p_X(x_2) + \cdots = 1,$$

then

$$F_X(x) = \sum_{x_i \leq x} p_X(x_i)$$

The properties of *cdf* F_X:

1. For any $a \leq b$, $F_X(a) \leq F_X(b)$. This property is an immediate consequence of the fact that the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$. F_X is a non-decreasing function.

2. Since $F_X(x)$ is a probability, the value of the *cdf* is always between 0 and 1. Moreover,

$$\lim_{x \to \infty} F_X(x) = 1 \qquad \lim_{x \to -\infty} F_X(x) = 0$$

3. F_X is right-continuous, i.e.,

$$\lim_{\varepsilon \to 0} F_X(x + \varepsilon) = F_X(x)$$

The properties of *cdf* F_X:

4.

$$\Pr(a < x \leq b) = F_X(b) - F_X(a)$$

Remark:

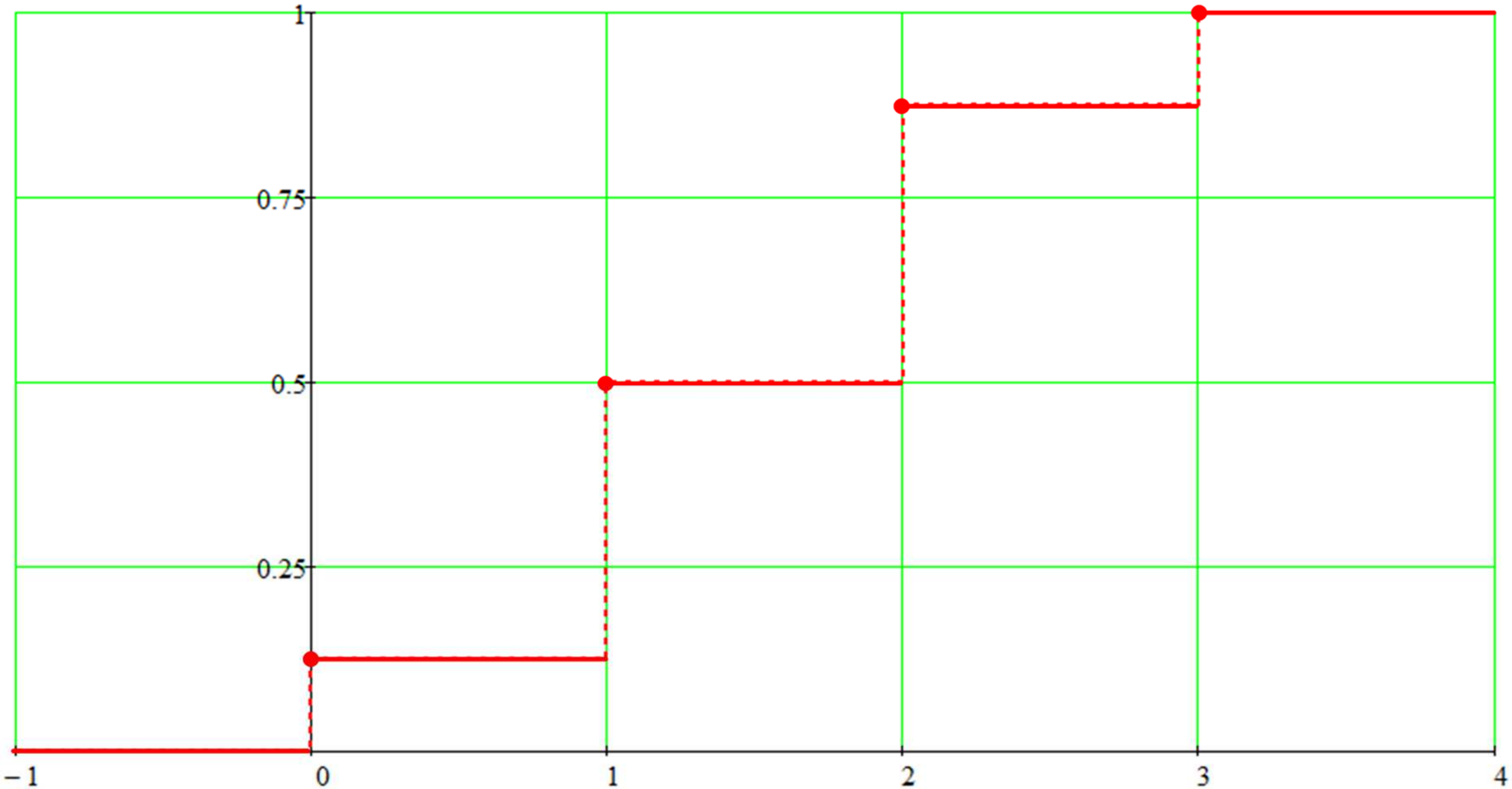
$$\Pr(a \leq x \leq b) = F_X(b) - F_X(a) + p_X(a) \quad \text{for discrete r.v.}$$

