Basic Statistical Analysis

Rodrigo Labouriau

Applied Statistics Laboratory
Department of Mathematics
Aarhus University

Module 1, Day 1 - Basic Probability - 2024

1

General Remark

This material is only for internal use in the course.

Please, do not circulate and **do not record**.

Outline

First Examples

The basic probabilistic model

Random variables, expectation and variance

Some common key distributions

Summary of the day

Course overview

Typical construction and use of a statistical model:

- Definition of a mathematical structure representing the phenomena of interest (statistical model)
- Use available data to infer the model or adjust details (*e.g.* results of an experiment, observational study or registers)
  (basic inference)
- Verify whether the model really adjusts
  (model control)
- If this process is successful then we might use this mathematical structure
  to characterise the phenomena of interest,
  to answer specific questions and
  to make predictions (in some cases)
  (concluding inference)

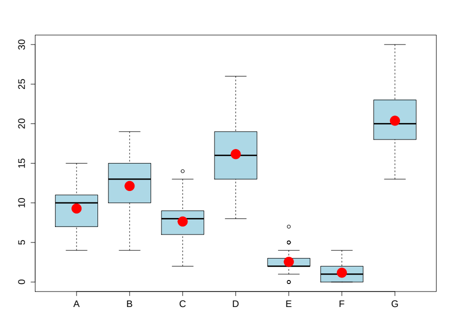
Some initial examples illustrating the schema of modelling

First example: Counts response one-way classification

- Experiment comparing 7 different treatments for parasites in chickens (termed "A", "B", . . . , "G")
- The chickens were slaughtered and the number of worms in their intestines were counted.
- Question:
  Do we have evidence that the treatments affect the number of worms infesting the intestines of the chickens.
- Data: 45 repetitions for each treatment, 315 observations (chickens)

Initial examples illustrating the schema of modelling

First example: Counts response one-way classification



Initial examples illustrating the schema of modelling

First example: Counts response one-way classification

- I claim that the data fits well to a Poisson distribution
  (a classical distribution for counts)
  (statistical model)
- Using the data it is possible to adjust a model that assumes one different mean number of worms per treatment (basic inference)
- A range of model control techniques can be applied:
  test of homogeneity, residual analyses, adherence to the Poisson distribution etc (model control)
- A test to verify whether the mean number of worms is different for the different treatments can be defined. We can estimate the mean number of worms, test equality of pairs of treatments etc (concluding inference)

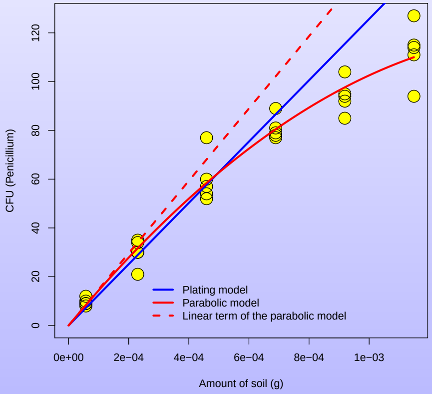
Initial examples illustrating the schema of modelling

Second example: Counts of colonies of *P. verrucosum*, regression

- Experiment:
  Determined the number of colonies of *P. verrucosum*in suspensions of soil with different concentrations
- Question:
  How many colony forming units of *P. verrucosum* are contained in a gram of soil?

Initial examples illustrating the schema of modelling

Second example: Counts of colonies of *P. verrucosum*, regression



Initial examples illustrating the schema of modelling

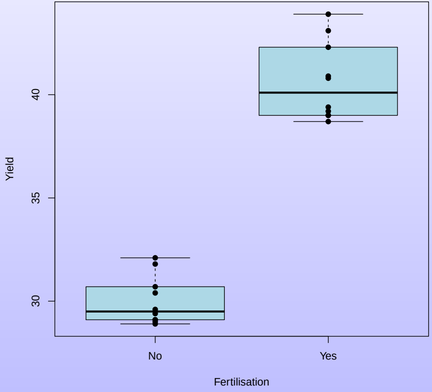
Second example: Counts of colonies of *P. verrucosum*, regressionSecond example: Counts of colonies of *P. verrucosum*, regression

- I claim that the data fits well to a Poisson distribution
  (this can be deduced!)
  The mean number of colonies increases quadratically with the amount of amended soil. (statistical model)
- Using the data it is possible to adjust a model this model (basic inference)
- A range of model control techniques can be applied:
  test of homogeneity, test of adequacy of the adjusted curve, residual analyses, adherence to the Poisson distribution etc (model control)
- We can conclude that there is a strong effect of competition among the in the same Petri dish, but we can anyway estimate the number of colonies present in a gram of soil (with and without a correction for the effect of competition) (concluding inference)

