Probability and Statistics
Review with R

Raquel Menezes
rmenezes@math.uminho.pt

Department of Mathematics and Applications

University of Minho, Portugal

October, 2020

Syllabus

Probability and Distributions

- Definitions and theorems
- Discrete and continuous random variables
- Binomial, Poisson, Exponential and Gaussian distributions
- Central limit theorem

Experiments

Our first step in constructing a mathematical model for probability studies is to describe the type of experiments on which probability studies are based.

Some types of experiments **do not yield the same results, no matter how carefully they are repeated under the same conditions**.

These experiments are called random experiments[1] (vs. deterministic experiments).

[1] For simplicity reasons, from now on, the word **experiment** is typically used to mean a random experiment

Experiments

Examples of random experiments:

1. flipping coins
2. rolling dice
3. observing the frequency of defective items from an assembly line
4. observing the frequency of deaths in a certain age group

Probability theory is a branch of mathematics that has been developed to deal with outcomes of random experiments, both real and conceptual.

Sample space and event

Definition

The set containing all possible outcomes of a random experiment is named its **sample space**, being usually represented by Ω.

Examples

1. Suppose one is flipping 2 coins, then $\Omega = \{(E,E),(E,N),(N,E),(N,N)\}$ where N="national side" e E="european/common side"
2. Suppose 3 six-sided dice are rolled and we are interested in the number of dots facing up, then $\Omega = \{(i,j,l) : i,j,l = 1,2,\ldots,6\}$

Definition

Any subset A of Ω is named an **event**. Each distinct outcome of the experiment is named **simple event**.

Certain event. Impossible event. Intersection of events

Definition

The **complement of an event A** consists of all outcomes of the experiment that do not result in event A. We write $\overline{A}$.

Definition

Let $\Omega = \{w_1, w_2, \ldots, w_n\}$ be a sample space.

Ω - represents a **certain event**

$\emptyset$ (empty set) - represents an **impossible event**

Definition

The **intersection of two events**, A and $B \subset \Omega$, is the event that both A and B occur when the experiment is performed. We write $A \bigcap B$.

Mutually exclusive events. Union of events

Definition

Two events are **mutually exclusive** (or disjoint or incompatible) if, when one event occurs, the other cannot, and vice versa (no elementary outcomes in common).

Definition

The **union of two events**, A and B, is the event that either A or B or both occur when the experiment is performed. We write $A \bigcup B$.

Interpretation of probability. Frequentist probability

One **intuitive way of computing the probability of an event** is by **using the relative frequency** (no knowledge, no assumptions).

- Suppose one can repeat an experiment an infinite number of times, always with the same conditions, and the result of the experiment is either the event A or not A (A is a simple event).
- Then, the probability of the event A is

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$$

i.e. limit of the relative frequency as the sample size goes to infinity.

Example: Toss a coin 100 times and define event A as "get National side". Suppose one gets 54 National and 46 European, then the relative frequency is 0.54.

Interpretation of probability. Frequentist probability

One **intuitive way of computing the probability of an event** is by **using the relative frequency** (no knowledge, no assumptions).

- Suppose one can repeat an experiment an infinite number of times, always with the same conditions, and the result of the experiment is either the event A or not A (A is a simple event).
- Then, the probability of the event A is

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

i.e. limit of the relative frequency as the sample size goes to infinity.

Example: Toss a coin 100 times and define event A as "get National side". Suppose one gets 54 National and 46 European, then the relative frequency is 0.54.

Interpretation of probability. Classical probability

The probability of an event A is a measure of our belief that the event A will occur. One rule to compute probability is to use

$$P(A) = \frac{n_A}{n} = \frac{\text{number of simple events in A}}{\text{total number of simple events}}$$

Assumptions: Ω is finite and equally likely outcomes.

Examples:

- Toss a die, 6 possible outcomes, numbers 1 to 6, so the probability that the upper face is 3 is 1/6.
- Toss a coin, 2 possible outcomes, National or European, so the probability of National is 1/2.