$$\Pr(a \leq x \leq b) = F_X(b) - F_X(a) \quad \text{for continuous r.v.}$$

x_i	0	1	2	3
$p_X(x_i)$	$^1/_8$	$^3/_8$	$^3/_8$	$^1/_8$

x_i	0	1	2	3
$F_X(x_i)$	$^1/_8$	$^4/_8$	$^7/_8$	1

Random Variables



We had defined continuous r.v. X as a function on an infinite uncountable sample space Ω. Here we have another definition.

A random variable X is continuous if for some function $f_X\colon \mathbb{R} \to \mathbb{R}$ and for any numbers a and b with $a \leq b$,

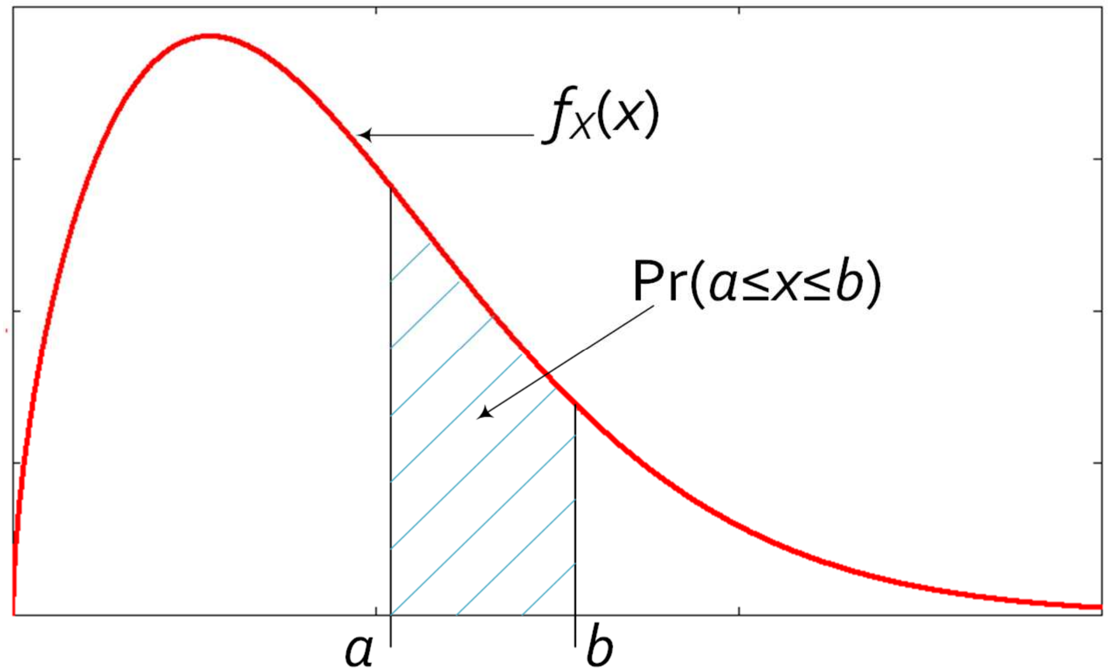
$$\Pr(a \leq x \leq b) = \int_a^b f_X(x)dx$$

The function f_X has to satisfy $f_X(x) \geq 0$ for all x and

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

We call f_X the probability density function (*pdf*) of X.

Note that the probability that X lies in an interval $[a, b]$ is equal to the area under the probability density function f_X of X over the interval $[a, b]$.



