2. Random variables

2.1. Discrete random variables

In Galileo's example we had $\Omega = \{1, 2, 3, 4, 5, 6\}^3$. The outcomes are of the form (i, j, k). The gamblers, however, were not so interested in the inner structure of the outcome but in the random number that was the sum of dots. We can imagine that random numbers are created through some process that involves chance. In probability we call such random numbers random variables. We denote them by letters X, Y, Z.

Technical note: Formally we understand random variables as functions from Ω to the real numbers $\mathbb{R}$. We imagine that some invisible „hand" chooses the outcome ω and the random variable X gives the random number $X(\omega)$.

Definition: A random variable X is a function $X : \Omega \to \mathbb{R}$ such that $X^{-1}((a, b])$ is an event for all $a < b$, $a, b \in \mathbb{R}$.

Note: The choice of intervals of the form $(a, b]$ is arbitrary. Intervals of the form (a, b), $[a, b]$ do the same.

Definition: A random variable X is discrete if it can only take values in a finite or countable set $\{x_1, x_2, \dots\}$.

We can see that for discrete random variables we can require $X^{-1}(\{x_k\})$ to be an event for all possible values. This definition is equivalent to the more general one.

To continue with Galileo's example gamblers were interested in the probabilities that the sum is 9 or 10. We can ask the question for any $k \in \{3, 4, \dots, 18\}$.

<u>A note on notation</u>: We will write $\{X = x_k\}$ for the event $X^{-1}(\{x_k\})$. When we write probabilities $P(\{X = x_k\})$ we will drop the curly brackets and write $P(X = x_k)$.

We obviously have $\bigcup_{k=3}^{18} \{X = k\} = \Omega$. Since the events are disjoint we have

$$\sum_{k=3}^{18} P(X = k) = P(\Omega) = 1.$$

The total probability 1 is "distributed" among all possible values of X.

We say that these probabilities determine the distribution of X.

Definition: Let X be a discrete random variable. The distribution of X is given by the probabilities $P(X = x_k)$ for all possible values of X.

There is a number of standard distributions in probability.

Binomial distribution

Suppose we toss a coin n times. Suppose the tosses are independent and the probability of heads is $p \in (0,1)$. Let X = # of heads in n tosses. This random variable has values $k = 0, 1, 2, \ldots, n$.

To describe the distribution we need to compute $P(X = k)$ for all k.

We have $\Omega = \{H, T\}^n$ and the event $\{X = k\}$ consists of all sequences HTHHT ..... H that contain exactly k heads. Every such outcome has the probability $p^k (1-p)^{n-k}$ because of independence. So we only need to compute how many such outcomes there are. But this is given by $\binom{n}{k}$ because we need to choose k positions for heads among n positions.

We have

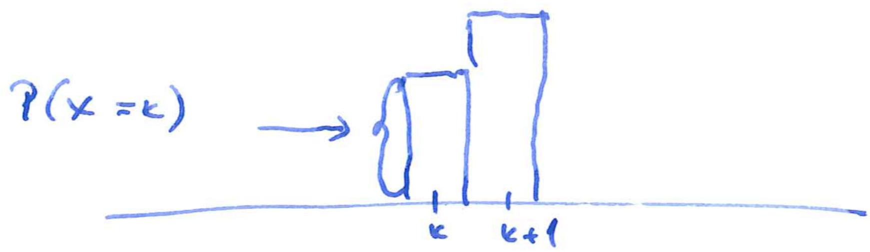
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots, n$. We say that X has binomial distribution with parameters n and p.

Notation: $X \sim \mathrm{Bin}(n, p)$.

The way to visualize a distribution is to draw a histogram. If X is a random variable with integer values we draw a column over a possible value k of X with base 1 and height $P(X=k)$ centered on k.

Figure:



Let us consider $X \sim Bin(n,p)$.

For $k \geq 1$ we can compute

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

$$= \frac{n-k+1}{k} \cdot \frac{p}{1-p}$$

If $\dfrac{P(X=k)}{P(X=k-1)} > 1$, then we have $P(X=k) > P(X=k-1)$, i.e. the column over k is taller than the column over $k-1$. This happens if

$$\frac{n-k+1}{k} \cdot \frac{p}{1-p} > 1 \Rightarrow$$

$$(n-k+1)p > k(1-p) \Rightarrow$$

$$(n+1)p > k$$

We have two cases:

1. If $(n+1)p$ is not an integer then the tallest column is over $k = \lfloor (n+1)p \rfloor$.

2. If $(n+1)p$ is an integer then for $k=(n+1)p$ we have

$$\frac{P(X=k)}{P(X=k-1)} = 1$$

This means that the two columns over $k = (n+1)p$ and $(n+1)p - 1$ are the tallest and equal.

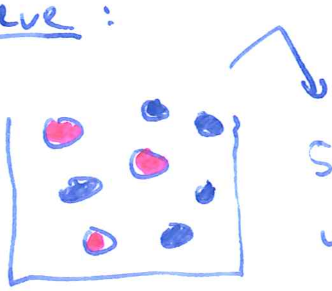
Coin tosses are methaphors for conting „successes" in identical and independent repetitions of the same experiment.

Hyper-geometric distribution

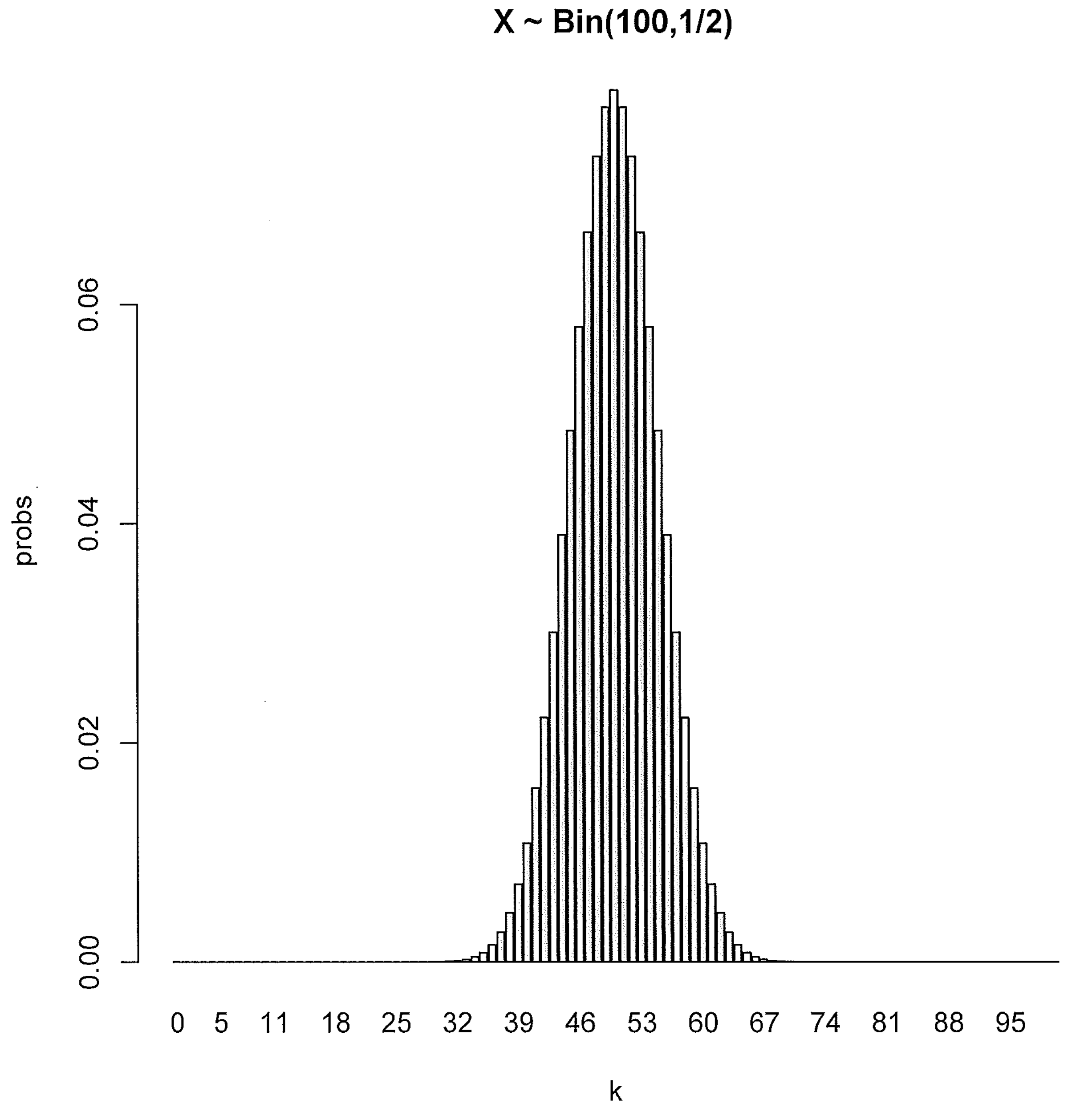
Suppose we have an urn with B black and R red balls.