Interpretation of probability. Classical probability

The probability of an event A is a measure of our belief that the event A will occur. One rule to compute probability is to use

$$P(A) = \frac{n_A}{n} = \frac{\text{number of simple events in A}}{\text{total number of simple events}}$$

Assumptions: Ω is finite and equally likely outcomes.

Examples:

- Toss a die, 6 possible outcomes, numbers 1 to 6, so the probability that the upper face is 3 is 1/6.
- Toss a coin, 2 possible outcomes, National or European, so the probability of National is 1/2.

Probability axioms

Kolmogorov's axioms

1. $P(\Omega) = 1$
2. $P(A) \geq 0, \ \forall A \subset \Omega$
3. If $A_1, A_2, \ldots, A_n, \ldots$ are mutually exclusive events then

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i)$$

These axioms allow us to prove important properties, such as:

- $P(A) \leq 1$
- $P(A) = 1 - P(\overline{A})$
- $P(\emptyset) = 0$
- Se $A \subset B$ então $P(A) \leq P(B)$
- $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$
- $P(A \bigcap \overline{B}) = P(A) - P(A \bigcap B)$

Exercise 1
A sample space Ω consists of 4 simple events with probabilities $P(e_1) = 0.15$, $P(e_2) = 0.2$ and $P(e_3) = 0.4$.

a) Find the probabilities for all simple events
b) Let $A = \{e_1, e_3\}$ and $B = \{e_2, e_3\}$, then calculate
   i. P(A) and P(B)
   ii. P(A does not occur)
   iii. P(at least one occurs)
   iv. P(both A and B occur)
   v. P(neither A nor B occurs)
   vi. P(A occurs and B does not occur)

Exercise 2
An electronic equipment consists of 2 components A and B. It is known that the probability of failure of component A is 0.2, the probability that only B fails is 0.15, and the probability that both simultaneously fail is 0.15.

a) Determine the probability of failure of at least one of the two components.
b) Determine the probability of only A fails.

Conditional probability

Suppose you want to calculate the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred.

Definition

If the events are A and B respectively, this is said to be "the conditional probability of A given B". It is commonly denoted by $P(A|B)$, being defined as:

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)}, \quad P(B) > 0$$

Note:
In statistical inference, the conditional probability is an update of the probability of an event based on new information.

Some important results

1. $P(A \bigcap B) = P(A)P(B|A) = P(B)P(A|B)$
2. $P(A|B) + P(\overline{A}|B) = 1$
3. $P(\emptyset|B) = 0$
4. If $A_1 \subset A_2$ then $P(A_1|B) \leq P(A_2|B)$
5. $P(A|B) = \dfrac{P(A)P(B|A)}{P(B)}$

Exercise 3
Consider the following table, which presents the classification results of given itens according to next criteria - porosity and dimension.

	Porous	Non-porous
Defective	2.1%	4.9%
Non-defective	18.1%	74.9%

a) One item was chosen at random, showing to be defective. Which is the probability of being porous?
b) If a given item is porous, which is the probability of being defective?

Law of total probability

Sometimes we do not know the probability of a given event, but we are aware of the conditional probabilities given other events.

Total probability theorem

Let A be an event in the sample space Ω and $B_1, B_2, \ldots, B_n$, a given partition on that space, i.e. $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_i \bigcap B_j = \emptyset, \forall i \neq j$,

$$P(A) = \sum_{i=1}^{n} P(A \bigcap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Exercise 4
A manufacturer buys items from two different suppliers, A and B. According to past experience, the probability of itens from A being defective is 0.001, while this probability is 0.005 for supplier B. Consider that 35% of the material comes from supplier A and the remaining 65% from supplier B. If one item is chosen at random, which the probability of being defective ?

Bayes's theorem

Theorem

Let $B_1, B_2, \ldots, B_n$ be a partition of Ω then

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \quad \forall k = 1, \ldots, n$$

Exercise 5

A manufacturer buys items from two different suppliers, A and B. According to past experience, the probability of itens from A being defective is 0.001, while this probability is 0.005 for supplier B. Consider that 35% of the material comes from supplier A and the remaining 65% from supplier B. If a randomly chosen item is defective, what is the probability of having been supplied by A ?