If the interval gets smaller and smaller, the probability will go to zero: for any positive ε

$$\Pr(a - \varepsilon \leq x \leq a + \varepsilon) = \int_{a-\varepsilon}^{a+\varepsilon} f_X(x)dx\,,$$

and sending ε to 0, it follows that for any a

$$\Pr(X = a) = 0$$

Probability density function $f_X(x)$ can be interpreted as a relative measure of how likely it is that X will be near x.

However, unlike *pmf* p_X, *pdf* doesn't represent probability!

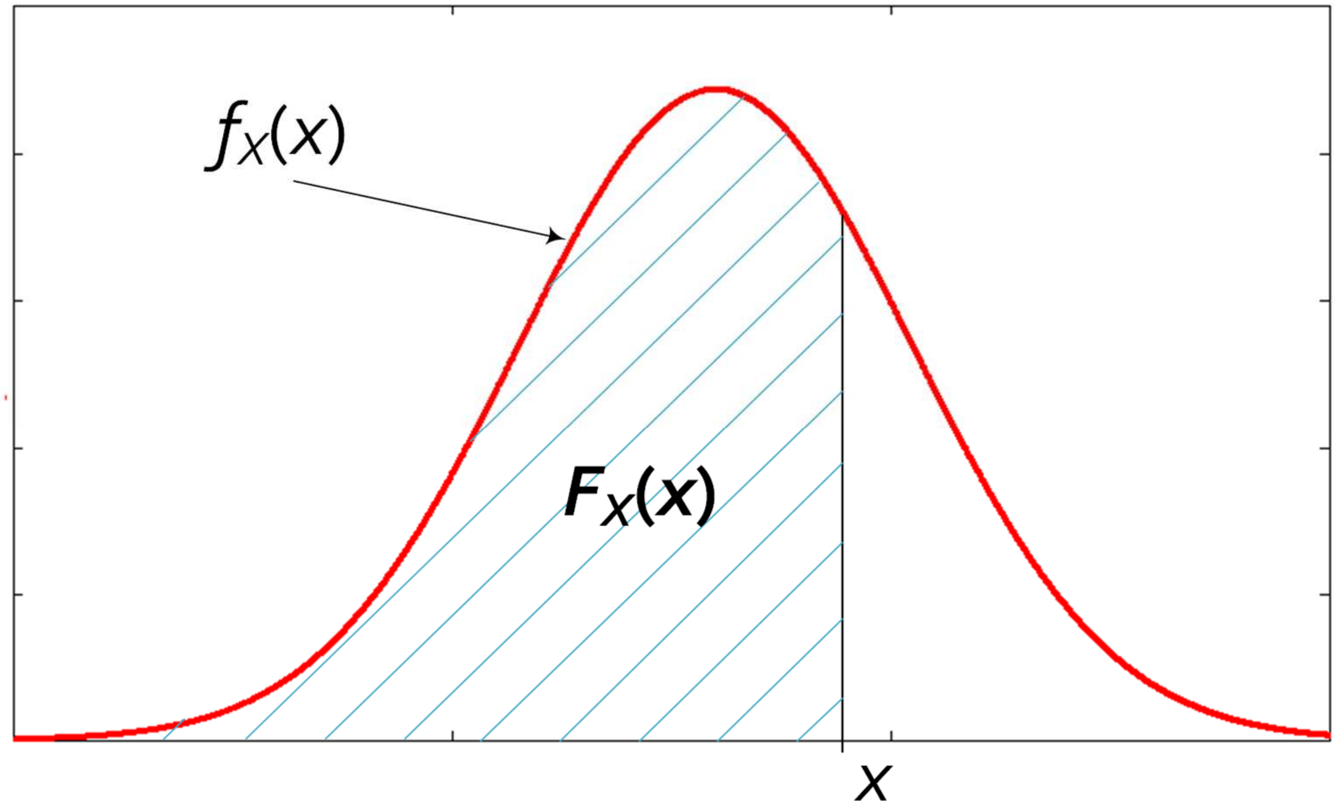
$$0 \leq p_X(x) \leq 1 \qquad f_X(x) \geq 0$$

There is a simple relation between the *cdf* $F_X(x)$ and the *pdf* $f_X(x)$ of a continuous random variable.