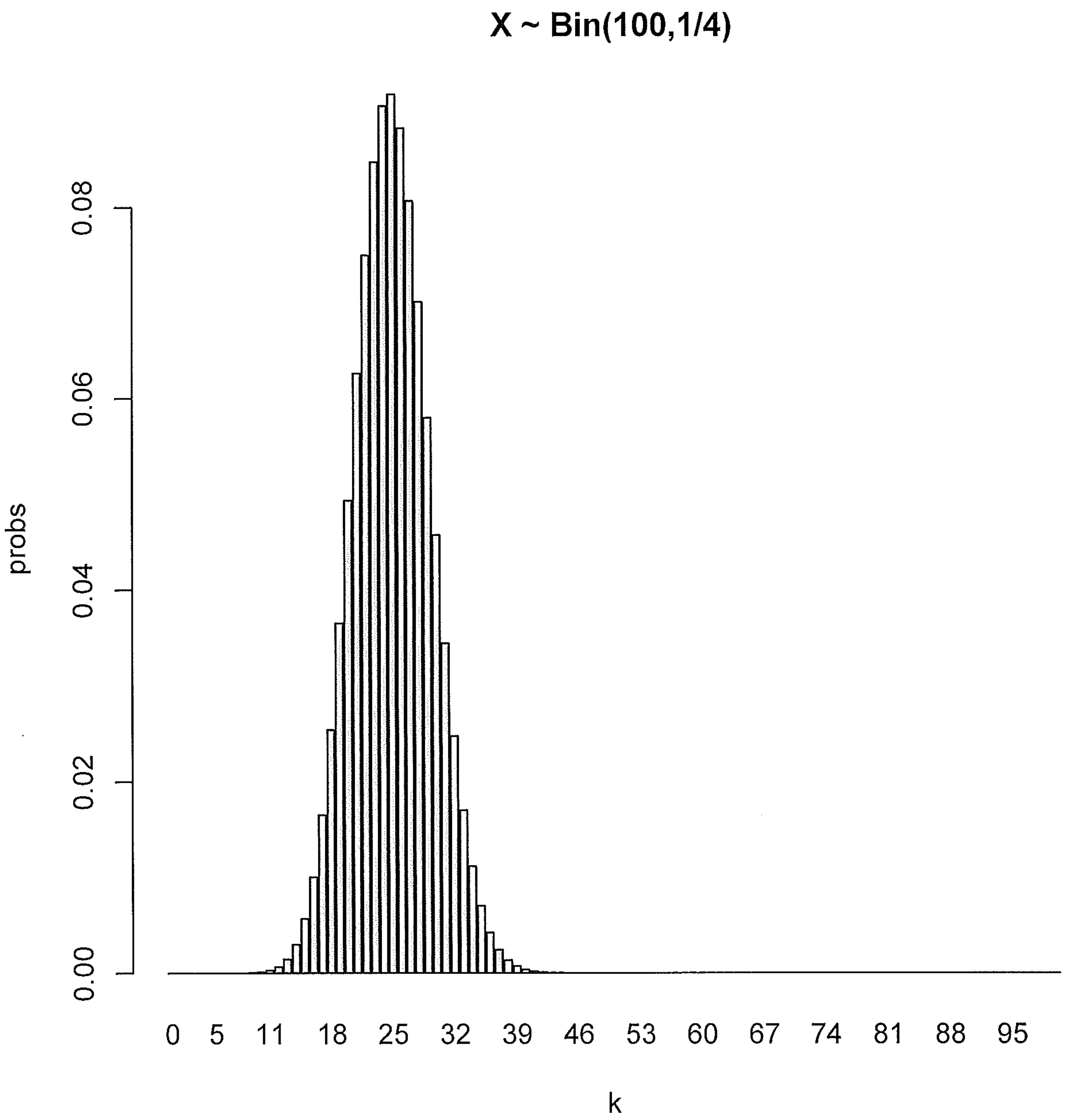
Denote $N = B + R$. Suppose we select $n \leq N$ balls at random from the urn. In mathematical terms this means that all $\binom{N}{n}$ possible selections of n balls are equally likely.