Initial examples Corn yield (one block)

- Study on the effect of nitrogen fertilisation on the yield of corn plants
- Observed plot yields of the plots **without** fertilisation:
  29.1 30.7 29.6 29.4 30.4 31.8 28.9 29.0 32.1 29.1
- Observed plot yields of the plots **with** fertilisation:
  43.1 38.7 40.8 43.9 39.0 39.4 39.2 39.0 40.9 42.3

Initial examples Corn yield (one block)



Initial examples illustrating the schema of modelling

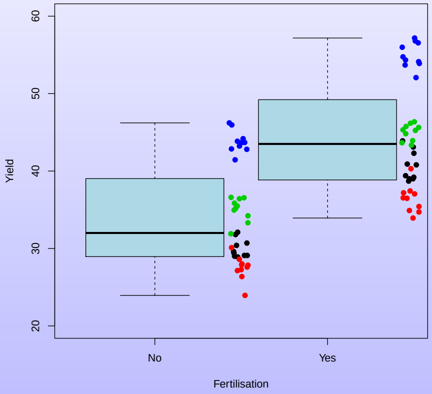
Third example: Yield of corn, simple one-way variance analysis model

- I claim that the data fits well to a Normal distribution
  (Central limit theorem: sums of small errors imply approximately normal distribution (under some regularity conditions!)
  The mean yields is the same for plots with the same fertilisation.
  Possibly different means when different fertilisation is used.
  The variance is constant (I will clarify this point latter) (statistical model)
- Using the data it is possible to adjust this model (basic inference)
- A range of model control techniques can be applied:
  test of homogeneity of variances, residual analyses, adherence to the normal distribution etc (model control)
- We can conclude that there is a strong effect of fertilisation, calculate confidence intervals, etc ... (concluding inference)

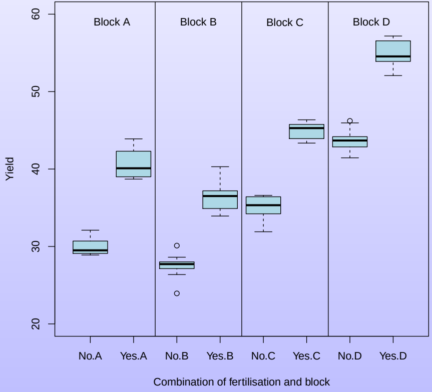
Details on the example of corn - Extended

- Study on the effect of nitrogen fertilisation on the yield of corn plants
- Yields of 40 plots fertilised with N and 40 plots without N fertilisation
- Divided the area into 4 blocks (sub-areas) each with 10 fertilised plots and 10 not fertilised plots

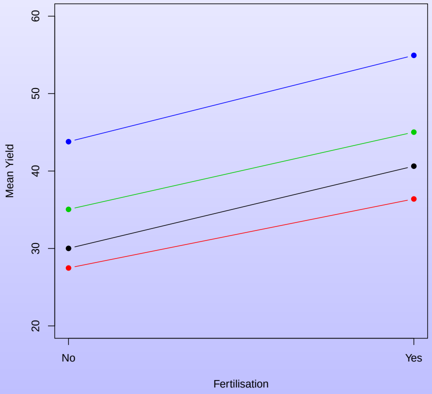
Details on the example of corn with blocks



Details on the example of corn with blocks



Details on the example of corn with blocks



Initial examples illustrating the schema of modelling

Third example: Yield of corn, simple two-ways variance analysis model

- I claim that the data fits well to a Normal distribution
  (Central limit theorem: sums of small errors imply approximately normal distribution (under some regularity conditions!)
  The mean yields is the same for all the plots receiving the same fertilisation. From the same block
  The means can be written as a sum of a quantity depending on the fertilisation and a quantity depending on the block
  The variance is constant (I will clarify this point latter) (statistical model)
- Using the data it is possible to adjust this model (basic inference)
- A range of model control techniques can be applied:
  test of homogeneity of variances, residual analyses, adherence to the normal distribution etc (model control)
- We can conclude that there is a strong effect of fertilisation, calculate confidence intervals, etc ... (concluding inference)

Basic probabilistic models

First very simple example: Binary trials

- *Binary trial*: an experiment with only two outcomes
  One of the most simple situations!
- Example: two boxes
  One containing an object an one empty
  It is not know which box contains the object
  Experiment: Choose one box and observe if we get the object
  Possible results: object or not
- The result of the experiment is a *random quantity*
  because we cannot predict it with certainty

Basic probabilistic models

First very simple example: Binary trials

- Other classical example:
  Toss a coin and observe the result
  Two possibilities: tail or head
- Is the result of this experiment a random quantity?

