University of St.Gallen

EFMD EQUIS ACCREDITED

Methods: Statistics (4,120)

4. Random Variables and Discrete Distributions

Spring 2021

Prof. Dr. Roland Füss

Swiss Institute of Banking and Finance (s/bf)

Contents

Learning Objectives

After that lecture, you know how:

- to calculate the **expected value** and **variance of** a discrete distribution by using the concept of random variables.
- discrete distributions can be visually represented by **probability functions** and **cumulative distribution functions.**
- and when certain discrete distributions should be applied in practice-oriented examples.

Literature

Levine, D.M., K. A. Szabat, and D.F. Stephan. (2016). *Business Statistics: A First Course*, 7th ed. United States: Pearson, **Chapter 5**.*

Stinerock, R. (2018). *Statistics with R*. United Kingdom: Sage. **Chapter 5**.*

Shira, Joseph (2012). *Statistische Methoden der VWL und BWL*, 4th ed. Munich et al.: Pearson Studium, **Chapter 9**.

Weiers, R. M. (2011). *Introductory Business Statistics*,7th ed., Canada: Thomson South-Western, **Chapter 6**.

*Mandatory literature

I. Random Variables

Starting point is the outcome (elementary events) of a random experiment:

- qualitative, e.g., "head" or "number" when tossing a coin
- quantitative, e.g., dice with the results 1, 2, 3, 4, 5, 6

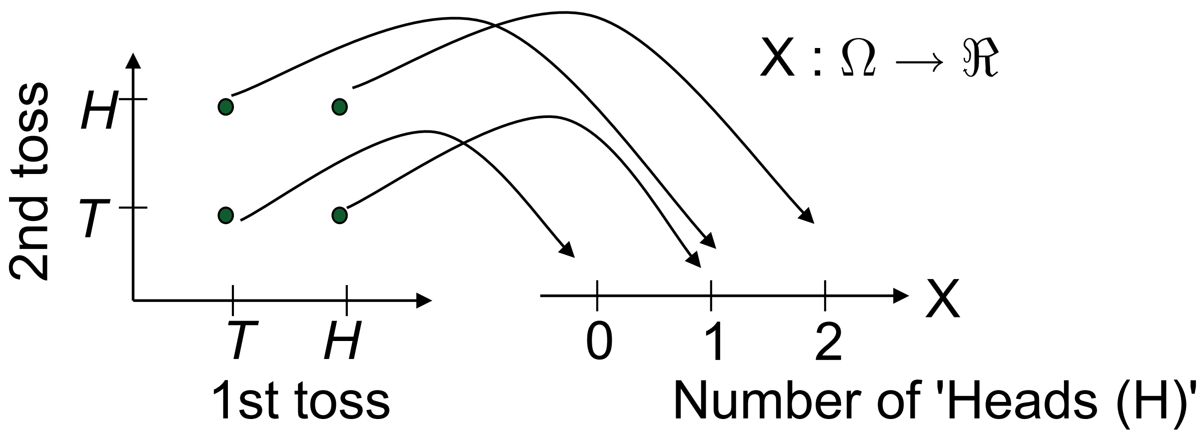
The **random variable** *X* is a function that assigns each outcome of a random experiment a numerical value. Formally, *X* is a mapping Ω to real numbers. The value that *X* takes is called **realization** (expressed in lowercase *x*).

Consider a random experiment with set Ω that collects all possible outcomes of the random experiment (e.g., throw a coin):

- observed outcome $\omega \in \Omega$ (either “heads (*H*)” or “tails (*T*)”)
- random variable *X* assigns the outcome a real number:
  e.g, $x(H) = 1$ and $x(T) = 0$

I. Random Variables (R-Example)

Two coin tosses, i.e., Ω = {*HH*, *HT*, *TH*, *TT*}.
We are interested in the "number of heads (*H*)"!



X is a mapping that assigns to each $\omega \in \Omega$ a real number:

$$\Omega = \{HH, HT, TH, TT\} \rightarrow X(\omega) \in \{0,1,2\}$$

I. Random Variables (R-Example)

Open the file "L4-Example_1.R" in R-Studio and reproduce the R-Code.

```
# throwing a coin twice:
# event space S={HH,HT,TH,TT}
# the random variable X is defined as the number of observed events "head" after 2 tosses.
# elementary events Sx={0,1,2}

x<-c(0,1,2) # vector x with elementary events
f<-c(1/4,1/2,1/4) # assigned probabilities

# mean
mu<-sum(x*f)
mu

# variance
sigma2<-sum((x-mu)^2*f)
sigma2

# standard deviation
sigma<-sqrt(sigma2)

# cumulative distribution function
F<-cumsum(f)
```

I. Random Variables

A **random variable** which maps outcomes to values of a:

1. countable* set is called **discrete**;
2. uncountable set is called **continuous**.

Intuition:

- A discrete random variable can take only "single" values (in particular, this involves the case that the sample space is finite).
- A continuous random variable can take **any** value in a continuum (real interval).

* A set is called **countable** if it has the same cardinality as some subset of the set of natural numbers.

II. Discrete Probability Functions

A **discrete random variable** X can take a value x_i with a probability:

$$0 \leq P(X = x_i) \leq 1,$$

$$\sum_i P(X = x_i) = 1$$

The function f, which assigns a probability to each x_i , $f(x_i) = P(X = x_i)$, is called **probability function** of the discrete random variable X (see Kolmogorov's Axioms). Only one value x_i can occur for a given experiment.