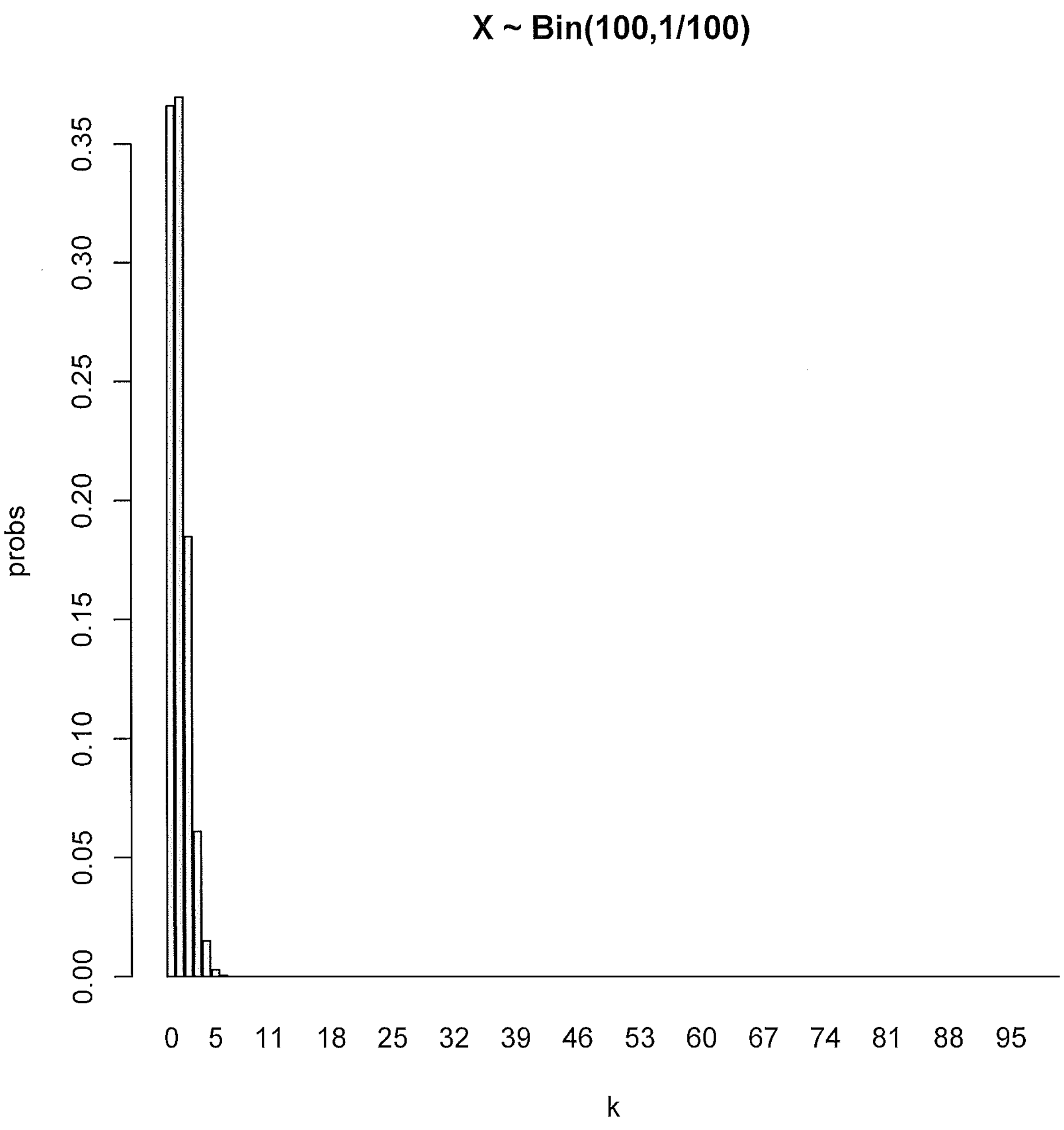
Figure:

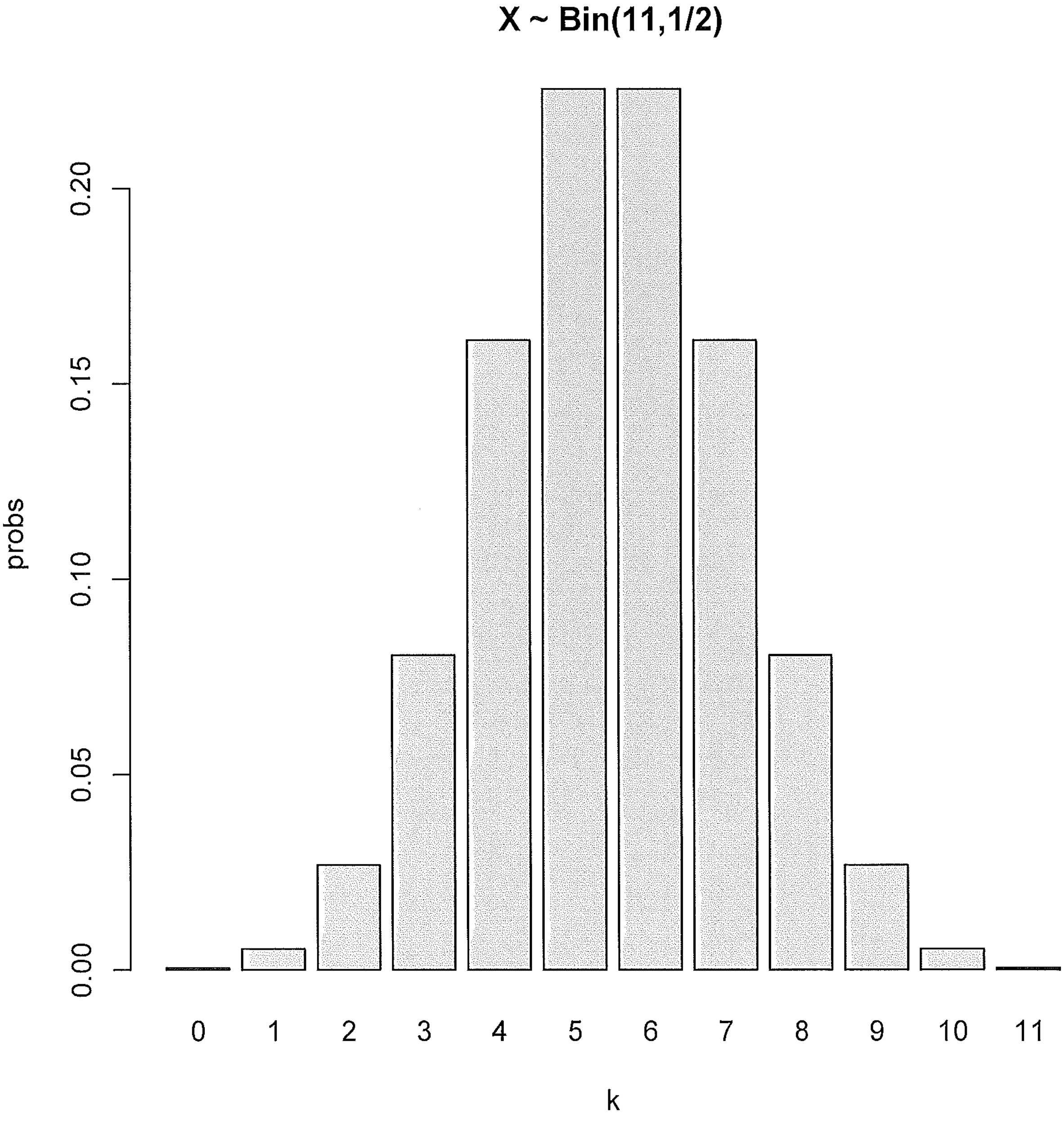


Select n balls at random without replacement.


X ~ Bin(100,1/2)
probs
0.00
0.02
0.04
0.06
0 5 11 18 25 32 39 46 53 60 67 74 81 88 95
k


X ~ Bin(100,1/4)
probs
0.00
0.02
0.04
0.06
0.08
0 5 11 18 25 32 39 46 53 60 67 74 81 88 95
k


X ~ Bin(100,1/100)
probs
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
5
11
18
25
32
39
46
53
60
67
74
81
88
95
k


X ~ Bin(11,1/2)
probs
0.00
0.05
0.10
0.15
0.20
0 1 2 3 4 5 6 7 8 9 10 11
k

Definition: If

$$P(X=k) = \frac{\binom{B}{k}\binom{R}{n-k}}{\binom{N}{n}} \quad \text{for}$$

$$\max(0, n-R) \leq k \leq \min(n, B)$$

we say that X has the hyper-geometric distribution with parameters n, B and $N = B+R$.