Basic probabilistic models

First very simple example: Binary trials

- In both examples we cannot predict the result with certainty
  But we still can observe some regularity
- Toss the coin many times
  Observe the relative frequency of the tails
  The relative frequency tends to stabilises around a value
- These regularity suggests that there is an intrinsic characteristic common to all the realisations of the experiment
- Statistics and probability theory takes advantage of these type of regularity

Basic Probability Theory

Basic ideas

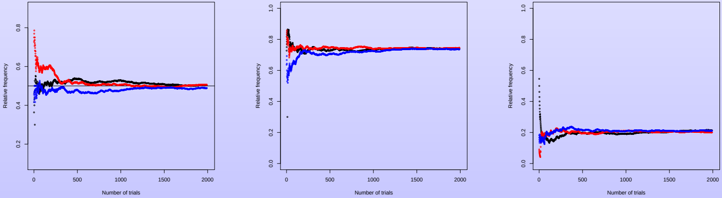
- Random event: an event such that its occurrence cannot be predicted with certainty
- Example: Flip a coin
- Probability: Limit of the relative frequency as the number of observations increases
- A event

  $P(A) = \text{Probability of } A$

  $$\text{Relative frequency of } A = \frac{\text{Number of occurrences of } A}{\text{Number of occurrences of trials}} \to P(A)$$

Basic Probability Theory

Example: Flip a coin and observe whether it turns a tail (simulations with three different probabilities)



Basic probabilistic models

First very simple example: Binary trials

- Convention for the binary trial:
  One of the outcomes is called "success"
  and the other "failure"
- We call this experiment the *basic binary trial*
- Probability of success $= p$
  Then the probability of failure $= 1 - p$
  Question: Why?

Basic probabilistic models

Second simple example: Binomial trials

- A more complex situation:
  perform twice the basic binary trial
  and count the number of successes
- Possible results: 0, 1 and 2
- Let Y be a variable representing the result of this experiment
- Y is called a *random variable*
  since its value is not known with certainty
- Notation (events):
  $[Y = 0]$ = not observing a success
  $[Y = 1]$ = observing one success
  $[Y = 2]$ = observing two successes
- Notation (probabilities):
  $P(Y = 0)$ = probability of not observing a success
  $P(Y = 1)$ = probability of observing one success

  $P(Y = 2)$ = probability of observing two successes

Basic probabilistic models

Second simple example: Binomial trials

- We calculate $P(Y = 0)$, $P(Y = 1)$ and $P(Y = 2)$
- Recall, the experiment is: repeat twice the basic binary trial
  The probability of success in the basic binary trial is p

- Calculation of $P(Y = 2)$
  $[Y = 2]$ = observing 2 successive successes

- Need an assumption:
  The results of the first binary basic trial does not influence at all the result of the second basic binary trial

  (think in the example of tossing a coin)

Basic probabilistic models

Second simple example: Binomial trials

- The relative frequency of the event "observing success in both binary basic trials" is given by the product of the relative frequencies of observing a success in the first and in the second basic binary trial
- Therefore the probability of observing two consecutive successes is the probability of success in the first trial times the probability of success in the second trial
- We say that the events "success in the first trial" and "success in the second trial" are **independent**
-

$$\begin{aligned} P(Y=2) &= P(\text{"success in the first trial"}) \cdot P(\text{"success in the second trial"}) \\ &= p^2. \end{aligned}$$

- An analogous argument yields: $P(Y=0) = (1-p)^2$

Basic probabilistic models

Second simple example: Binomial trials

- We calculate $P(Y=1)$
- Two possibilities for $[Y=1]$ occur:
  $A=$ [Observe success and then failure] **or**
  $B=$ [Observe failure and then success]

- Lets calculate the probability of the event A

  (Observe success and then failure)

  $P(A)=P(success)\cdot P(failure)=p(1-p)$

Basic probabilistic models

Second simple example: Binomial trials

- Analogously
  $P(B) = P(\textit{failure}) \cdot P(\textit{success}) = (1-p)p = p(1-p)$

- A and B are mutually exclusive

  (if A occurs then Bdoes not occur and vice-versa)
- If the events A and B are mutually exclusive, then the probability of A and B is the sum of the probability of A and the probability of B
  if $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$
- $P(Y = 1) = P(A \cup B) = P(A) + P(B) = 2p(1-p)$