It follows from integral calculus that

$$F_X(x) = \int_{-\infty}^{x} f_X(\tau)d\tau \qquad f_X(x) = \frac{d}{dx}F_X(x)$$

$$F_X(x) = \int_{-\infty}^{x} f_X(\tau)d\tau$$



Both the probability density function (*pdf*) and the distribution function (*cdf*) of a continuous random variable X contain all the probabilistic information about X; the probability distribution of X is described by either of them.

The Bernoulli distribution

The Bernoulli distribution models an experiment with only two possible outcomes, often referred to as “success” and “failure”, usually encoded as 1 and 0.

A discrete r.v. X has a *Bernoulli distribution* with parameter p, where $0 \leq p \leq 1$, if its *pmf* is given by

$$p_X(1) = p; \quad p_X(0) = 1 - p.$$

We denote this distribution by $Ber(p)$.

Binomial distribution

Binomial distribution allows obtaining the probability of k "successes" in n Bernoulli trials, $Ber(p)$.

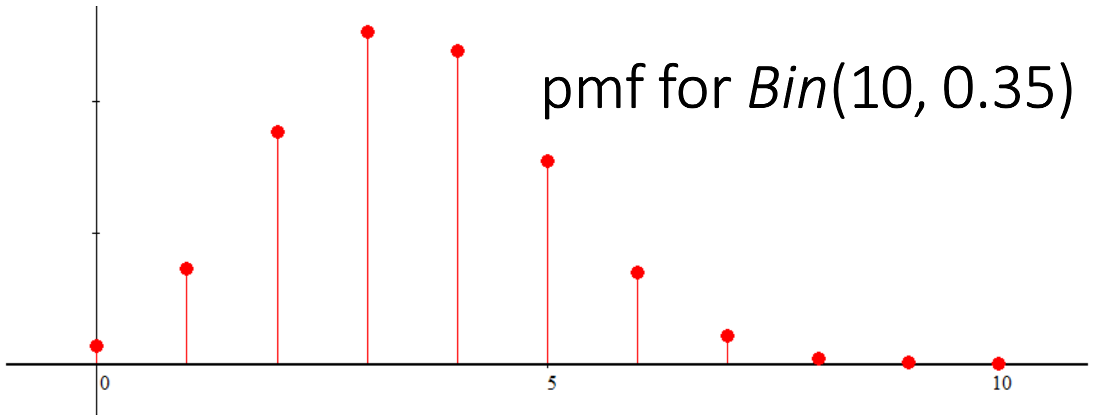
A discrete r.v. X has a *binomial distribution* with parameters n and p, where $n = 1,2, \ldots$ and $0 \leq p \leq 1$, if its *pmf* is given by

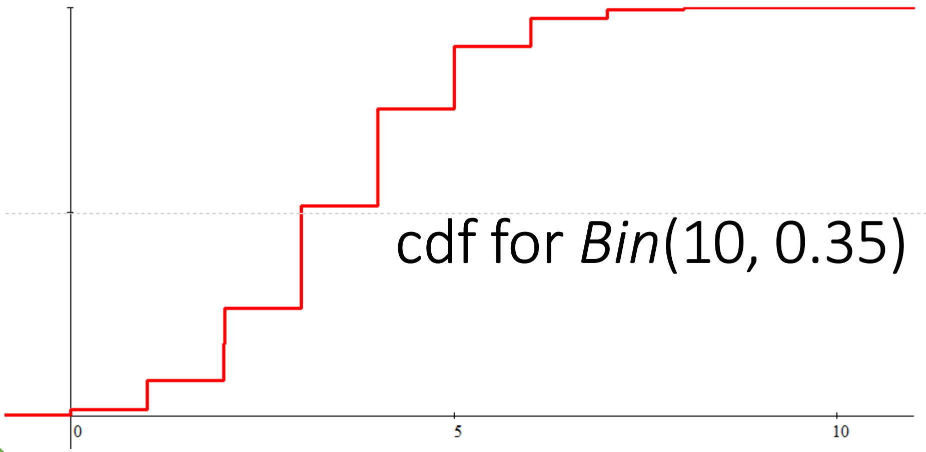
$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0,1, \ldots, n$$

We denote this distribution by $Bin(n, p)$.