Shorthand:

$$X \sim \text{HiperGeom}(n, B, N).$$

Let X = number of black balls among the n selected. X is a random variable with values k that must satisfy

$$\max(0, n-R) \le k \le \min(n, B).$$

We have

$$P(X=k) = \frac{\binom{B}{k}\binom{R}{n-k}}{\binom{N}{n}}$$

The denominator is the number of all possible selection and the numerator is the number of selections with <u>exactly</u> k black and $n-k$ white balls. As with the binomial distribution we can calculate

$$\frac{P(X=k)}{P(X=k-1)} = \frac{(B-k+1)}{k} \cdot \frac{(n-k+1)}{R-n+k}$$

After some calculation we find that

$$\frac{P(X=k)}{P(X=k-1)} > 1 \quad \text{if} \quad k < \frac{(B+1)(n+1)}{N+2}$$

Again we have two cases:

1. $\dfrac{(B+1)(n+1)}{(N+2)}$ is not an integer.

   Then $k = \left\lfloor \dfrac{(B+1)(n+1)}{N+2} \right\rfloor$ is the tallest column.

2. $\dfrac{(B+1)(n+1)}{N+2}$ is an integer.

   Then $k = \dfrac{(B+1)(n+1)}{N+2}$ is still the largest probability but is equal to $P(X=k-1)$.

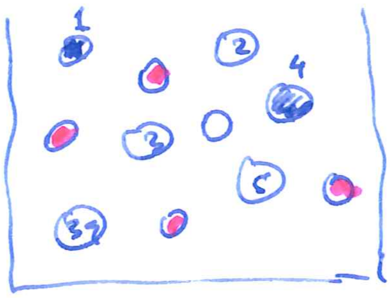
<u>Example</u> : Lottery.

A lottery ticket has 39 numbers. We cross out m numbers where $m = 8, 9, \ldots, 17$;

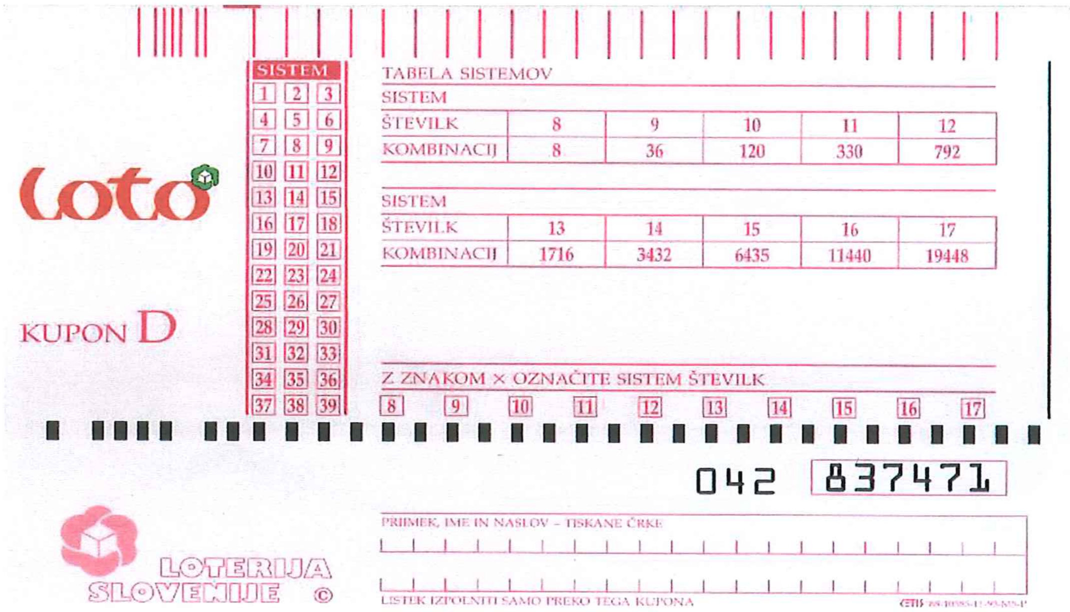
The winnings depend on the draw.
Each week 7 numbers are drawn.
If all the numbers are among the ones we crossed out we win a large amount of money. We can translate this problem into a problem involving the hyper-geometric distribution

Imagine you have balls numbered 1 - 39. You put them into an urn. When we cross the numbers on the Lottery ticket we paint those numbers black. The others we paint red. When 7 balls are drawn the number of correct guesses is $X \sim$ HiperGeom$(7, m, 39)$ where m is the number of balls we painted black.

<u>Figure</u>:

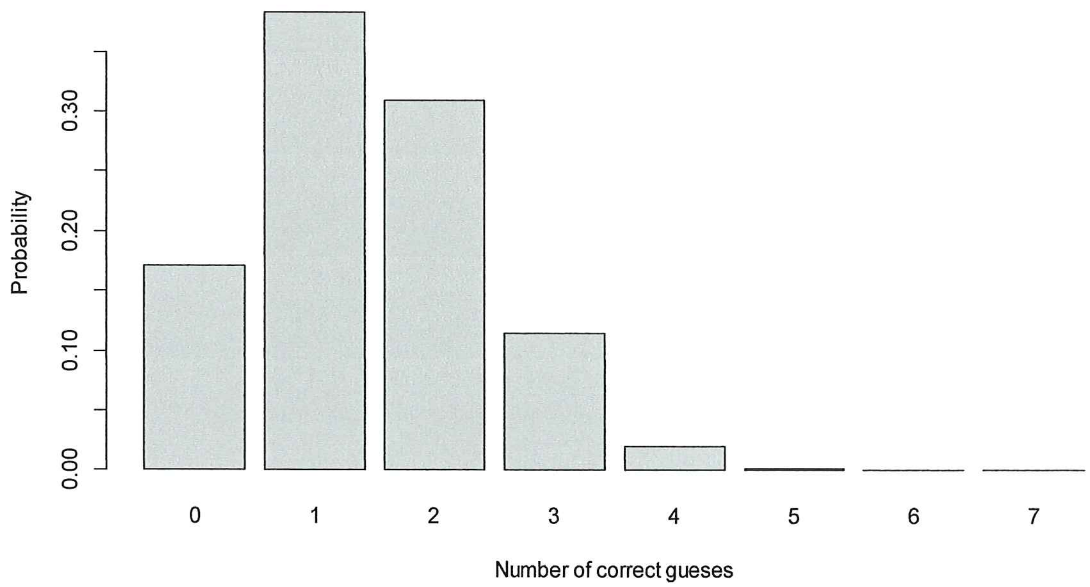


Below are distributions for correct number of guesses in Lottery for m=8,13,17. We translated the winning odds in Lottery to a question about the hyper-geometric distribution. The Lottery ticket looks like



