Statistics for Master's students

Random variables and their expected values and (co)variances

Dirk Metzler

April 21, 2026

Contents

1 **Random Variables and Distributions** **2**

2 **Conditional Probabilities and the Bayes-Formula** **3**

3 **The binomial distribution** **6**

4 **Expected value** **8**

5 **Variance and Correlation** **11**

6 **Normal distribution** **18**

7 **Normal approximation** **22**

You sample an indivdual from a population and measure its length X.

X is a **random variable** because it depends on random sampling.
Its **expected value** is in this case the population mean μ:

$$\mathbb{E}X = \mu$$

If you n individuals, all their lengths X_1, X_2,...,X_n are random variables.

Also their mean value $\overline{X} = \frac{1}{n}\sum_i X_i$ and $s = \sqrt{\frac{1}{n-1}\sum_i (X_i - \overline{X})^2}$ are random variables.

Assume a small population of 100 individuals, and a neutral allele A that has frequency 0.3 in this generation.

What will be the frequency X of A in the next generation?

We don't know, as X is a random variable .

However, we can ask, for example, for

$\mathbb{E}X$, the expected value of X, or for

$\Pr(X = 0.32)$, the probability that X takes a value of 0.32.

Even these values (especially the second on) depend on our model assumptions.

Contents

1 Random Variables and Distributions

We start with a simpler Example: Rolling a dice, W is the result of the next trial.

$$\mathcal{S} = \{1, 2, \dots, 6\} \ \Pr(W = 1) = \dots = \Pr(W = 6) = \tfrac{1}{6} \ (\Pr(W = x) = \tfrac{1}{6} \text{ for all } x \in \{1, \dots, 6\} \ )$$

A Random Variable is a result of a random incident or experiment.

The state space $\mathcal{S}$ of a random variable is the set of possible values.

The distribution of a random variable X assigns to each measurable[1] set $A \subseteq \mathcal{S}$ the probability $\Pr(X \in A)$ that X takes a value in A.

In general, we use capitals for random variables ($X, Y, Z, \dots$), and small letters ($x, y, z, \dots$) for (possible) fixed values.

Writing events like sets

The event that X takes a value in A can be written with curly brackets:

$$\{X \in A\}$$

We can interpret this as the set of results (elementary events) for which the event is fulfilled. The intersection

$$\{X \in A\} \cap \{X \in B\} = \{X \in A, X \in B\}$$

is then the event that X takes a value that is in A and in B.
Sometimes curly braces are not always written:

$$\Pr(X \in A, X \in B) = \Pr(\{X \in A, X \in B\}) = \Pr(X \in A \cap B)$$

The join

$$\{X \in A\} \cup \{X \in B\} = \{X \in A \cup B\}$$

is the event that the event that X takes a value in A or in B (or both).

Of course, we can also give events names, e.g.:

$$\begin{aligned} U &:= \{X \in A\}, \qquad V := \{X \in B\} \\ &\Rightarrow \ U \cap V = \{X \in A \cap B\} \end{aligned}$$

Note that if two events contradict each other, e.g.

$$U = \{X \in \{1, 2\}\} \quad V = \{X \in \{3, 4\}\}$$

then

$$U \cap V = \varnothing = \{X \in \emptyset\}$$

where $\varnothing$ is the (impossible) empty event.

Example Rolling a dice W:

$$\begin{aligned} \Pr(W \in \{2, 3\}) &= \frac{2}{6} = \frac{1}{6} + \frac{1}{6} \\ &= \Pr(W = 2) + \Pr(W = 3) \\ \Pr(W \in \{1, 2\} \cup \{3, 4\}) &= \frac{4}{6} = \frac{2}{6} + \frac{2}{6} \\ &= \Pr(W \in \{1, 2\}) + \Pr(W \in \{3, 4\}) \end{aligned}$$

Caution:

$$\begin{aligned} \Pr(W \in \{2, 3\}) + \Pr(W \in \{3, 4\}) &= \frac{2}{6} + \frac{2}{6} = \frac{4}{6} \\ &\neq \Pr(W \in \{2, 3, 4\}) = \frac{3}{6} \end{aligned}$$

[1] Whether non-measurable sets exist goes beyond the scope of this lecture.