Basic Probability Theory

Mutually exclusive events

- A and B events

  A and B are mutually exclusive when:

  A occurs implies that B does not occur and

  B occurs implies that A does not occur

- Example: Toss a coin, $A =$ "observe a head" and $B =$ "observe a tail"

- If A and B are mutually exclusive then

$$P(A \text{ or } B) = P(A) + P(B)$$

Basic Probability Theory

Complementary events

- If A is an event,
  the complement of A is the event "A does not occur"

  Notation: A^c

- Example:
  Toss a coin, $A =$ "observe a head" and $A^c =$ "not observe a head"

- Clearly,

$$P(A^c) = 1 - P(A)$$

Basic Probability Theory

Independency of events

- The events A and B are *independent* when

$$P(A \text{ and } B) = P(A).P(B)$$

- Example:
  Toss a coin twice:
  $A =$ "observe a head in the first trial" and
  $B =$ "observe a head in the second trial"

Basic probabilistic models

Another example: Uniform distribution

- Another example of experiment with random result:
  One person places an object in a path between two locations
  No additional information is available
  The result of the experiment is the position of the object
- The result is random because it cannot be predicted with certainty
- For simplicity assume the path with length 1
  Possible results: any number between 0 and 1
- Notation: $X \sim U[0, 1]$
  We say X is uniformly distributed in the interval $[0, 1]$
- Typical question:
  What is the probability of the result lies in a certain region?
  Specific question:
  What is the probability of a result between 0 and $1/2$?

Basic probabilistic models

Another example: Uniform distribution

- $X \sim U[0,1]$
- Specific question: What is the probability of a result between 0 and 1/2?
- Intuitive approach:
  If all the regions of the interval $[0,1]$ are equiprobable, then repeating the experiment many times, the chosen point will lie between 0 and 1/2 in approximately half of the cases
- That is, the relative frequency of the event
  "the chosen points lies between 0 and 1/2" approaches 1/2
- As before, the probability of
  "chose a point between 0 and 1/2" is 1/2
- Notation: $P(X \text{ in } [0,\ 1/2]) = 1/2$

Basic probabilistic models

Another example: Uniform distribution

- $P(X \text{ in } [0,\ 1/2]) = 1/2$
- Analogously,
  $P(X \text{ in } [0,\ 1/4]) = 1/4$
  $P(X \text{ in } [1/2,\ 1]) = 1/2$
- More general:
  given an interval $[a, b]$ in $[0, 1]$
  $(0 \leq a < b \leq 1)$
  $P(X \text{ in } [a,\ b]) = b - a$
  $b - a$ is the length of $[a, b]$

Basic probabilistic models

The general idea of probability

- There are some common features between the two examples studied
- In both we found a set of possible values, the *sample space*
  ("success" or "failure" in the case of the binomial trial and a number between 0 and 1 in the case of the uniform distribution)
  Denote the sample space by Ω
- There is a class of events that we can attribute a probability
- The probability was defined as a limit of the relative frequencies when many repetitions of the experiment where performed

Basic probabilistic models

The general idea of probability

- The probability was defined as a limit of the relative frequencies when many repetitions of the experiment where performed
- Three immediate consequences:
  1. Given an event A, its probability is a positive number, *i.e.* $P(A) \geq 0$;
  2. The probability of the sample space is one, *i.e.* $P(\Omega) = 1$;
  3. Given a sequence of mutually exclusive (disjoint) events, say $A_1, A_2, \ldots$, the probability that one of the events occurs is the sum of the probability of each of the events, i.e. $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$.
- The three points above are called the Kolmogorovs axioms

Basic probabilistic models

The general idea of probability

- Two events, say A and B, are *independent* if the probability that A and B occurs is the product of the probability of A and the probability of B, i.e. $P(A \cap B) = P(A) \cdot P(B)$
- Kolmogorovs axioms for probability:
  1. Given an event A, its probability is a positive number, i.e. $P(A) \geq 0$;
  2. The probability of the sample space is one, i.e. $P(\Omega) = 1$;
  3. Given a sequence of mutually exclusive (disjoint) events, say $A_1, A_2, \ldots$, the probability that one of the events occurs is the sum of the probability of each of the events, i.e. $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$.
- The three points above (*i.e.* the Kolmogorovs axioms) together with the notion of independency can be used to build the theory of probability!