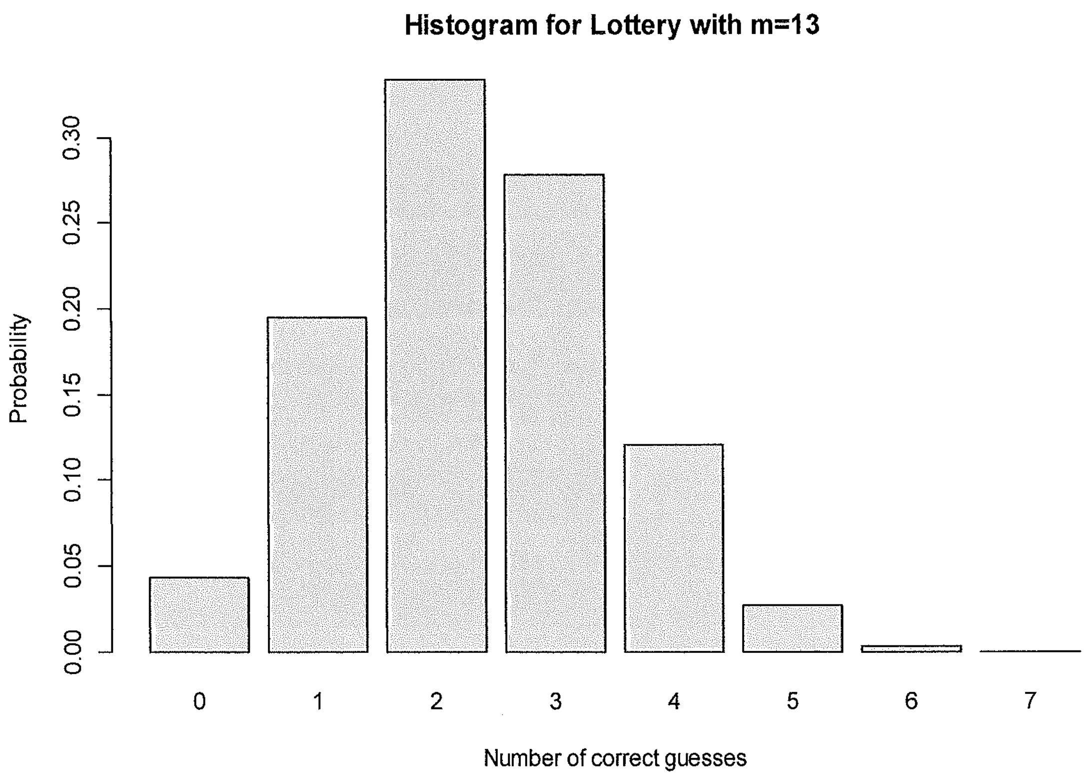
On the ticket the player can cross from m=8 to m=17 numbers. The number of correct guesses is the basis for determining the winnings. The correct guesses are a random variable X. The probability $P(X = 7)$ is the most interesting as it is the probability of jack-pot.

Histogram for Lottery with m=8



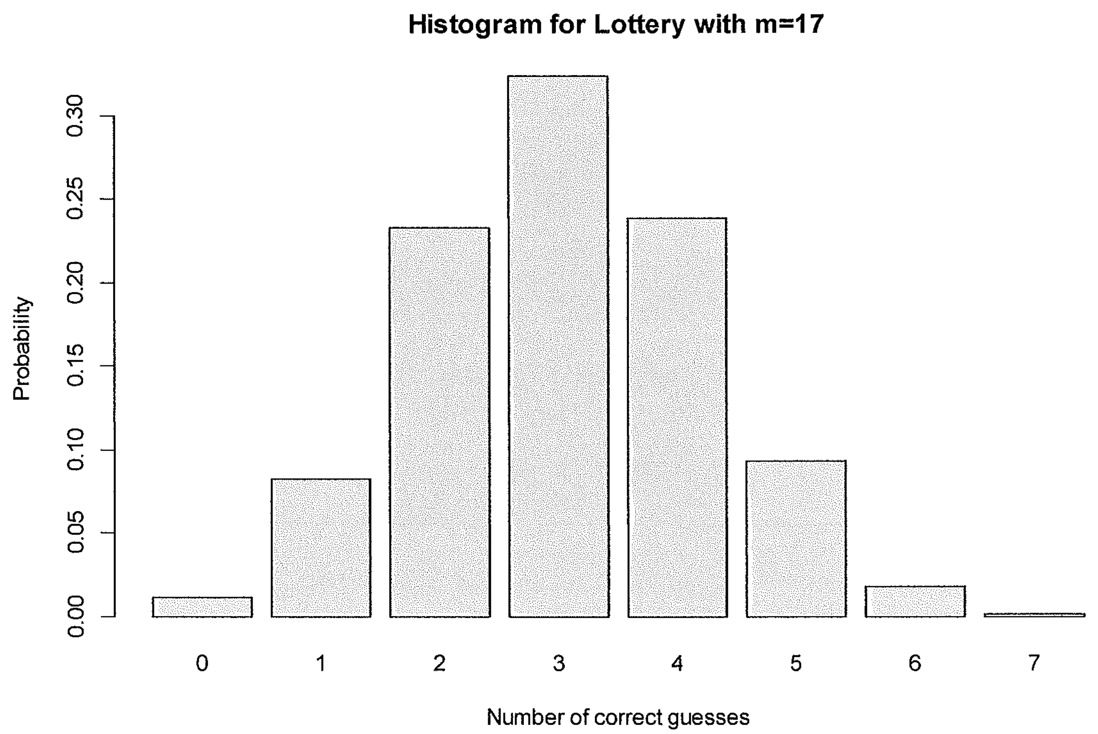
The numerical values of the above probabilities are:

1.709633e-01 3.829577e-01 3.093120e-01 1.145600e-01 2.045714e-02 1.693005e-03 5.643349e-05 5.201244e-07



The numerical values are:

0.0427672254 0.1945908757 0.3335843584 0.2779869653 0.1208638980 0.0271943770 0.0029007336 0.0001115667



Numerical values:

0.011088011 0.082467082 0.232848233 0.323400323 0.238294975 0.092935040 0.017701912 0.001264422

Geometric and negative binomial distribution

Suppose we toss a coin. The tosses are independent and the probability of heads is p.

Let X be the number of tosses until the first heads.

X is a random variable with values in $k = 1, 2, \ldots$

__Remark__: The values of X can be arbitrarily large. The event $\{X = k\}$ happens if we get $\underbrace{TT\ldots T}_{(k-1) \text{ times}} H$. By independence this implies

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

Definition: If

$$P(X=k) = (1-p)^{k-1} \cdot p$$

we say that X has geometric distribution with parameter p.

Shorthand: $X \sim \mathrm{Geom}(p)$.

Example: We play roulette and wait for the number 17 to appear. What is the probability that 17 will not appear in the first n tosses?

$$P(X>n) = P(\underbrace{TT\cdots T}_{n\text{-times}}) = (1-p)^n$$

Because $X \sim \mathrm{Geom}(1/37)$ this means

$$P(X>n) = \left(\frac{36}{37}\right)^n$$

If instead of waiting for heads we wait for the appearance of m heads then X is a random variable with values $k = m, m+1, \ldots$.