A discrete probability function illustrates all possible outcomes of an experiment, along with their corresponding probabilities.

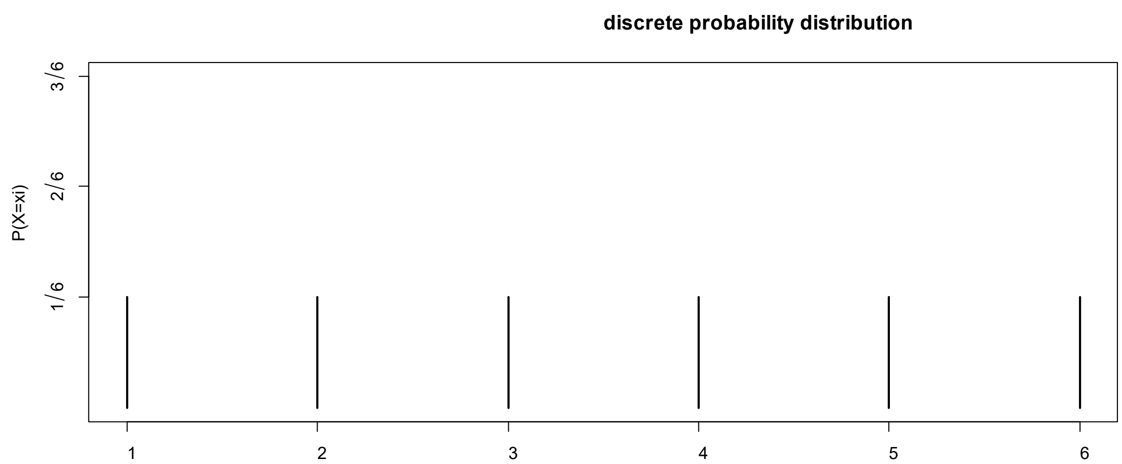
We define a probability function as the **relative** frequency distribution that should theoretically occur for observations from a given population.

III. Discrete Probability Function (Example)

Let X be the number of pips after rolling a fair dice. The **probability function** of X reads (discrete uniform distribution):

$$P(X = x_i) = 1/6 \quad i = 1, \dots, 6$$

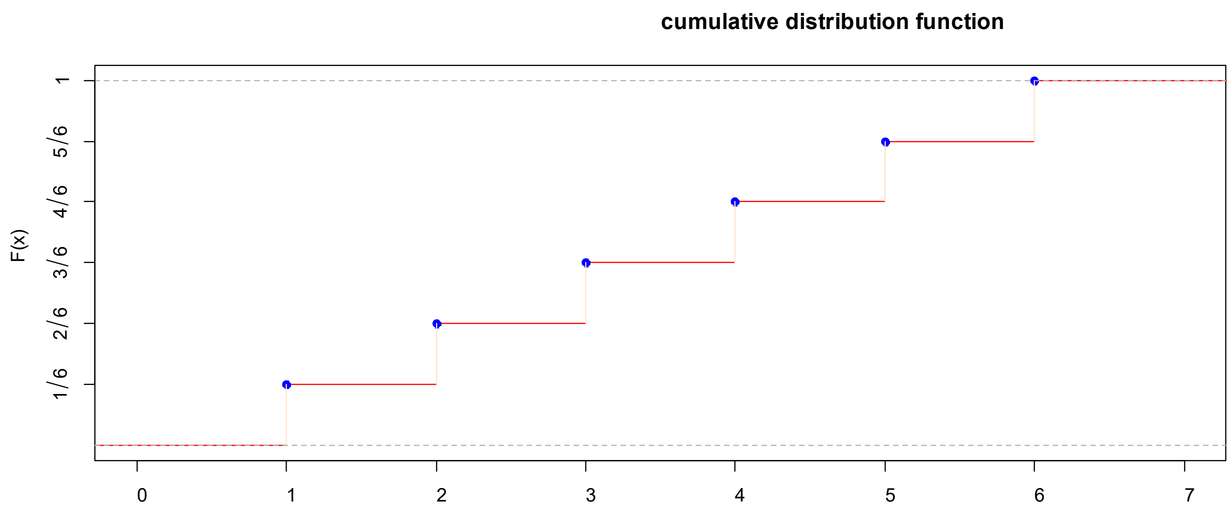
Probability functions are generally illustrated by **bar charts.**



III. Discrete Distribution Function (Example)

We are often interested in the probability that a random variable X takes values that are not greater than a specified value x. This is done by the (cumulative) **distribution function** $F(x)$ which is described by:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i).$$



IV. Expected Value and Variance of Probability Functions

Measures of central tendency and **dispersion** can be used to describe the discrete probability distributions of random variables

The **mean** or expected value $E(X)$ of a discrete random variable is defined as:

$$\mu = E(X) = \sum_i x_i \cdot P(X = x_i) = \sum_i x_i \cdot f(x_i)$$

The **variance** of a discrete random variable is defined as:

$$Var(X) = \sigma^2 = E\left[(X-\mu)^2\right] = \sum_i (x_i - \mu)^2 \cdot f(x_i)$$

IV. Expected Value and Variance of Probability Functions

The **variance formula** (with $E(X)=\mu$)

$$Var(X)=\sigma^2=E\left[(X-\mu)^2\right]=\sum_i (x_i-\mu)^2\cdot f(x_i)$$

can be rewritten as follows:

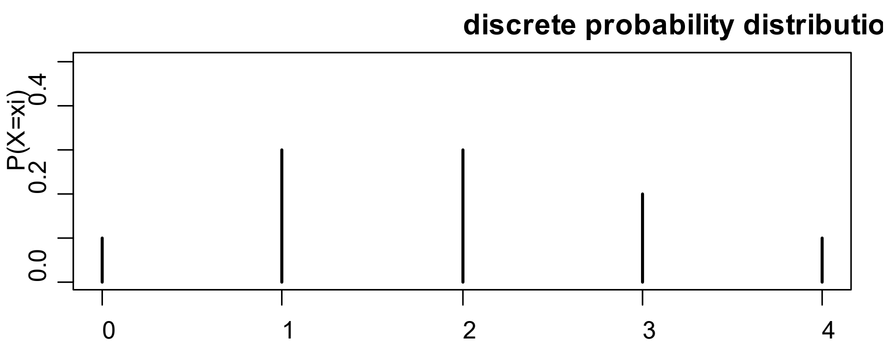
$$Var(X)=E(X^2)-[E(X)]^2=\sum_i x_i^2\cdot f(x_i)-\mu^2$$

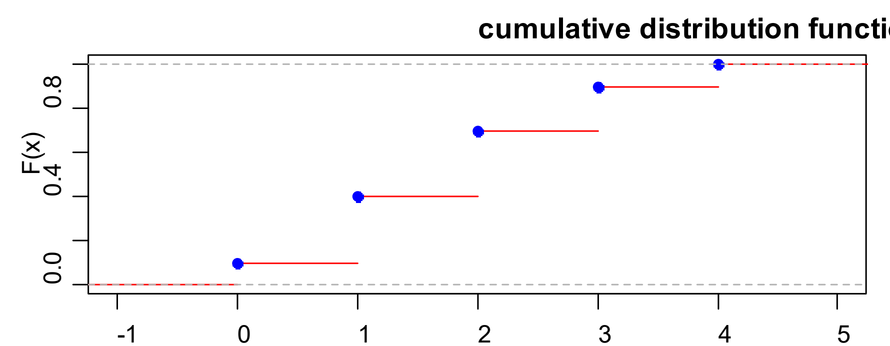
$$\begin{aligned}
Var(X)&=\mathrm{E}\left[\left(X-\mathrm{E}(X)\right)\right]^2\\
&=\sum_i\left(x_i-\mathrm{E}(X)\right)^2 f(x_i)\\
&=\sum_i x_i^2 f(x_i)+\sum_i\left[\mathrm{E}(X)\right]^2 f(x_i)-2\sum_i x_i\,\mathrm{E}(X)f(x_i)\\
&=\sum_i x_i^2 f(x_i)+\left[\mathrm{E}(X)\right]^2-2\,\mathrm{E}(X)\sum_i x_i f(x_i)\\
&=\mathrm{E}(X^2)+\left[\mathrm{E}(X)\right]^2-2\,\mathrm{E}(X)\mathrm{E}(X)\\
&=\mathrm{E}(X^2)-\left[\mathrm{E}(X)\right]^2
\end{aligned}$$

IV. Expected Value and Variance (R-Example)

Determine the expected value and variance of the following discrete probability distribution:

x_i	0	1	2	3	4
$P(xi)$	0.1	0.3	0.3	0.2	0.1





$$\mu = \mathrm{E}(X) = \sum x_i P(x_i) = 0(0.1) + 1(0.3) + 2(0.3) + 3(0.2) + 4(0.1) = 1.9$$

$$\sigma^2 = \mathrm{E}[(X-\mu)^2] = \sum (x_i - \mu)^2 P(x_i) = (0-1.9)^2(0.1) + (1-1.9)^2(0.3) +$$

$$+(2-1.9)^2(0.3) + (3-1.9)^2(0.2) + (4-1.9)^2(0.1) = 1.29$$

IV. Location and dispersion mass (R-Example)

Open the file "L4-Example_2.R" in R-Studio and reproduce the R-Code, which leads to the previous graphs of the distributions.

```
# the following elementary events in a random experiment are given:
events <- c(0,1,2,3,4)
# events can occur with the following probabilities:
prob <- c(0.1,0.3,0.3,0.2,0.1)

# mean
mu<-sum(events*prob)
mu

# variance
sigma2<-sum((events-mu)^2*prob)
sigma2

# standard deviation
sigma<-sqrt(sigma2)
sigma


# discrecte probability density function:
names(prob) <- events
prob <- as.table(prob)
op <- par(mfrow = c(2, 1), mgp = c(1.5, 0.8, 0), mar =  .1+c(3,3,2,1))
plot(prob,main="Discrecte Probability Density Function",xlab="",ylab="P(X=xi)",ylim=c(0,0.5))

# cumulative distribution function
F1 <- cumsum(prob)
plot(F1, main="Cumulative Distribution Function",xlab="",ylab="F(x)", type = "s")
par(op)
```

V. Chebyshev Inequality

Chebyshev's inequality provides a **minimum value** for the probability that the random variable takes realizations within an interval symmetric around the mean (in terms of multiples of a standard deviation). Given a random variable X with expected value μ and standard deviation of σ it holds that:

$$P\left(\left|X-\mu\right|<k\sigma\right)\geq 1-\frac{1}{k^2}$$

$$P\left(\left|X-\mu\right|\geq k\sigma\right)\leq \frac{1}{k^2}$$

Applies **regardless** of the distribution of X!