Basic probabilistic models

Tutorials and exercises (To be done after the lecture)

- Kolmogorovs axioms for defining probability:
  1. Given an event A, its probability is a positive number, i.e. $P(A) \geq 0$;
  2. The probability of the sample space is one, i.e. $P(\Omega) = 1$;
  3. Given a sequence of mutually exclusive (disjoint) events, say $A_1, A_2, \ldots$, the probability that one of the events occurs is the sum of the probability of each of the events, i.e. $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$.
- Exercise 1: Justify the three Kolmogorovs axioms in terms of limits of relative frequency.
- Exercise 2: Show, using the axioms of probability that the probability of the complementary of an event is given by one minus the probability of the event, i.e. $P(A^c) = 1 - P(A)$
- Please, study and discuss the tutorial 02

Random variables

The idea of distribution of a r.v.

- The results of a probabilistic experiment as described before are usually registered by mean of a variable taking values at random, called a *random variable*
- What is the probability of a random variable assumes a certain value or take values in a certain region?
- We can attribute a *probability law* or *distribution* to random variables using the close correspondence of the results of the basic probabilistic experiment and the values of the random variable

Random variables

The idea of distribution of a r.v.

- The way to describe the distribution of a random variable differs slightly according to the type of the variable
- A random variable taking a values in an countable set is called a *discrete random variable*
  (*e.g.* Y, or the results of counting something)
- A random variable that takes real number values
  (*e.g.* $X = 0.333$, $Y = 7/8$ etc)
  is called a *continuous random variable*

Random variables

The distribution of discrete r.v.

- A random variable taking a values in an countable set is called a *discrete random variable*
- In this case the law of probability of the random variable is described by a function that associates each possible value of the random variable with its probability
- This function is called the *probability function* of the random variable

Random variables

The distribution of discrete r.v.

For example, in the binomial experiment the probability function is given by, for each possible value t,

$$f_Y(t) = P(Y = t) \begin{cases} p^2 & , \quad \text{if } t = 2, \\ 2p(1-p) & , \quad \text{if } t = 1, \\ (1-p)^2 & , \quad \text{if } t = 0\,. \end{cases}$$

Random variables

The distribution of continuous r.v.

- A random variable X is called *absolutely continuous* or simply *continuous* if there is a function f taking positive values (i.e. $f(x) \geq 0$, for each x) such that, for each real number x,

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

- For those not acquainted with the integration, the right hand sign in the expression above is the area between the graph of the function f and the horizontal axis, measured up to the point x
- The function f is called the *probability density of* X, or in short the *density of* X.

Random variables

The distribution of continuous r.v.

- In the example of the uniform distribution (i.e. the example on continuous variable) the probability density of the random variable Y is the function

$$f(y) = \begin{cases} 1 & , \quad \text{if } 0 \leq y \leq 1\,; \\ 0 & , \quad \text{otherwise.} \end{cases}$$

- The calculation of the probability of Y takes a value between 0 and $1/2$ can be alternatively calculated in the following way:

$$P([0 \leq Y \leq 1/2]) = P([Y \leq 1/2]) \\ = \int_{-\infty}^{1/2} f(y)dy = \int_{-\infty}^{1/2} dy = 1/2\,.$$

Random variables

The distribution of continuous r.v.

- Changing the form of the density, the law of probability changes also

- There are two restrictions:
  1) the density function should assume only non-negative values and
  2) the density should integrate 1

  (that is, the area between the graph of the density and the horizontal axis should be 1)

Random variables

The notion of independency

- It is possible to define the notion of independency of two random variables in a somehow similar way as we did for the independency of events
- This is important because several forms of independency of the observations will be a customary assumption in the models that we will study
- Intuitive idea:
  If the knowledge of the value taken by a random variable does not affect the distribution of another random variable, then they are independent.
- Formally we say that the random variables X and Y are independent when, for each par of numbers x and y,

$$P(X \leq x \text{ and } Y \leq y) = P(X \leq x) \cdot P(Y \leq y) \ .$$

That is, the event "X is less or equal x" is independent of the event "Y is less or equal y", for each x and each y.