We have

$$P(X=k) = P(\{\text{exactly } m-1 \text{ heads in first } k-1 \text{ tosses}\} \cap \{\text{heads of } k\text{-th toss}\})$$

(independence)

$$= P(\{\text{exactly } m-1 \text{ heads in first } (k-1) \text{ tosses}\}) \times P(\text{heads on toss } k)$$

$$= \underbrace{\binom{k-1}{m-1} p^{m-1} \cdot (1-p)^{(k-1)-(m-1)}}_{\text{by binomial distribution}} \times p$$

$$= \binom{k-1}{m-1} p^{m} (1-p)^{k-m},$$

$$k = m, m+1, \ldots .$$

<u>Definition</u>: we say that X has negative binomial distribution with parameters m and p if

$$P(X = k) = \binom{k-1}{m-1} p^m (1-p)^{k-m},$$

$$k = m, m+1, \ldots .$$

<u>Shorthand</u>: $X \sim \mathrm{NegBin}(m, p)$.

Example: The Polish mathematician Stefan Banach (1897 – 1945) was a chain smoker. He always carried two boxes of matches in his pockets. Assume Banach starts with two boxes of n matches. Then he randomly reaches into the left or right pocket at random with probability $1/2$. At some stage Banach will take the last match from a box but will not notice it. The first time Banach pulls an empty matchbox from his pockets, the number of matches in the other box is random. Call it X. Possible values for X are $k = 0, 1, \dots, n$. We would like to compute the distribution of X.

Let us define

$A = \{X = k\} \cap \{$ Banach pulls the empty box from left pocket $\}$

By symmetry and the law of total probabilities we have

$$P(X=k) = 2 \cdot P(A).$$

Let us picture Banach cigarettes.

The event A happens when Banach lights his $(n + (n-k) + 1)$-st cigarette and at that point has pulled a box from his left pocket exactly $(n+1)$-st time.

Declare: reach into the left pocket = „success".

Poisson distribution.

Let us look at the binomial distribution $\mathrm{Bin}(n, \frac{\lambda}{n})$ for a given $\lambda > 0$. What happens if n is „large"? Let k be fixed. For $n \geq k$ we have for $X_n \sim \mathrm{Bin}(n, \frac{\lambda}{n})$ that

$$P(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k},$$

What happens when $n \to \infty$?

From Analysis we know that

$$\left(1 + \frac{x}{n}\right)^n \to e^x \quad \text{for} \quad x \in \mathbb{R}.$$

Rewrite

$$P(X_n = k) = \frac{\lambda^k}{k!} \cdot \overset{\nearrow 1}{\frac{n(n-1)\cdots(n-k+1)}{n^k}} \times$$

$$\times \overset{\nearrow e^{-\lambda}}{\left(1 - \frac{\lambda}{n}\right)^n} \times \overset{\nearrow 1}{\left(1 - \frac{\lambda}{n}\right)^{-k}}$$

We find that

$$\lim_{n \to \infty} P(X_n = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}.$$

This motivates the following definition

<u>Definition</u>: If $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$

for $\lambda > 0$ and $k = 0, 1, 2, \ldots$

we say that X has the Poisson distribution with parameter $\lambda > 0$.

Shorthand: $X \sim Po(\lambda)$.

<u>Remark</u>: From analysis we know that

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}.$$

If $X \sim Po(\lambda)$ we do get

$$\sum_{k=0}^{\infty} P(X = k) = 1.$$

2.2. Continuous distributions

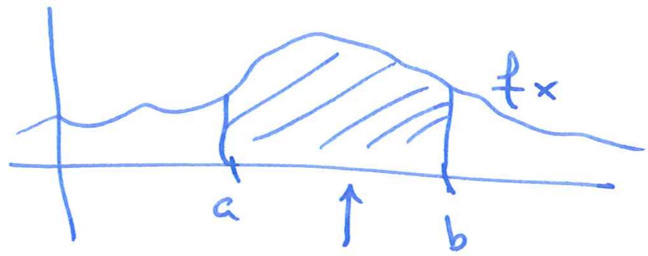
We can imagine „random numbers" that can take any real number as a value. Examples are lifetimes of components, a randomly chosen point on the interval $[0,1]$, ...

Technically, X is still a function on Ω and we require $X^{-1}((a,b])$ to be an event for all $a < b$.

<u>Definition:</u> The distribution of a random variable is given by the probabilities $P(X \in (a,b])$ for all $a < b$.

The idea of continuous random variables is to describe probabilities $P(X \in (a,b])$ by integrals of a single function.

<u>Figure</u>:



Area $= P(a < X \leq b)$

<u>Definition</u>: The random variable X has continuous distribution if there is a non-negative function $f_X(x)$ called the density such that

$$P(a < X \leq b) = \int_a^b f_X(x)\,dx$$

for all $a < b$.

Continuous distributions are used in financial modelling and statistics because of their practicality.

There are standard distributions that are often used.

Normal distribution

The random variable X has normal distribution with parameters μ a $\sigma > 0$ if the density is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Remark: For densities we must have $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

We believe mathematics that f_X is a density.

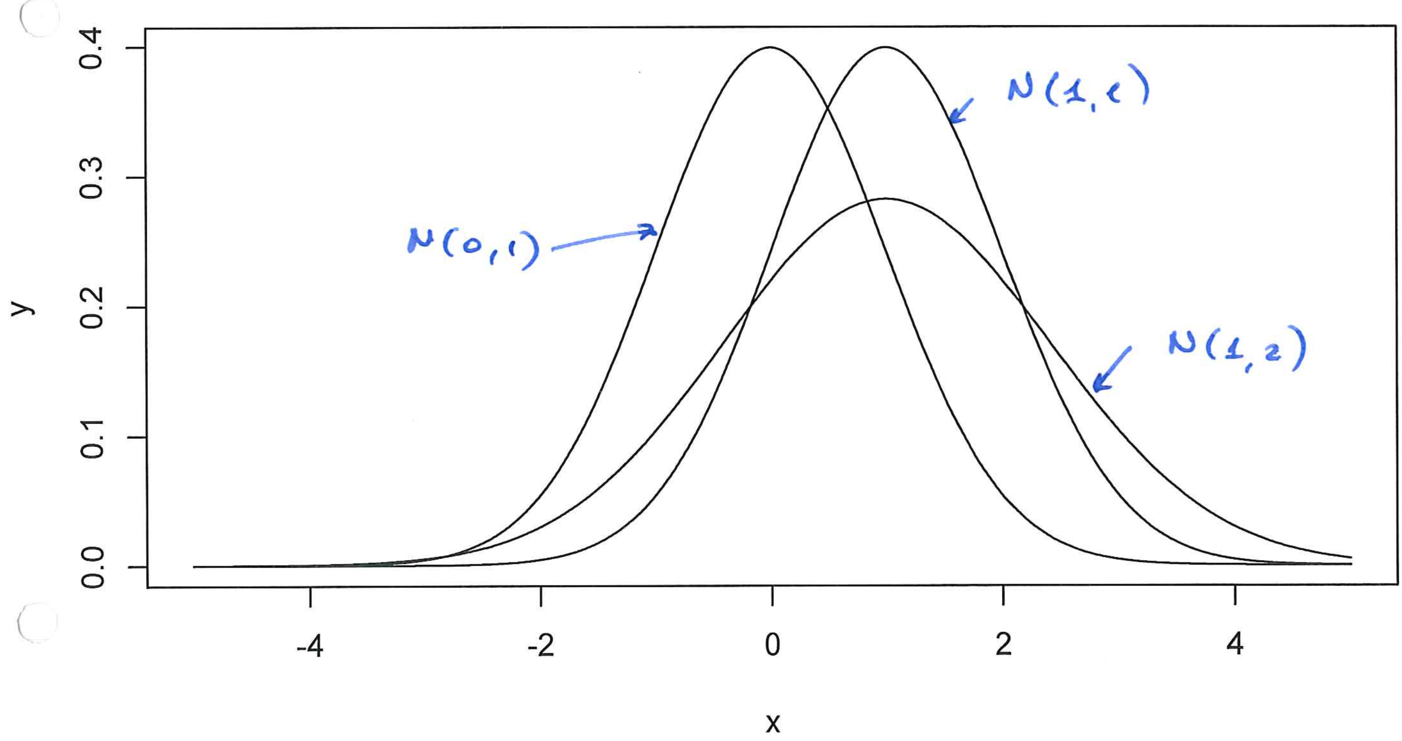
We will use the notation:

$$X \sim N(\mu, \sigma^2).$$

Remark: The name normal distribution was chosen by the Belgian statistician Adolphe Quetelet because the histograms of human characteristics such as IQ, height and others are close to normal.