Specifies a very **conservative lower limit**!

V. Chebyshev Inequality (Example)

A random variable X has the expected value $\mu = 6$ and the variance $\sigma^2 = 9$. There is no further information available on the distribution of X. A meaningful **upper** limit for the probability that a realization of X lies **outside** the interval from -3 to 15 is found by a reformulation of the earlier given form of Chebyshev's inequality:

$$P\left(|X - \mu| > k\sigma\right) \leq \frac{1}{k^2}$$

In this example, the interval is $[-3;\ 15]$ symmetrical around the expected value. It must have the value $k = 3$, so that the interval limits are defined as $\mu - k\sigma = -3$ and $\mu + k\sigma = 15$. It follows:

$$P\left(X \notin [-3;15]\right) \leq \frac{1}{3^2} = \frac{1}{9}$$

VI. Binomial Distribution

The Binomial distribution is based on the **Bernoulli experiment**, which is characterized by the following properties:

- The experiment consists of a **sequence** of consecutive trials.
- In each trial, there are **two possible** outcomes *A* and *B*, with the probabilities π and $(1-\pi)$, respectively.
- The individual trials are statistically **independent** of each other (e.g., sequence of *n* coin tosses).

Then, the probability of *x* times occurring the event *A*, and thus, $(n-x)$ times the event *B* in *n* independent repetitions of the Bernoulli-experiment, is determined with the help of the **Binomial distribution**.

VI. Binomial Distribution

Multiplication Rule (independence of events):

- Probability of x times occurring A:

  $P(\text{‚}x\text{-times A‘}) = P(\text{A})\cdot P(\text{A})\cdot \ldots \cdot P(\text{A}) = (P(\text{A}))^x = \pi^x$
- Correspondingly, the probability of the occurrence B:

  $\text{‚}(n - x)\text{-times B‘} = (1 - \pi)^{n-x}$

Due to the multiplication rule for independent events, the following applies to the occurrence of a **particular sequence**:

$$P(\text{‚}x\text{-times A‘; ‚}(n - x)\text{-times B‚}) = \pi^x (1 - \pi)^{n-x}$$

with $\binom{n}{x}$ possible sequences, in which event A happens x times.

VI. Binomial Distribution

An experiment, in which either A occurs with probability *P(A)* = π or B with probability *P(B) = 1 -* π*,* is independently repeated *n* times.

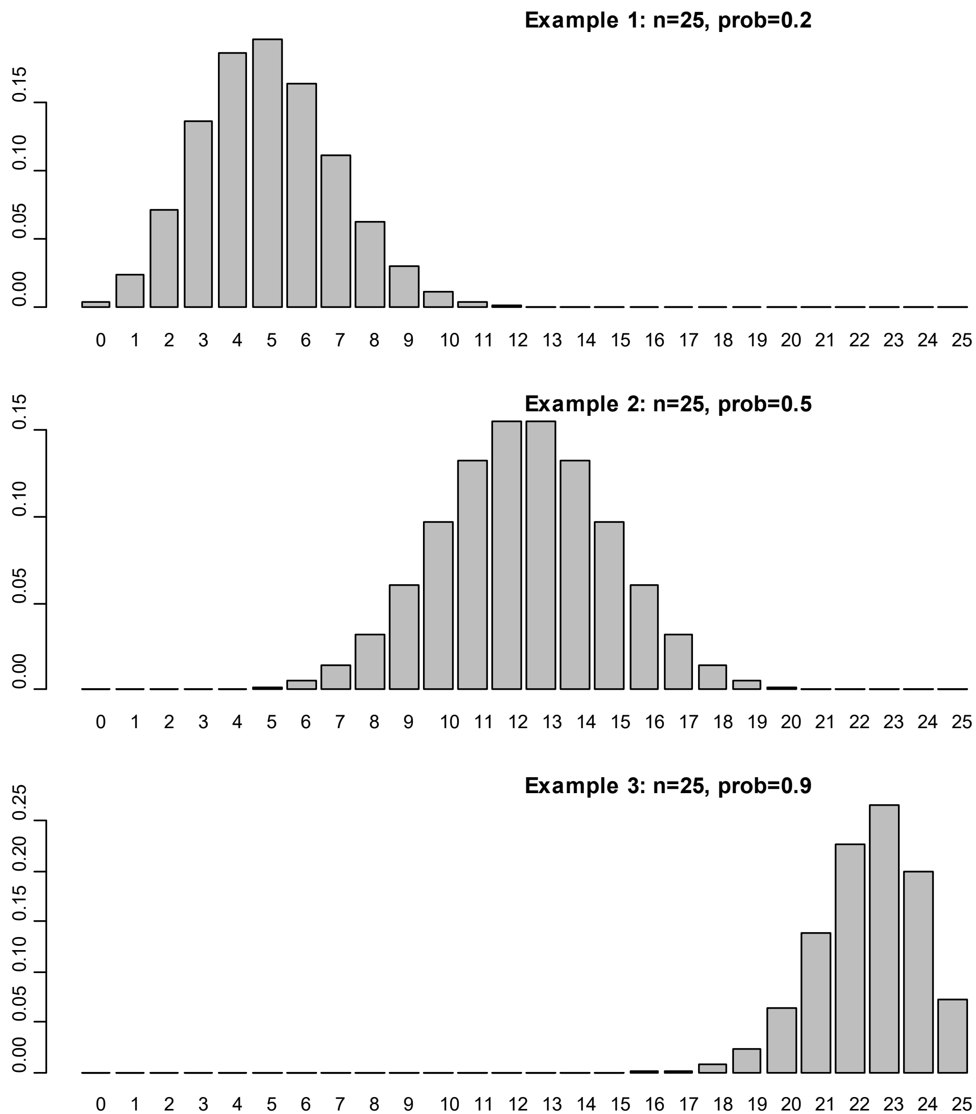
The random variable *X* states the number of random experiments resulting in *A*, and follows a Binomial distribution. This distribution has the following probability function:

$$P(X = x) = \frac{n!}{x!(n-x)!}\pi^{x}(1-\pi)^{n-x} \quad = \quad \binom{n}{x}\pi^{x}(1-\pi)^{n-x}$$

Expected value: $\mu = E(x) = n\pi$

Variance: $\sigma^2 = E[(x-\mu)^2] = n\pi(1-\pi)$

VI. Binomial Distribution (Example)



VI. Binomial Distribution (Example)

The proportion of women in a lecture is estimated to be 30%. Suppose 3 people are randomly selected from this lecture and *X* is the number of women in this sample. What is $P(X=0)$, $P(X=1)$, $P(X=2)$, $P(X=3)$?

$$P(X=3)=\frac{3!}{3!(3-3)!}(0.3)^3(0.7)^{3-3}=0.027$$

$$\boxed{\begin{aligned}&P(X=x)\\&=\frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x}\\&=\binom{n}{x}\pi^x(1-\pi)^{n-x}\end{aligned}}$$

$$P(X=2)=\frac{3!}{2!(3-2)!}(0.3)^2(0.7)^{3-2}=0.189$$

$$P(X=1)=\frac{3!}{1!(3-1)!}(0.3)^1(0.7)^{3-1}=0.441$$

$$P(X=0)=\frac{3!}{0!(3-0)!}(0.3)^0(0.7)^{3-0}=0.343$$

VI. Binomial Distribution (R-Example)

Open the file "L4-Example_3.R" in R-Studio and reproduce the R-Code.

```
# example: lottery
#-------------------------------------
# How likely is it to win a lottery where 6 correct numbers need to be drawn from 49?
p <- 1/(factorial(49)/(factorial(6)*factorial(49-6)))
# check:
p <- 1/choose(49,6) # same result


# example: Women's share in the lecture
#-------------------------------------
# the percentage of women in the lecture is about 30%.
# suppose 3 people are chosen  randomly.
# let the random variable Y be the number of women in the sample.
# determine the probability P(y=0),P(y=1),P(y=2),P(y=3)!

p<-0.3
n<-3
x<-0:n

px<-choose(n,x)*p^x*(1-p)^(n-x)
barplot(px,names=as.character(x),main="Example 1: n=3, p=0.3")

p0<-choose(3,0)*p^0*(1-p)^3
p1<-choose(3,1)*p^1*(1-p)^2
p2<-choose(3,2)*p^2*(1-p)^1
p3<-choose(3,3)*p^3*(1-p)^0

# same result with the specific R-funtion
dbinom(3,size=3,prob=0.3)
dbinom(2,size=3,prob=0.3)
dbinom(1,size=3,prob=0.3)
dbinom(0,size=3,prob=0.3)
```

VII. Hypergeometric Distribution

The **Hypergeometric distribution** models the number of successes in a given number of consecutive trials. However:

- The consecutive trials are **NOT independent** of each other.
- The probability of success **changes** from one trial to the next (selected objects are not returned to population).

Binomial distribution: sampling *with* replacement

Hypergeometric distribution: sampling *without* replacement

VII. Hypergeometric distribution

A discrete random variable *X* with the probability function:

$$P(X=x)=\frac{\binom{s}{x}\binom{N-s}{n-x}}{\binom{N}{n}}$$

is distributed **hypergeometrically**, with

- the population size *N,*
- the sample size *n,*
- the number of successes *s* in the population,
- the number of successes *x* in the sample.

VII. Hypergeometric distribution

The probability function is only defined for $x \leq S$ and $n - x \leq N - s$

$$P(X=x) = \frac{\binom{s}{x}\binom{N-s}{n-x}}{\binom{N}{n}}$$

Expected value: $\mu = E(X) = \frac{ns}{N}$

$\frac{s}{N}$ Probability of success for the first trial

Variance: $\sigma^2 = E\left[(X-\mu)^2\right] = \frac{ns(N-s)}{N^2} \cdot \frac{(N-n)}{(N-1)}$ with correction factor $\frac{(N-n)}{(N-1)}$

VII. Hypergeometric Distribution (Example)

Processors: Consider the quality control of a firm for a shipment of 250 processors. In the shipment, there are 17 defective processors. The quality control collects 5 processors for inspection. Let the random variable *X* be the number of defective processors in the sample. What is the probability of exactly 3 defective processors in the sample?