The notion of expectation First simple example

- Consider an experiment that produces the results: 10 and 20 with probability 1/2 for each of these results
- X is a random variable representing the results of this experiment
  $P(X = 10) = 1/2$ and $P(X = 20) = 1/2$
- If we repeat this experiment many times, say n times
  in approximately 1/2 of the repetitions we will get the value 10
  and in approximately 1/2 of the repetitions we will get the value 20
- We may say that the expected value of the experiment is approximately

$$1/2 * 10 + 1/2 * 20 = 15$$

- The *expectation* or *expected value* of X,
  denoted by $\mathrm{E}(X)$, is $1/2 * 10 + 1/2 * 20 = 15$
- Changing the values of the probabilities, changes expectation of X

$$\mathrm{E}(X) = 10.P(X = 1) + 20.P(X = 2)\ .$$

The notion of expectation

- The idea of expectation can be easily understood for discrete variables taking a finite number of values
- X is a random variable taking n values, $x_1, x_2, \ldots, x_n$, with probabilities $p_1, p_2, \ldots, p_n$, respectively.
- The *expectation* or *expected value* of X is the sum of the possible values of X multiplied by their probabilities
- We use the symbol $\mathrm{E}(X)$ to denote the expectation of X and write

$$\mathrm{E}(X) = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n \ .$$

The notion of expectation

Simple examples

- Example: binary trial
  X takes the values 0 and 1 with probabilities $(1-p)$ and p, respectively.
  The expectation of X is then,

$$E(X) = (1-p)\,0 + p\,1 = p\ .$$

- Example: the binomial trial
  X takes the values $0, 1$, and 2
  with probabilities $(1-p)^2$, $2p(1-p)$ and p^2.
  The expectation is

$$E(X) = (1-p)^2 0 + 2p(1-p)1 + p^2 2 = 2p\ .$$

- Remark: The random variable X represents the number of successes.

Basic Probability Theory

Expectation of continuous random variables

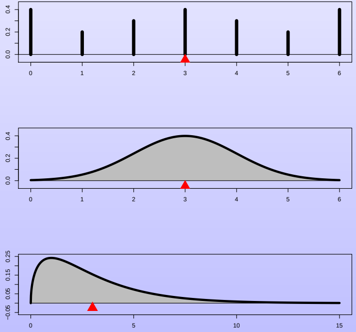
- If X continuous with density f then

$$E(X) = \int xf(x)dx$$

- The *expected value* of a continuous random variable X is the centre of mass of the graph of the density function

Basic Probability Theory

Physical interpretation of the expectation: Centre of gravity



Basic properties of the expectation

The expectation has the following basic properties:

1. If the random variable X is equal to a constant c with probability 1, then $\mathrm{E}(X) = c$;
2. If X and Y are random variables (with expectation well defined) and a, b are constants, then $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$;
3. If X and Y are random variables (with expectation well defined) such that $X \leq Y$ with probability 1, then $\mathrm{E}(X) \leq \mathrm{E}(Y)$.
4. (Jensens inequality) If ϕ is a convex real function and X is a random variable with finite expectation, then

$$\mathrm{E}\left\{\phi(X)\right\} \geq \phi\left\{\mathrm{E}(X)\right\} .$$

The notion of variance

The variance of a random variable X is defined by

$$\mathrm{Var}(X) = \mathrm{E}\{X - \mathrm{E}(X)\}^2 = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 \ .$$

Clearly, $\{X - \mathrm{E}(X)\}^2$ is a measure of the distance between the random variable X and its expectation.
Therefore, the expected value of this distance, i.e. the variance, is a measure of the dispersion of the data around its expected value.
The larger is the variance the more disperse is the data.

The variance of the binary variable X taking values 0 and 1 with probabilities $(1.p)$ and p is

$$\mathrm{Var}(X) = \mathrm{E}\{X - \mathrm{E}(X)\}^2 = \mathrm{E}\{X - p\}^2 = \cdots = \mathrm{E}(X^2) - p^2 \ .$$

To complete the calculation above we must compute the expectation of the random variable X^2. Note that $X^2 = X$, since X takes only the values 0 and 1. Therefore $\mathrm{E}(X^2) = \mathrm{E}(X)$. Replacing that in the last equation yields

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - p^2 = p - p^2 = p(1-p) \ .$$

The notion of variance

The variance has the following basic properties:

1. If the random variable X is equal to a constant with probability 1, then $\mathrm{Var}(X) = 0$;
2. If the random variable X has finite variance and b is a constant, then $\mathrm{Var}(bX) = b^2 \mathrm{Var}(X)$;
3. If the random variables X and Y are independent, then $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Three key distributions

- We studied three key distributions that will be the basic building blocks of (most of) the statistical models we will study
- Binomial distribution: study the occurrence of events, frequencies etc
- Poisson distribution: counting data
- Normal distribution: continuous measurements
- There are **many** other important distributions ...