Independent events

Definition

Two events **A and B are independent** if the fact that A occurs does not affect the probability of B occurring, which means

$$P(A \bigcap B) = P(A)P(B)$$

Notes:

- If A and B are independents then $P(A|B) = P(A)$ and $P(B|A) = P(B)$
- Any event is independent of $\emptyset$ and Ω

Exercise: If A and B are independents, then the independence also happens among A and $\overline{B}$; $\overline{A}$ and B; $\overline{A}$ and $\overline{B}$.

Solutions

Exercise 1:
a) $P(e_4) = 0.25$
b) *i*. P(A)=0.55 and P(B)=0.6   *ii*. $P(\overline{A}) = 0.45$ *iii*. $P(A \cup B) = 0.75$
iv. $P(A \cap B) = 0.4$ *v*. $P(\overline{A} \cap \overline{B}) = 0.25$ *vi*. $P(A \cap \overline{B}) = 0.15$

Exercise 2: *a*) 0.35   *b*) 0.05

Exercise 3: *a*) 0.3   *b*) 0.104

Exercise 4: 0.0036

Exercise 5: 0.0972

Random variable

- The outcome of an experiment need not be a number, for example, the outcome when a coin is tossed can be National or European.
- However, we often want to represent outcomes as numbers. A **random variable**, usually written X, is a function that associates a unique numerical value with every outcome of an experiment.
- The value of the random variable (r.v.) will vary from trial to trial as the experiment is repeated.

Examples:

1. A coin is tossed ten times. The r.v. X is the number of National sides that are noted. X can only take the values 0, 1, ..., 10.
2. A light bulb is burned until it burns out. The r.v. X is its lifetime in hours. X can take any positive real value.

Random variable

- The outcome of an experiment need not be a number, for example, the outcome when a coin is tossed can be National or European.
- However, we often want to represent outcomes as numbers. A **random variable**, usually written X, is a function that associates a unique numerical value with every outcome of an experiment.
- The value of the random variable (r.v.) will vary from trial to trial as the experiment is repeated.

Examples:

1. A coin is tossed ten times. The r.v. X is the number of National sides that are noted. X can only take the values 0, 1, . . . , 10.
2. A light bulb is burned until it burns out. The r.v. X is its lifetime in hours. X can take any positive real value.

Types of random variable

Discrete - It may take any of a specified finite or countable list of values such as 0, 1, 2, 3, 4, . . .

More examples:

- number of children in a family
- Friday night attendance at a cinema
- number of patients in a doctor's surgery
- number of defective light bulbs in a box of ten

Continuous - It may take any real value in IR or subset of IR.

More examples:

- height or weight of individuals
- amount of sugar in an orange
- time required to run a kilometre

Discrete random variable

Let D be the set of all possible values for the discrete random variable X.

Definition

The **probability mass function** of X is defined as:

$$f(a) = \begin{cases} P(X = a) & \text{if } a \in D \\ 0 & \text{otherwise} \end{cases}$$

Properties:

- $f(a) \geq 0,\ \forall a \in IR$
- $\sum_{a\in D} f(a) = 1$

Definition

The function F, with domain IR, defined as:

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

is named the **distribution function** of X, with $\lim_{a\to-\infty} F(a) = 0$ and $\lim_{a\to+\infty} F(a) = 1$.

Exercise 1:
In a given store of computer itens, the daily sale of hard drives of type X has the following probability function:

a	0	1	2
P(X=a)	0.2	0.65	0.15

Calculate P($X \leq 1$), P($X > 1.3$), P($X \leq 1.5$), P($0 \leq X < 2$) and P($0 \leq X \leq 1$).