$$P(X=3)=\frac{\binom{17}{3}\binom{250-17}{5-3}}{\binom{250}{5}}=0.00235$$

VII. Hypergeometric Distribution (R-Example)

Open the file "L4-Example_4.R" in R-Studio and reproduce the R-Code.

```
# example: black/white balls
#--------------------------------------
# a jar is filled with 7 white balls and 5 black balls.
# take a sample (without putting back) and draw a ball K times from the jar with M white balls and N black balls
# random variable X: number of white balls in the sample

# X~hyper(m=M,n=N,k=K)
# valid for K<=M and K<=N

# randomly draw 4 balls from the jar without putting them back
# -> sample of  K=4
# -> M=7 white balls
# -> N=5 black balls

# probability of drawing 3 white balls
dhyper(3,m=7,n=5,k=4)


# example: processors
#--------------------------------------
# assume there are 17 damaged processors in a shipment of 250 processors.
# the quality control department will take 5 processors for inspection.
# the random variable X corresponds to the number of damaged processors in the sample.

# probability to find exactly 3 damaged processors in the sample (P(X=3))?
dhyper(3,m=17,n=233,k=5)

# probability to find exactly 5 damaged processors in the sample (P(X=5))?
dhyper(5,m=17,n=233,k=5) #very small chance

# Probability to find a maximum of 2 damaged processors in the sample P(X<=2)?
# P(X<=2)=P(X=0,1,2)
dhyper(0,m=17,n=233,k=5)+dhyper(1,m=17,n=233,k=5)+dhyper(2,m=17,n=233,k=5)
phyper(2,m=17,n=233,k=5) # P(X<=2)

# P(X>1)=1-Prob(X<=1)=1-F(1)
phyper(1,m=17,n=233,k=5,lower.tail=FALSE) # upper.tail-Option fuer (P(X>x)
```

VIII. Poisson Distribution

The Poisson distribution is applied to events for which the probability of occurrence over a given span is **extremely small**.

In numerous applications, which formally base on a Bernoulli experiment, the probability that an event *A* may occur in a single experiment is very low. In such cases it is beneficial to approximate the Binomial distribution by a distribution that results when $\pi \to 0$ and n $\to\infty$, in such a way that the expected value $\lambda = n\pi$ stays constant. It holds:

$$\lim_{\pi \to 0} \binom{n}{x} \pi^x (1-\pi)^{n-x} = \frac{(n\pi)^x}{x!} e^{-n\pi} \qquad n\pi = \lambda \text{ is constant}$$

VIII. Poisson Distribution

A discrete random variable *X* with the probability function

$$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$$

Is called **Poisson-distributed** with the parameter λ. $P(X=x)$ is the probability that the event will occur exactly *x times* within a given time period. The expected value and the variance of a Poisson distribution are equal:

Expected value = Variance = λ

The Poisson distribution is usually used when *n* and π are not known individually, but only the mean value $\lambda = n\pi$ and π is very small (< 0.05 according to Weiers).

VIII. Poisson distribution (Example)

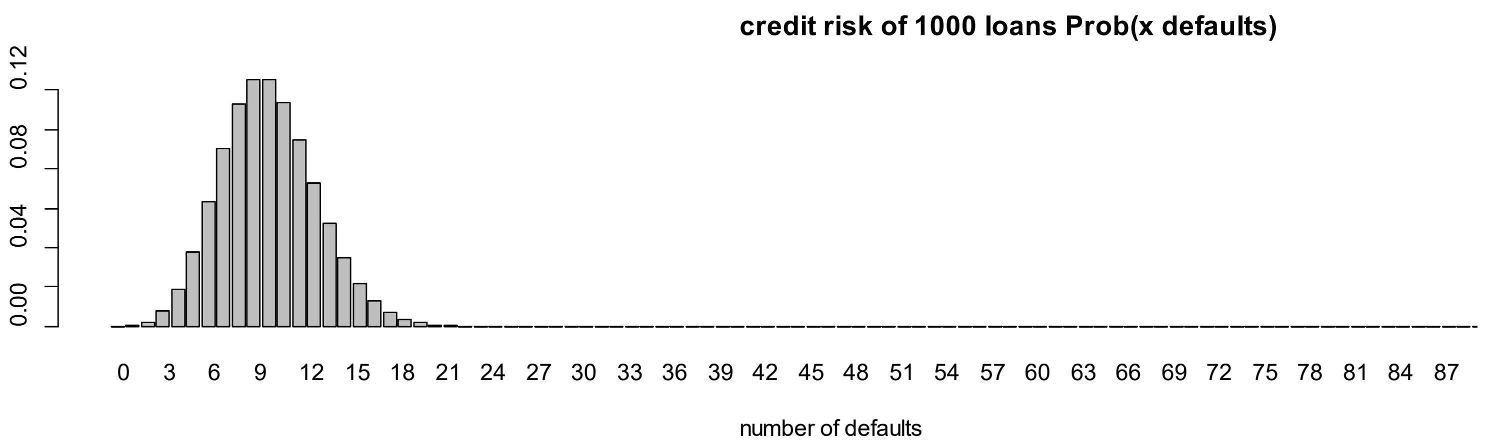
Railroad Accidents: Average number (resp. long-time mean) of railroad accidents per year is $\lambda = 1.25$. What is the probability of 2 accidents in 2020?

$$P(X = 2) = \frac{\lambda^2 e^{-\lambda}}{2!} = \frac{1.25^2 e^{-1.25}}{2!} = 0.224$$