Three key distributions:

The binomial Distribution

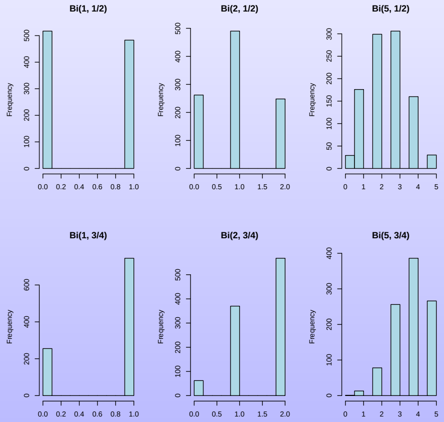
- Binomial distribution: Perform independently n times a basic binary trial with probability p of success and count the number of successes.
- Notation: $X \sim Bi(n, p)$

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

  for $x = 0, 1, \ldots, n$.
- $E(X) = np$, $Var(X) = np(1-p)$
  The variance can be expressed as a function of the mean.

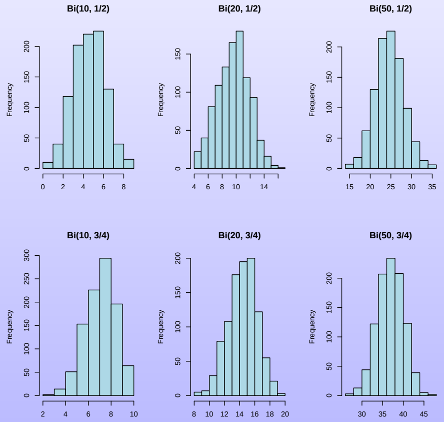
Three key distributions:

1000 simulations of the binomial distribution



Three key distributions:

1000 simulations of the binomial distribution



Three key distributions:

The Poisson distribution

- The Poisson distribution: describes the number of events
  (number of accidents, number of mutations in a fragment of DNA, number of worms in a portion of soil, etc.)
- This distribution was first used by Siméon-Denis Poisson
  Poisson, S.D., 1838. *Recherches sur la probabilité des jugements en matières criminelles et matière civile*
  (Study on the Probability of Judgments in Criminal and Civil Matters)
  to study the number of occurrences of an event during a time-interval of a given length, specifically the number of criminal and civil judgments
- The Poisson distribution takes positive integer values (*i.e.* 0,1,2, ...) and depends on a single parameter, called the *intensity parameter* and usually denoted by λ

Three key distributions:

The Poisson distribution

- The Poisson distribution: describes the number of events
(number of accidents, number of mutations in a fragment of DNA, number of worms in a portion of soil, etc.)
- This distribution was first used by Siméon-Denis Poisson
Poisson, S.D., 1838. *Recherches sur la probabilité des jugements en matières criminelles et matière civile*
(Study on the Probability of Judgments in Criminal and Civil Matters)
to study the number of occurrences of an event during a time-interval of a given length, specifically the number of criminal and civil judgments
- The Poisson distribution takes positive integer values (*i.e.* 0,1,2, . . .) and depends on a single parameter, called the *intensity parameter* and usually denoted by λ

Three key distributions:

The Poisson distribution

- A random variable Y is said to follow a *Poisson distribution* with parameter λ ($\lambda > 0$) if

$$P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!},$$

for $y = 0, 1, 2, \ldots$.
Here $y! = y \cdot (y-1) \cdot \cdots \cdot 1$ and $0! = 1$.

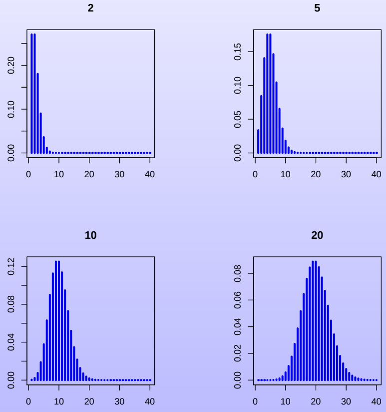
- A Poisson variable takes only non-negative integer values.
The Poisson distribution describes typically counts
(but there exist many other distributions for counts!!!).

- Notation: $Y \sim Po(\lambda)$

- $E(Y) = Var(Y) = \lambda$

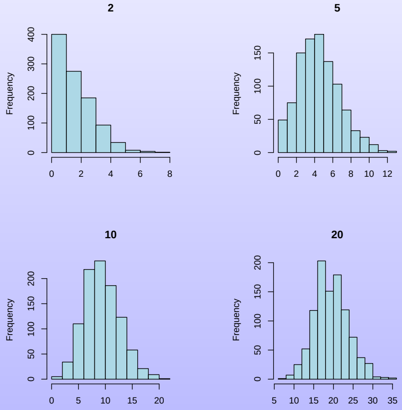
Three key distributions:

The probability function of the Poisson distribution



Three key distributions:

Simulated 1000 Poisson random variables



Three key distributions:

A classical example of Poisson distribution - Counts of alpha-particles

- Frequency of counts of alpha-particles emitted by the radioactive decay of a source of polonium, registered in time-intervals of 72 seconds

Counts:	0	1	2	3	4	5	6	7
Frequency:	57	203	383	525	532	408	273	139
Counts:	8	9	10	11	12	13	14	+ 15
Frequency:	45	27	10	4	0	1	1	0

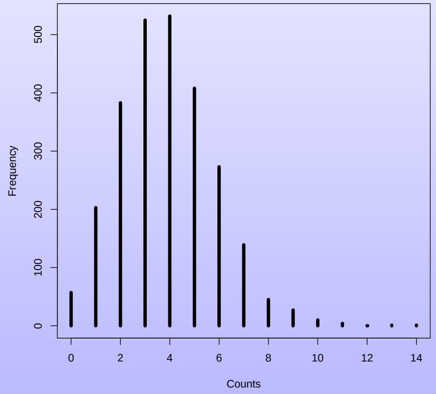
  Rutherford, E. and Geiger, M. (1910).

- Mean of counts: 3.87
  Variance of counts: 3.74
- A reasonable estimate of λ is 3.87

  (is the maximum likelihood estimate that we will study latter in this lecture)

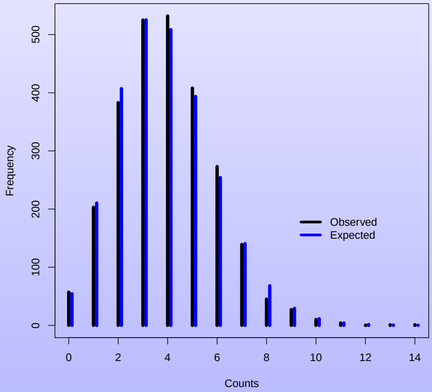
Three key distributions:

A classical example of Poisson distributed data - Counts of alpha-particles



Three key distributions:

A classical example of Poisson distributed data - Counts of alpha-particles



Three key distributions:

The normal distribution

- The normal distribution is one of the most used (and misused) distributions
- The normal distribution was used by Gauss to describe errors in astronomical measurements
  and is sometimes called the gaussian distribution
- The normal distribution was in fact used before Gauss by De Moivre and Laplace

Three key distributions:

The normal distribution

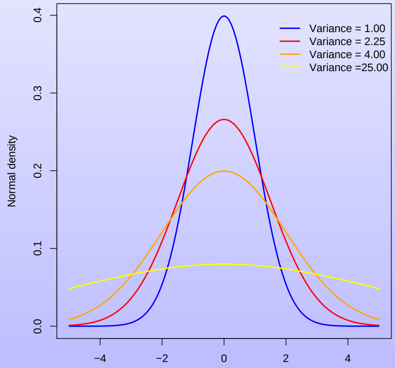
- Normal distribution: continuous distribution depending on two parameters, μ and σ^2 and probability density given by, for each real number x,

$$\phi(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}.$$

  Here μ is a real number and σ is a positive number ($\sigma > 0$).
- $E(X) = \mu$, $Var(X) = \sigma^2$
  The variance is not a function of the mean.

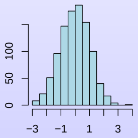
Three key distributions:

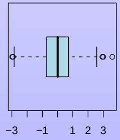
The density of the normal distribution



Three key distributions:

1000 simulated normal distributed variables





Tutorials on distributions expectations and variances (To be done after the lecture)

- Tutorial 03 on probability distributions
- Tutorial 04 - on expectations and variances

Summary and exercises

What you should know

- The notion of random quantity, probability and independency
- What is a random variable and the distribution of a random variable
- The notion and the basic properties of expectation and variance
- The very basic properties of the binomial, Poisson and normal distributions

Exercises and Tutorials (To be done after the lecture)

- Suggested exercises: 1.1, 1.2, 1.4, (1.6), 1.7, 1.8 and 1.9
- Please study the tutorials:
  Tutorial-01- Deterministics X randomness
  Tutorial-02- Probability Distributions
  Tutorial-03- Expectations and Variances

The Master Quiz problem (To be done after the lecture)

- Three boxes:
  One contains a BIG check, the other two are empty
- You choose one box,
  before you open the box the Master-Quiz says
  '*I give you a hint, the check is* **not** *here*'
  and he opens one of the remaining boxes, which is empty
- The Master-Quiz continues:
  Would you like to change and choose the other closed box?
- Question: Is it advantageous to change?