Remark: We will give the right interpretation to parameters μ in σ^2 later.

X ~ N(0,1),N(1,1),N(1,3)



Exponential and gamma distribution

The density of the exponential distribution is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else.} \end{cases}$$

We say that X has the exponential distribution with parameter λ.

<u>Notation</u>: $X \sim \exp(\lambda)$

The exponential distribution is used to mode lifetimes of electronic components.

To define the gamma distribution recal the definition of the gamma function.

$$\Gamma(x) = \int_0^\infty u^{x-1} \cdot e^{-u} \, du.$$

The most important properties are:

(i) $\Gamma(x+1) = x \cdot \Gamma(x)$

(ii) $\Gamma(n) = (n-1)!$

(iii) $\Gamma(1/2) = \sqrt{\pi}$.

The density

$$f_X(x) = \begin{cases} \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, & x > 0 \\ 0 & \text{else.} \end{cases}$$

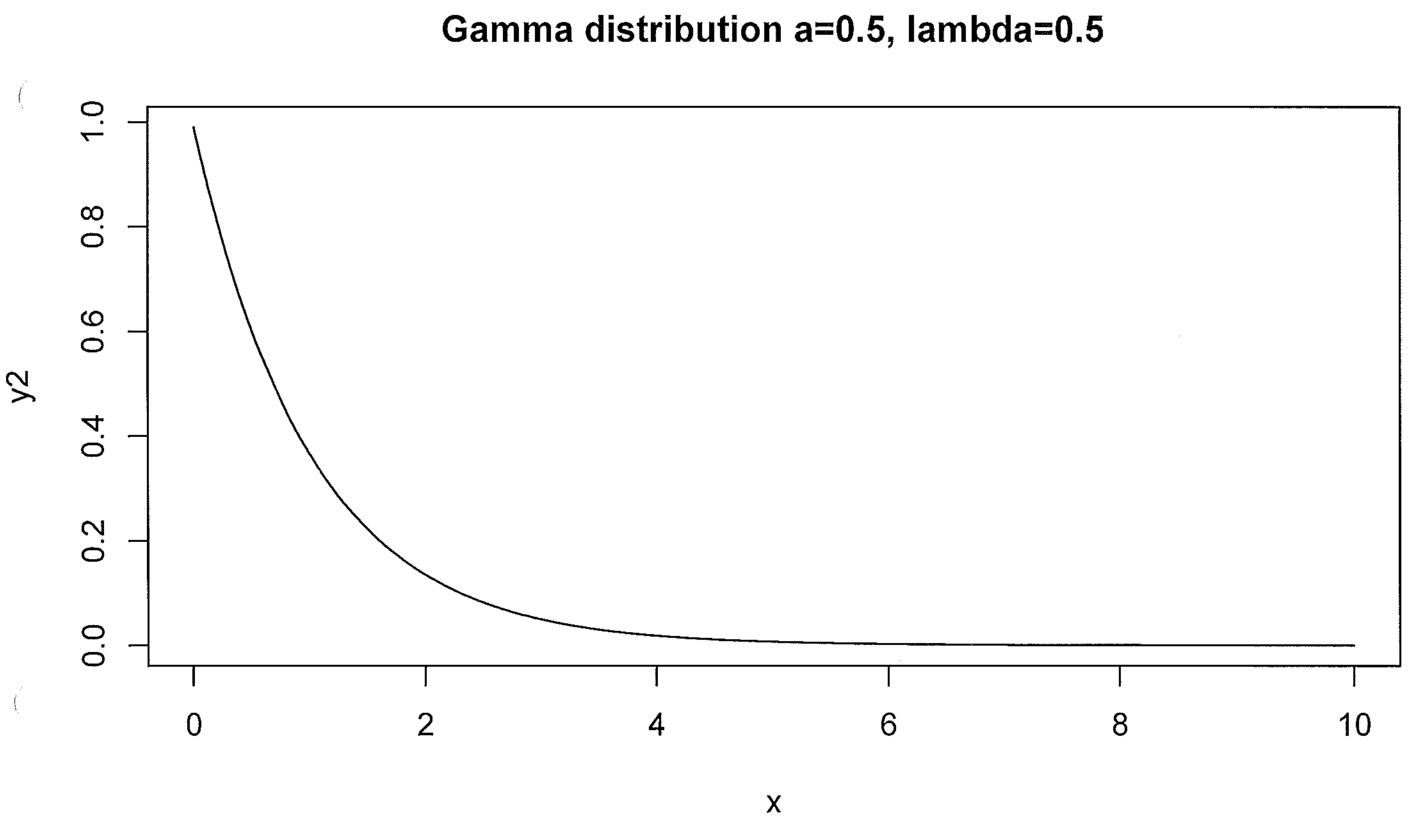
is called the gamma density with parameters a and λ.