VIII. Poisson Distribution (R-Example)

Credit risk: Consider a financial institution that holds a portfolio of 1000 loans. For small default probabilities of the loans, the frequency of default risk can be approximated by a Poisson distribution. For a portfolio of 1000 loans and an independent default probability of 1% ($\lambda = n \cdot \pi = 10$). The probability of *x* failures can be modelled as follows:

$$P(x) = \frac{10^x e^{-10}}{x!}$$



VII. Poisson Distribution (R-Example)

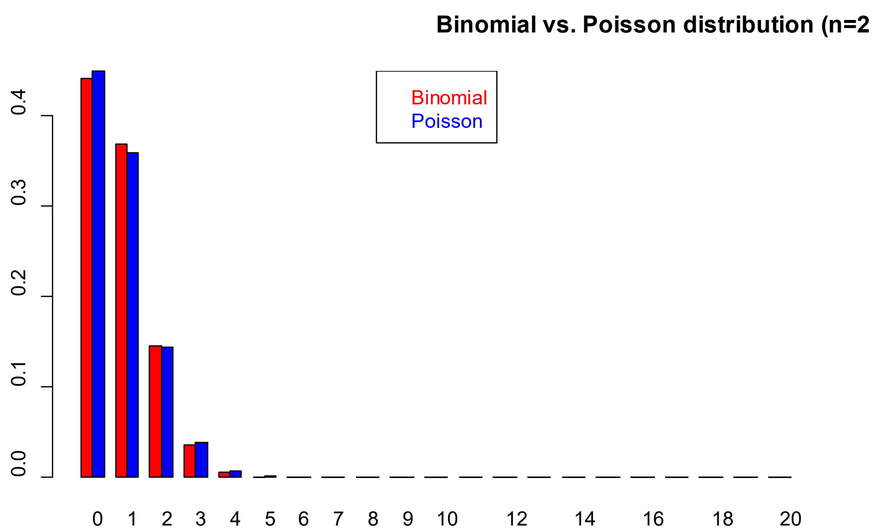
Open the file "L4-Example_5.R" in R-Studio and reproduce the R-Code for the previous example.

```
# example: credit risk
#--------------------------------------
# A financial institution holds a portfolio of 1000 loans. in the case of low default probabilities,
# the distribution can be approximated by a Poisson distribution given the portfolio of
# 1000 loans and an independent probability of default of 1% (lambda=n*p=10).
p<-0.01
n<-1000
np<-n*p
x<-0:n
px2<-(np^x*exp(-np))/factorial(x)
# displaying default risk in a graph.
barplot(px2,beside=T,names=as.character(x),main="Default Risk of 1000 Loans: Prob(x Defaults)",
        xlab="Number of Defaults",xlim=c(0,100))


# example: car wash
#--------------------------------------
# a car wash has the following opening hours: 08:00 to 18:00
# the random variable Y is given by the number of customers appearing during this period.

# since the period covers 10 hours, we model the following distribution:
# Y~pois(lambda=5*10=50)

# determine the probability of 48 to 50 customers (including) using the car wash.
# we want to know P(48<=Y<=50)=P(X<=50)-P(X<=47).
diff(ppois(c(47,50),lambda=50))
```

VIII. Approximation of the Binomial Distribution

The Poisson distribution is usually used when n and π are not known individually, but the mean value $\lambda = n\pi$ is known.

In practice, the Binomial distribution can be approximated by the Poisson distribution for sufficiently **large** n and sufficiently **small** π (rule of thumb by Weiers: $n \geq 20$, $\pi \leq 0.05$ *or by* Schira: $n \geq 100$, $\pi \leq 0.10$).



Appendix: Binomial Coefficient and Pascal's Triangle

The values of the binomial coefficient can be subsequently derived from **Pascal's triangle.** The values on the side are 1, all other values are obtained as the sum of the two values to the left and right above:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

The sum of each row is a power of two (2^n):

n																				**Sum**
0										1										*1*
1									1		1									*2*
2								1		2		1								*4*
3							1		3		3		1							*8*
4						1		4		6		4		1						*16*
5					1		5		10		10		5		1					*32*
6				1		6		15		20		15		6		1				*64*
7			1		7		21		35		35		21		7		1			*128*
8		1		8		28		56		70		56		28		8		1		*256*
9	1		9		36		84		126		126		84		36		9		1	*512*

Appendix: Binomial Distribution (Example)

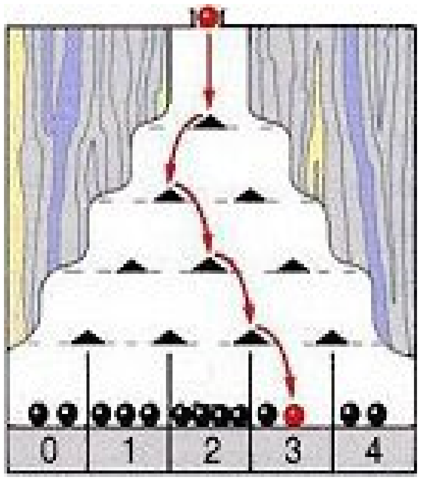
Galton Board: How many balls fall into the containers 0 to 4?

Possible paths (16 in total):

- Bin 0: {L, L, L, L}
- Bin 4: {R, R, R, R}
- Bin 1: {L, L, L, R}, {L, L, R, L}, {L, R, L, L}, {R, L, L, L}
- Bin 3: {R, R, R, L}, {R, R, L, R}, {R, L, R, R}, {L, R, R, R}
- Bin 2: {L, L, R, R}, {L, R, L, R}, {L, R, R, L}, {R, L, R, L}, {R, L, L, R}, {R, R, L, L}

Probability $p(B_k)$ *that* a ball will fall into bin k.