Notable Probability Distributions

Binomial distribution





$$F_X(x) = \begin{cases} 0, & if\ x < 0; \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}, & if\ 0 \le x < n; \\ 1, & if\ x \ge n. \end{cases}$$

Geometric distribution

Suppose we perform independent Bernoulli trials ($Ber(p)$) with parameter p, until we obtain a success.

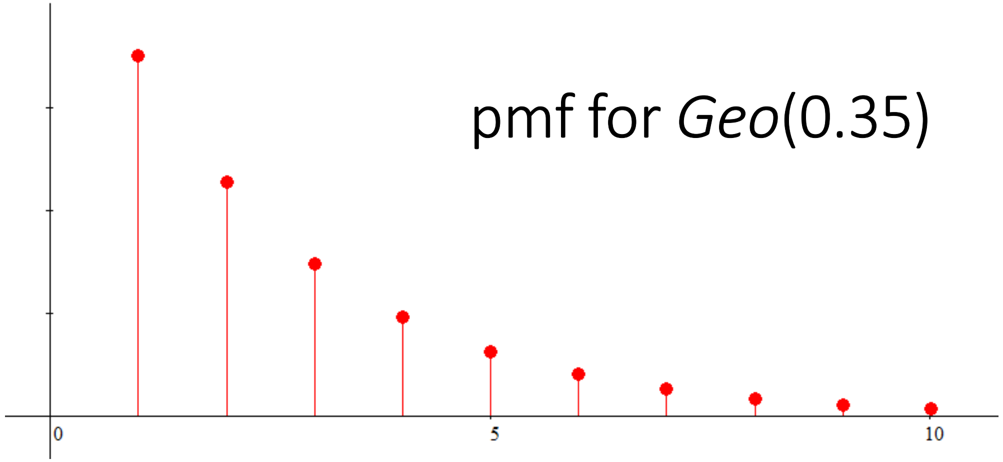
The number X of trials is called a geometric random variable with parameter p. Its *pmf* is given by

$$p_X(k) = p(1-p)^{k-1}, \qquad k = 1,2,\ldots$$

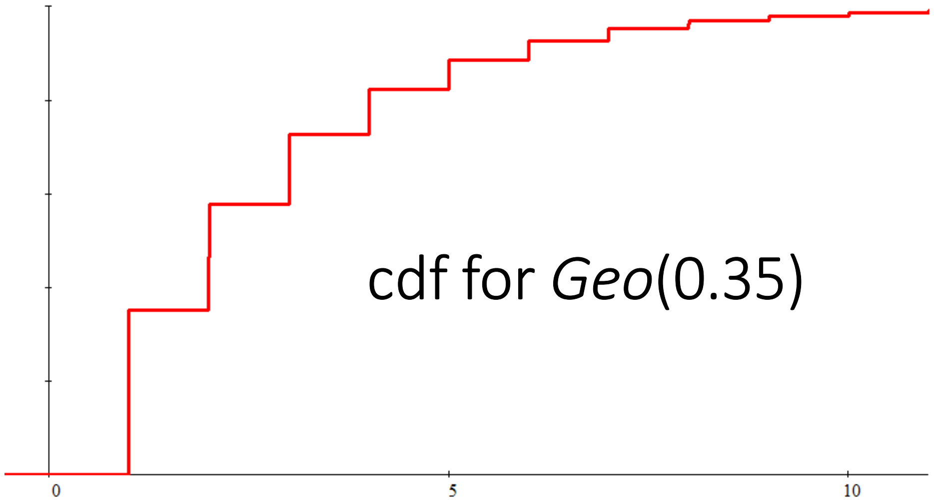
We denote this distribution by $Geo(p)$.