Exercise 2:
Consider the random variable X with distribution function:

$$F(a) = P(X \leq a) = \begin{cases} 0 & \text{if } a < 0 \\ 1/8 & \text{if } 0 \leq a < 1 \\ 1/2 & \text{if } 1 \leq a < 2 \\ 7/8 & \text{if } 2 \leq a < 3 \\ 1 & \text{if } a \geq 3 \end{cases}$$

a) Define the corresponding probability function.
b) Calculate P($0 < X \leq 2$), P($0 \leq X \leq 2$), P($X < 4$) and P($X > 1$).

Continuous random variable

Definition

A random variable is continuous if and only if there is a real function $f(x)$, non negative, such that:

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

where $f(x)$ is called the **probability density function** and $F(a)$ is the corresponding **distribution function**.

Properties:

- $f(x) \geq 0, \forall x \in IR$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$
- $f(x) = \frac{d\ F(x)}{dx}$

Some remarks:

- $P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) =$
  $= P(a < X \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$
- $P(X > a) = P(X \geq a) = 1 - P(X \leq a) = 1 - F(a)$

Exercise 3:
Consider the random variable X with next density probability function:

$$f(x) = \begin{cases} 1+x & \text{se } -1 \leq x < 0 \\ 1-x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate:

- $\mathrm{P}(X \leq 0)$
- $\mathrm{P}(X > -0.5)$
- $\mathrm{P}(0 \leq X < 2)$
- $\mathrm{P}(0 \leq X \leq 1)$

Parameters of a distribution

Definition

The **expected value** or **population mean** of X is defined as:

$$\mu = E[X] = \begin{cases} \sum_i x_i f(x_i) & \text{discrete r.v.} \\ \int_{-\infty}^{+\infty} x\ f(x)dx & \text{continuous r.v.} \end{cases}$$

assuming that the sum or the integral is absolutely convergent.

Notes:

- The expected value gives a general impression of the behaviour of a r.v. without giving full details of its probability distribution.
- Two r.v.'s with the same expected value can have very different distributions.

Exercise 4:
For the random variables defined in exercises 1 and 3, calculate $E[X]$.

Parameters of a distribution

Properties of the expected value:

Let a and b be two real constants and X a random variable, then

- $E[a] = a$
- $E[a\ X + b] = aE[X] + b$

Let $X_1, X_2, \ldots$ be random variables, then

- $E[X_1 + X_2 + \ldots] = E[X_1] + E[X_2] + \ldots$

Let X and Y be independent random variables, then

- $E[XY] = E[X]E[Y]$

Mode: the mode of a distribution is the value mo where f(x) achieves its maximum.

Median: the median of a distribution is the value M where $P(X \leq M) \geq 0.5$ and $P(X \geq M) \geq 0.5$

Parameters of a distribution

Definition

The **variance** of the random variable X is defined as:

$$\sigma^2 = Var[X] = E[(X-\mu)^2] = \begin{cases} \sum_i (x_i-\mu)^2 f(x_i) & \text{discrete r.v.} \\ \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx & \text{continuous r.v.} \end{cases}$$

assuming that the sum or the integral is absolutely convergent.

Note: In practice, it is usual to calculate the variance as $Var[X] = E[X^2] - E[X]^2$

Parameters of a distribution

Exercise 5:
For the random variables defined in exercises 1 and 3, calculate $Var[X]$.

Properties of the variance:
Let a and b be two real constants and X a random variable, then

- $Var[X] \geq 0$
- $Var[a] = 0$
- $Var[a\,X + b] = a^2 Var[X]$

Let $X_1, X_2, \ldots$ be **independent** random variables, then

- $Var[X_1 + X_2 + \ldots] = Var[X_1] + Var[X_2] + \ldots$

Binomial distribution

Definition

Bernoulli trial (or binomial trial) is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted.

Definition

The Binomial distribution considers a **sequence of a fixed number of independent Bernoulli trials**, instead of only one bernoulli trial.

Examples of binomial experiments:

- Asking 200 people if they watch ABC news.
- Rolling a die to see if a 5 appears.

Examples which aren't binomial experiments:

- Rolling a die until a 6 appears (not a fixed number of trials).
- Asking 20 people how old they are (not two outcomes).
- Drawing 5 cards from a deck for a poker hand (done without replacement, so not independent).