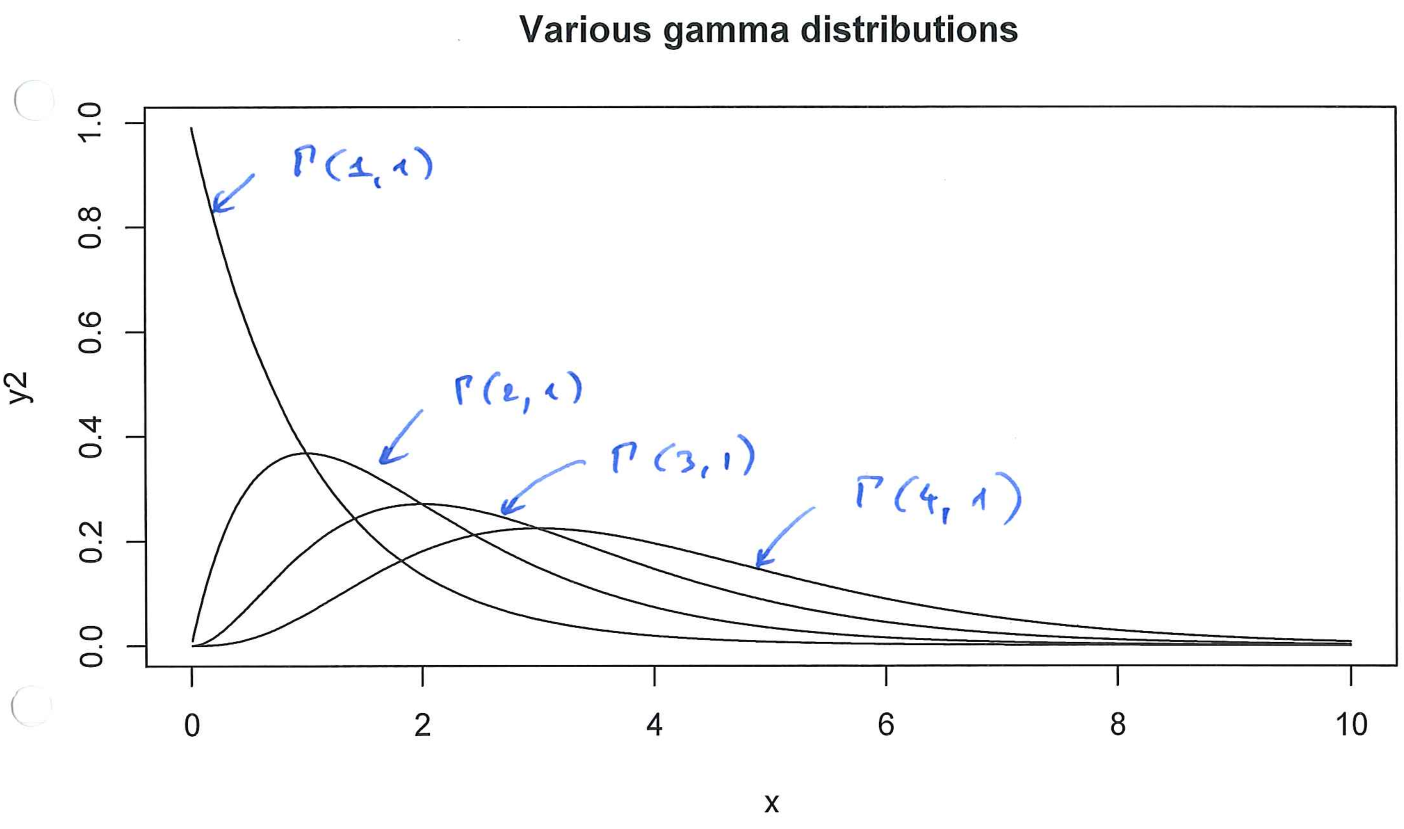
a : shape parameter

λ : scale parameter.

Remark: Sometimes $1/\lambda$ is used instead of λ.

Notation: $X \sim \Gamma(a, \lambda)$


Gamma distribution a=0.5, lambda=0.5
y2
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
x



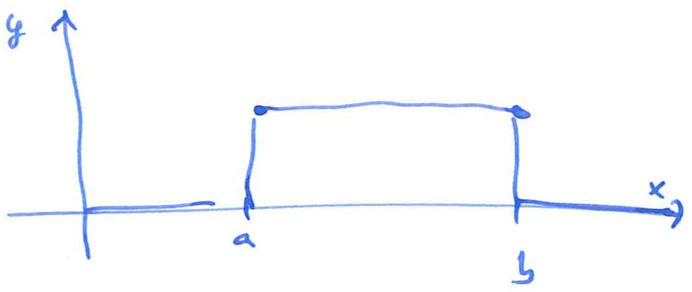
Various gamma distributions
Γ(1,1)
Γ(2,1)
Γ(3,1)
Γ(4,1)
y2
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
x

Uniform distribution

The uniform distribution models the choice of a point at random uniformly on the interval (a,b). The density is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b. \\ 0 & , \text{else.} \end{cases}$$

Figure :



We use the notation :

$$X \sim \mathcal{U}(a,b)$$

Most computer generated random numbers are uniform on $[0,1]$.

2.3. Functions of random variables

Let X be a random variable.

We have that $\{X \leq x\}$ is an event. So the probability is defined.

Definition: The distribution function of X is defined as the function

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

Theorem 2.1: Let X be a random variable with distribution function F_X.

(i) F_X is nondecreasing.

(ii) $\lim_{x \to \infty} F_X(x) = 1, \quad \lim_{x \to -\infty} F_X(x) = 0$

(iii) F_X is right continuous.

Proof :

(i) For $x < y$ we have $\{X \le x\} \subseteq \{X \le y\}$.

It follows that $P(X \le x) \le P(X \le y)$.

(ii) We have $\Omega = \bigcup_{n=1}^{\infty} \{X \le n\}$.

The sets in the union are increasing so

$$1 = P(\Omega) = \lim_{n\to\infty} P(X \le n)$$

$$= \lim_{n\to\infty} F_X(n)$$

The conclusion follows because F_X is nondecreasing. The other limit is proved similarly.

(iii) Fix $x \in \mathbb{R}$. Let $x_n \downarrow x$.

We have $\{X \le x\} = \bigcap_{n=1}^{\infty} \{X \le x_n\}$.

The sets $\{X \le x_n\}$ are decreasing.

It follows

$$P(X \le x) = \lim_{n\to\infty} P(X \le x_n) \quad \text{or}$$

$$F_X(x) = \lim_{n\to\infty} F_X(x_n)$$

The last statement is equivalent to right continuity.

If X has density $f_X(x)$ then

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du.$$

Inversely, if

$$F_X(x) = \int_{-\infty}^{x} g(u)\, du$$

for all $x \in \mathbb{R}$ and a nonnegative g then g is (a version of) the density. If g is continuous at x then

$$F_X'(x) = g(x) = f_X(x).$$

Example : Let $X \sim N(0,1)$ i.e.

the density of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Denote $F_X(x)$ by $\Phi(x)$, i.e.

$$\Phi(x) = \int_{-\infty}^{x} f_X(u) \, du.$$

Let $Y = aX + b$ for $a > 0$. What is the density of Y? We have

$$P(Y \le y) = P(aX + b \le y)$$

$$= P(aX \le y - b)$$

$$= P\left(X \le \frac{y-b}{a}\right)$$

$$= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \, du$$

New variable :

$$\frac{v-b}{a} = u \qquad \frac{dv}{a} = du$$

Example: Let $X \sim N(0,1)$ and $Y = X^2$. Density of Y? For $y > 0$

$$P(Y \le y) = P(X^2 \le y)$$
$$= P(-\sqrt{y} \le X \le \sqrt{y})$$
$$= \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$$

Comment: In general

$$\boxed{P(a < X \le b) = F_X(b) - F_X(a)}.$$

For continuous random variables the probabilities $P(X = x) = 0$ for all x and it is not relevant whether we write $<$ or $\le$. We have

$$P(Y \le y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}).$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-u^2/2} \, du$$

$$= \frac{2}{\sqrt{2\pi}} \int_{0}^{\sqrt{y}} e^{-u^2/2} \, du$$

New variable: $u^2 = v \Rightarrow$

$$2u\,du = dv$$

$$du = \frac{dv}{2\sqrt{v}}$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^y \frac{1}{2\sqrt{v}} e^{-v/2} dv$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^y \frac{1}{\sqrt{v}} e^{-v/2} dv$$

Since $P(Y \geq 0) = 1$ we have

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi} \cdot \sqrt{y}} e^{-y/2}, & y > 0 \\ 0, & \text{else.} \end{cases}$$

We recognize: $Y \sim \Gamma(1/2, 1/2)$.