- $p(B_0) = p(B_4) = 1/16$,
- $p(B_1) = p(B_3) = 4/16$,
- $p(B_2) = 6/16$



Appendix: Binomial Distribution (Example)

1. Consider now the general case of a board with *n* stages, then there are *n + 1* bins (0 to *n*).
2. Probability $p(B_k)$ that a ball falls into bin k ($k = 0, \ldots, n$):

$$p(B_k) = \frac{\text{Number of paths to bin } k}{\text{Total number of paths}}$$

3. Number of paths to container *k* : The number of possible combinations to place *k* objects on *n* fields ($n \geq k$) <u>without</u> consideration of the order is

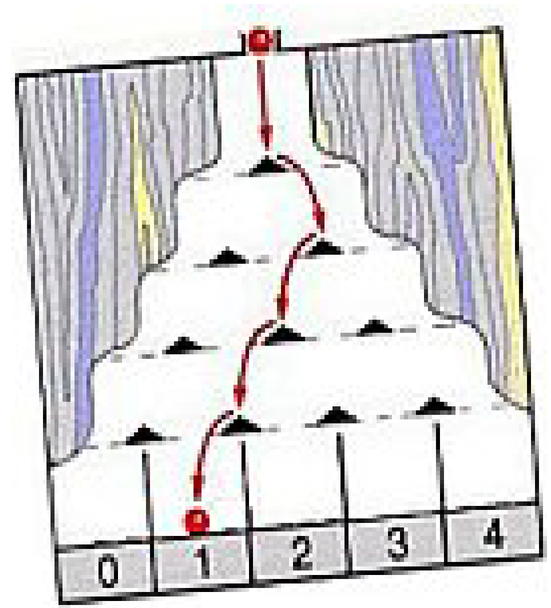
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$\Rightarrow p(B_k) = \binom{n}{k} \Big/ 2^n \qquad \text{as } \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

Appendix: Different Probabilities

1. The probability that a ball will fall into each bin to the left and to the right is no longer 0.5
2. Let $p(right) = \pi$ and $p(left) = 1 - \pi$

$$\Rightarrow p(B_k) = \binom{n}{k} \cdot \pi^k \cdot (1-\pi)^{n-k}$$



Appendix: Moments of Binomial Distribution

Expected value: $\mu = E(X) = n\pi$

Variance: $\sigma^2 = E[(X-\mu)^2] = n\pi(1-\pi)$,

wherein $X \sim B(n, \pi)$

Derivation of the expected value:

$$X = X_1 + X_2 + \ldots + X_n$$

$$X_i = \begin{cases} 1 & \text{if the result in the } i\text{-th trial is a success} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{E}\left[X_i\right] = 1 \cdot \pi + 0 \cdot (1-\pi) = \pi$$
$$\mathrm{E}\left[X\right] = \mathrm{E}\left[X_1 + X_2 + \ldots + X_n\right] = \mathrm{E}\left[X_1\right] + \mathrm{E}\left[X_2\right] + \ldots + \mathrm{E}\left[X_n\right] = n\pi$$

Appendix: Moments of Binomial Distribution

Derivation of variance:

$$E\left[X_i^2\right] = 1^2 \cdot \pi + 0^2 \cdot (1-\pi) = \pi$$

$X_i, X_j (i \neq j)$ are independent random variables, i.e.,

$$P(X_i = r, X_j = s) = P(X_i = r) \cdot P(X_j = s) \xrightarrow{\text{it follows}} E\left[X_i \cdot X_j\right] = E\left[X_i\right] \cdot E\left[X_j\right]$$

$$Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2 = E[(X_1 + X_2 + ... + X_n)^2] - (n\pi)^2$$

$$= E\left[\sum_{i=1}^{n} X_i^2 + \sum_{j=1}^{n} \sum_{\substack{i=1 \\ i \neq j}}^{n} X_i X_j\right] - (n\pi)^2 = \sum_{i=1}^{n} E[X_i^2] + \sum_{j=1}^{n} \sum_{\substack{i=1 \\ i \neq j}}^{n} E[X_i X_j] - (n\pi)^2$$

$$= \sum_{i=1}^{n} E[X_i^2] + \sum_{j=1}^{n} \sum_{\substack{i=1 \\ i \neq j}}^{n} E[X_i] \cdot E[X_j] - (n\pi)^2$$

$$= n\pi + n(n-1)\pi \cdot \pi - (n\pi)^2 = n\pi + (n\pi)^2 - n\pi^2 - (n\pi)^2$$

$$= n\pi(1-\pi)$$

Appendix: Derivation of the Poisson Distribution

Relation between Binomial and Poisson distribution

The value of a Poisson-distributed random variable for a given *k* is the limit $n \to \infty$ of a binomial distribution with $p = \lambda / n$ for the parameter *k*.

$$\lim_{n\to\infty} P(X=k) = \lim_{n\to\infty} \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n\to\infty} \left(\frac{\lambda^k}{k!}\right)\left(\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \frac{\lambda^k}{k!}\lim_{n\to\infty} \underbrace{\left(\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\cdots\frac{n-k+1}{n}\right)}_{\to 1} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1}$$

$$= \frac{\lambda^k e^{-\lambda}}{k!}$$