Well-known discrete distributions | Binomial

Binomial distribution

Definition

Bernoulli trial (or binomial trial) is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted.

Definition

The Binomial distribution considers a **sequence of a fixed number of independent Bernoulli trials**, instead of only one bernoulli trial.

Examples of binomial experiments:

- Asking 200 people if they watch ABC news.
- Rolling a die to see if a 5 appears.

Examples which aren't binomial experiments:

- Rolling a die until a 6 appears (not a fixed number of trials).
- Asking 20 people how old they are (not two outcomes).
- Drawing 5 cards from a deck for a poker hand (done without replacement, so not independent).

28 / 44

Binomial distribution

Consider $X =$ "number of successes in a sequence of n independent yes/no experiments" (each of which yielding success with probability p). This variable takes random values from the finite set $\{0, 1, ..., n\}$ and has probability function

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n$$

Definition

Under previous conditions, one can state that X has a Binomial distribution or $X \sim Bin(n, p)$, with $0 < p < 1$, being

$$\mathrm{E}[X] = np \quad \text{e} \quad \mathrm{Var}[X] = np(1-p)$$

Notes:

- $Bin(1, p) \equiv Ber(p)$;
- $X = \sum_{i=1}^{n} X_i$, where each $X_i =$"outcome of the ith trial", has a Bernoulli distribution.

Binomial distribution in R environment

Let $X \sim Bin(n, p)$

R function	Returns
`dbinom(x,size=n,prob=p)`	$P(X = x)$
`pbinom(x,size=n,prob=p)`	$P(X \leq x)$
`pbinom(x,size=n,prob=p,lower.tail=F)`	$P(X > x)$
`qbinom(p,size=n,prob=p)`	Q such that $P(X \leq Q) =$p
`rbinom(n,size=n,prob=p)`	n random numbers

Examples:

- How to obtain $P(X = 2)$ when $X \sim Bin(10, 0.3)$?
  > dbinom(2,size=10,prob=0.3), or simply "dbinom(2,10,0.3)"
- How to generate a sample of 20 tosses of a balanced coin?
  > rbinom(20,1,0.5), where 1 means National and 0 European
- How to generate a sample of size 10 of the total number of National sides obtained in 100 tosses ?
  > rbinom(10,100,0.5)

Poisson distribution

The **Poisson model** has many applications, being typically used to model phenomena related to the number of events occurring in a fixed interval of time or space, such as distance, area or volume.

These events occur with a **known average rate** and independently of the time since the last event (or location of the nearest event) .

- Examples in **time:**
  1. number of hits to your web site in a day;
  2. client entries in a supermarket per week-day;
  3. number of telephone calls that arrive each day on a call center.
- Examples in **space:**
  1. number of typos on a printed page;
  2. number of Alaskan salmon caught in a squid driftnet.

Poisson distribution

In a Poisson process, consider:

- $\lambda > 0$, the average number of events per unit interval;
- $X =$"number of events per time/space unit interval". Then this r.v. takes values in set $\{0, 1, ...\}$ and the corresponding probability mass function is

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, ...$$

Definition

Under previous conditions, one can state that X has a Poisson distribution or $X \sim P(\lambda)$, with $\lambda \in IR^+$, being

$$\mathrm{E}[X] = \mathrm{Var}[X] = \lambda$$

Poisson distribution in R environment

Let $X \sim P(\lambda)$

R function	Returns
dpois(x,lambda=λ)	$P(X = x)$
ppois(x,lambda=λ)	$P(X \leq x)$
qpois(p,lambda=λ)	Q such that $P(X \leq Q) =$p
rpois(n,lambda=λ)	n random numbers

Examples:

- How to obtain $P(X = 2)$ when $X \sim P(3)$?
  > dpois(2,lambda=3), or simply "dpois(2,3)"
- Obtain the graphics for r.v. $X \sim P(\lambda)$, assuming $\lambda = 0.5$, $\lambda = 1$ and $\lambda = 4$. Comment the symmetry of these probability functions ?

```
> par(mfrow=c(1,3))
> x<-seq(0,6); prob<-dpois(x,0.5); plot(x,prob,main="lambda 0.5",t="h")
> x<-seq(0,10); prob<-dpois(x,1); plot(x,prob,main="lambda 1",t="h")
> x<-seq(0,10); prob<-dpois(x,4); plot(x,prob,main="lambda 4",t="h")
```


Poisson distribution

Exercise 6:
During lunch break (from 12:00 to 14:00), the average number of cars parking in the main Park of a given town is 360. What is the probability, during one minute, of the arrival of $x = 0, 1, 2, \ldots$ cars ?

Exercise 7:
The number of telephone calls that arrive, in average, each hour on a call center of a given enterprise is 45. What is the expected value and the standard deviation of the number of calls arriving per minute ?

Exponential distribution

The **exponential distribution** is a model for some lifetime or time intervals between two consecutive random events. Examples:

1. Lifetime of a certain electronic component (in hours);
2. Time between two consecutive failures of a machine.

The exponential distribution is strongly related to the Poisson distribution:

- For example, suppose the number of failures per month of a given machine follows the Poisson distribution $P(30)$. As $\lambda = 30$, it means the average number of failures per month is 30. Then, the average time between two random failures is $\frac{1}{\lambda} = \frac{1}{30} = 0.033$ month, i.e. about 1 day.

Exponential distribution

Let $X =$"time between two consecutive events", then the probability density function and the distribution function of this r.v. are, respectively,

$$f(x) = \lambda e^{-\lambda x} \quad \text{and} \quad F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

Definition

Under previous conditions, one can state that X has a Exponential distribution or $X \sim Exp(\lambda)$, with $\lambda \in IR^+$, being

$$\mathrm{E}[X] = \frac{1}{\lambda} \quad \text{and} \quad \mathrm{Var}[X] = \frac{1}{\lambda^2}$$

Exercise 8:
According to Exercise 6, the average number of cars parking during lunch break is 3 per minute. What is the probability of time between the arrival of two cars being larger than 1 minute ? And smaller than 10 seconds ?

Hint: In R, use > ?pexp

Normal distribution

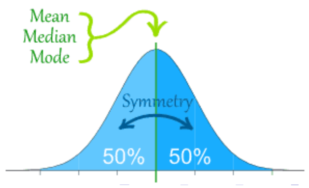
The **Normal** (or Gaussian) distribution is quite popular in applications, being adopted when data tends to be around a central value with no bias left or right.

For example, it may be used to model:

- heights of people
- size of things produced by machines
- errors in measurements
- blood pressure
- marks on a test

Characteristics of the normal distribution:

- mean = median = mode
- symmetry about the center
- 50% of values less than the mean and 50% greater than the mean



Normal distribution

A r.v. X follows the normal distribution iff the corresponding probability density function is defined as

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \ x \in IR$$

Definition

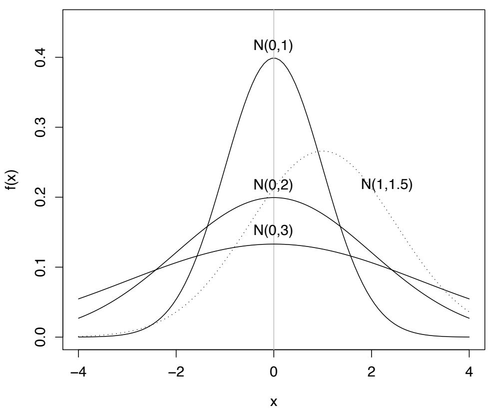
Under previous conditions, one can state that X has a normal distribution, or $X \sim N(\mu, \sigma^2)$, with $\mu \in IR, \sigma \in IR^+$, being

$$\mathrm{E}[X] = \mu \quad \text{and} \quad \mathrm{Var}[X] = \sigma^2$$

Notes:

- The simplest case of a normal distribution is known as the standard normal distribution. This is a special case where $\mu = 0$ and $\sigma = 1$.
- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Some graphical representations