Notable Probability Distributions

Geometric distribution



pmf for *Geo*(0.35)



cdf for *Geo*(0.35)

$$F_X(x) = \begin{cases} 0, if\ x < 1; \\ 1 - (1-p)^{\lfloor x \rfloor}, if\ x \geq 1. \end{cases}$$

Discrete Uniform distribution

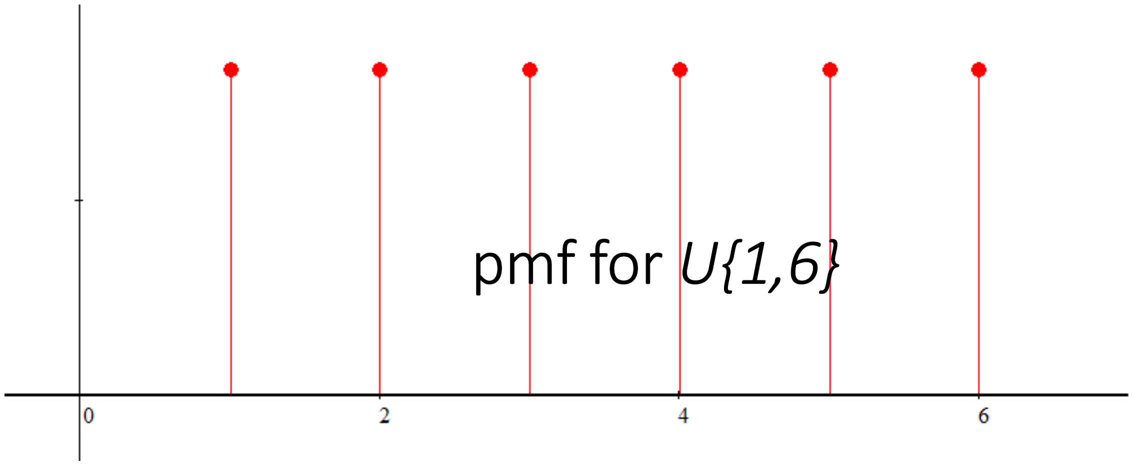
A random variable X and its distribution are called discrete uniform if X has a finite number of possible values, say $x_1, x_2, \dots, x_n$, for any positive integer n, and

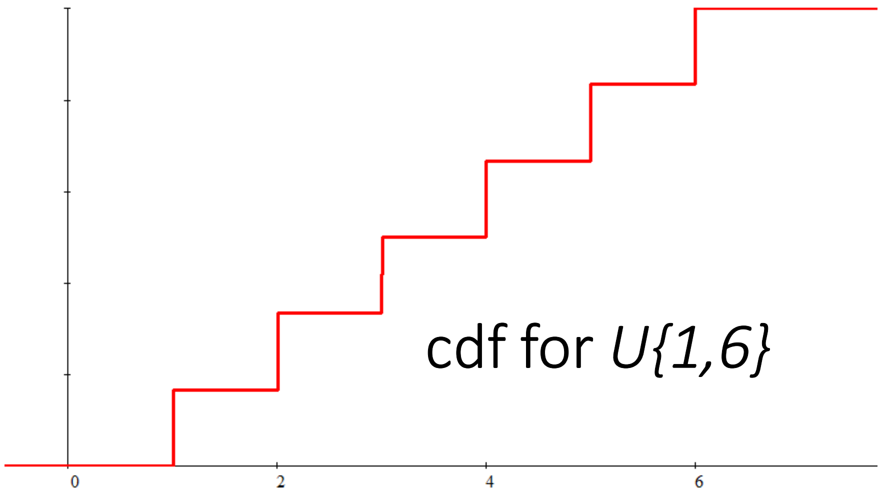
$$p_X(k) = \frac{1}{n}, \qquad k = 1,2, \dots, n.$$

It's common practice to denote $x_1 = a, x_n = b$. Then, the notation for discrete uniform distribution is $U\{a, b\}$.

Notable Probability Distributions

Discrete Uniform distribution





$$F_X(x) = \begin{cases} 0, & if\ x < a; \\ \dfrac{\lfloor x \rfloor - a + 1}{n}, & if\ a \le x < b; \\ 1, & if\ x \ge b. \end{cases}$$

Lecture 4

Textbook Assignment

Géza Schay. *Introduction to Probability…*

❖Chapter 5. 105-114 pp.

F.M. Dekking et al. *A Modern Introduction to…*

❖Chapter 4. 41-55 pp.

❖Chapter 5. 56-60 pp.