In R**:**

```
> ?dnorm
> x<-seq(-4,4,0.1); pr<-dnorm(x,mean=0,sd=1); plot(x,pr,t="l",ylab="f(x)")
> pr<-dnorm(x,mean=0,sd=2); lines(x,pr); text(list(x=-0.1,y=0.21),"N(0,2)")
> pr<-dnorm(x,mean=0,sd=3); lines(x,pr); text(list(x=-0.1,y=0.14),"N(0,3)")
```


Properties of normal distribution

If $X \sim N(\mu, \sigma^2)$ with distribution function $F(x)$, then

- $\forall a, b \in IR$ $\quad P(a < X < b) = P(a \leq X \leq b) = F(b) - F(a)$
- $F(x)$ is symmetric with respect to μ, thus
$$F(x) = 1 - F(-x)$$
- If $Y = aX + b$, where a and b are constants, then
$$Y \sim N(a\mu + b, \ a^2\sigma^2)$$

Moreover, if $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are **independents**, then

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Exercise 9:
The size of beer bottles sold by a given supplier can be considered normally distributed with $\mu = 0.33$ and $\sigma^2 = 10^{-5}$. The department of quality control decided to only accept for distribution bottles with a capacity between 0.32 and 0.34 l.

- What is the probability of a bottle being rejected?
- What is the probability of finding a bottle with less than 0.32 l ? And less than 0.33 l (**no calculations needed**) ?

Some Solutions

Exercise 6: f(0)=0.050, f(1)=0.149, f(2)=0.224, ...

Exercise 7: 45/60=0.75 and $\sqrt{45/60}$=0.866

Exercise 8: 0.05 and 0.39

Exercise 9: 0.0016, 0.0008 and 0.5

Central Limit Theorem - CLT

The central limit theorem states that the distribution of **the sum (or average) of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution**.

This distribution is as closer to the normal, the greater the number of r.v.'s in the sum.

Theorem

Let $X_1, X_2, ..., X_n$ be a set of n independents r.v.'s and each X_i have an arbitrary probability distribution $P(x_1, ..., x_n)$ with mean μ and a finite variance σ^2. When n $\rightarrow \infty$

$$\frac{(\sum_{i=1}^{n} X_i) - n\mu}{\sigma\sqrt{n}} \sim N(0,1) \quad \Leftrightarrow \quad \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Approximations for discrete distributions

The CLT may be applied in many different situations, namely:

1. According to the properties of the Binomial distribution, we can interpret a r.v. $X \sim Bin(n,p)$ as the sum of n Bernoulli's r.v. $X_i \sim Ber(p)$.
   Thus, for large n, distribution $Bin(n,p)$ can be approximated by $N(np, np(1-p))$.
   In practice, this approximation may be considered for $n > 30$, $np > 5$ and $n(1-p) > 5$.

2. According to the properties of the Poisson distribution, we can interpret a r.v. $x \sim P(\lambda)$ (with λ integer) as a sum of λ r.v.'s of Poisson with parameter 1.
   Thus, for large λ, $P(\lambda)$ can be approximated by $N(\lambda, \lambda)$.
   In practice, this approximation may be considered for $\lambda > 50$.

Central Limit Theorem - CLT

Simulation example for CLT:

1. Generate 100 random values $X_i \sim Exp(10)$, i.e. $E[X_i] = \frac{1}{10}$ and $Var[X_i] = \frac{1}{10^2}$. Let $S = \sum_{i=1}^{100} X_i$, i.e. $E[S] = 10$ and $Var[S] = 1$.
   According to CLT, r.v. $S \sim N(10,1)$.
2. Repeat previous step 200 times, obtaining sample $\{s_1, \ldots, s_{200}\}$
3. Plot the histogram for $\{s_1, \ldots, s_{200}\}$ and comment.

In R:

```
> allSums <- rep(0,200)
> for (k in 1:200) allSums[k] <- sum(rexp(100,rate=10))
> hist(allSums,freq=F); x <- seq(7,12,0.1); lines(x,dnorm(x,10,1))
```