Example: rolling two dice (W_1, W_2): Let W_1 and W_2 the result of dice 1 and dice 2.

$$\begin{aligned}
&\Pr(W_1 \in \{4\}, W_2 \in \{2,3,4\}) \\
&= \Pr((W_1, W_2) \in \{(4,2),(4,3),(4,4)\}) \\
&= \frac{3}{36} = \frac{1}{6} \cdot \frac{3}{6} \\
&= \Pr(W_1 \in \{4\}) \cdot \Pr(W_2 \in \{2,3,4\})
\end{aligned}$$

In general:

$$\Pr(W_1 \in A, W_2 \in B) = \Pr(W_1 \in A) \cdot \Pr(W_2 \in B)$$

for all sets $A, B \subseteq \{1, 2, \dots, 6\}$

Calculation rules:

We consider events from a sample space $\mathcal{S}$.

- $0 \leq \Pr(U) \leq 1$ for all events U
- $X \in \mathcal{S}$ and the impossible event $\varnothing$ are events, and $\Pr(X \in \mathcal{S}) = 1$ and $\Pr(\varnothing) = 0$.
- If $U, V \subset \mathcal{S}$ are disjoint, that is $U \cap V = \varnothing$, in other words, they contradict each other, then $U \cup V$ is also an event and:

$$\Pr(U \cup V) = \Pr(U) + \Pr(V)$$

- If $U \cap V \neq \varnothing$, then still $U \cup V$ is also an event and the inclusion-exclusion formula holds:

$$\Pr(U \cup V) = \Pr(U) + \Pr(V) - \Pr(U \cap V)$$

2 Conditional Probabilities and the Bayes-Formula

Definition of conditional probabilities

The probability of U under the condition V

$$\Pr(U|V) := \frac{\Pr(U, V)}{\Pr(V)}$$

"Conditional probability of U given V" Note:

$$\Pr(U, V) = \Pr(V) \cdot \Pr(U|V)$$

Example: Medical Test

Data about breast cancer mammography:

- 0.8% of 50 year old women have breast cancer.
- The mammogram detects breast cancer for 90% of the diseased patients.
- For 7% of the healthy patients, the mammogram gives false alarm.

In an early detection examination with a 50 year old patient, the mammogram indicates breast cancer. What is the probability that the patient really has breast cancer?

This background information was given and the question was asked to 24 experienced medical practitioners. [2].

- 8 of them answered: 90%

[2] Hoffrage, U. & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, **73**, 538-540

- 8 answered: 50 to 80%
- 8 answered: 10% or less.

This is a question about a *conditional probability*: How high is the *conditional* probability to have cancer, *given* that the mammogram indicates it.[2cm]

We can compute conditional probabilities with the Bayes-Formula.

A, B events

The conditional probability of A, given B (assuming $\Pr(B) > 0$):

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

($A \cap B$:= A and B occur)

The theorem of the total probability (with B^c:={B does not occur}):

$$\Pr(A) = \Pr(B)\Pr(A|B) + \Pr(B^c)\Pr(A|B^c)$$

Thomas Bayes,
1702–1761

Bayes-Formula:

$$\Pr(B|A) = \frac{\Pr(B)\Pr(A|B)}{\Pr(A)}$$

Now back to mammography. Define events:

A: The mammogram indicates breast cancer.

B: The patient has breast cancer.

The (unconditioned) probability $\Pr(B)$ is called *prior* probability of B, i.e. the probability that you would assign to B *before* seeing "the data" A. In our case $\Pr(B) = 0.008$ is the probability that a patient coming to the early detection examination has breast cancer.

The conditional probability $\Pr(B|A)$ is called *posterior* probability of B. This is the probability that you assign to B *after* seeing the data A.

The conditional probability that a patient has cancer, given that the mammogram indicates it, is

$$\begin{aligned} \Pr(B|A) &= \frac{\Pr(B) \cdot \Pr(A|B)}{\Pr(A)} = \frac{\Pr(B) \cdot \Pr(A|B)}{\Pr(B) \cdot \Pr(A|B) + \Pr(B^C) \cdot \Pr(A|B^C)} \\ &= \frac{0.008 \cdot 0.9}{0.008 \cdot 0.9 + 0.992 \cdot 0.07} \approx 0.0939. \end{aligned}$$

Thus, the probability that a patient for whom the mammogram indicates cancer has cancer is only 9.4%.

Hoffrage and Giegerenzer (1998) recommend using integers for communication: Among 1000 women, 8 have cancer and it is diagnosed for 7 of them. Among the 992 others there are $992 \cdot 0.07 \approx 69$ with false positive test. $\Rightarrow$ Less than 10 % of the approx. 69+7=76 women with positive test have cancer.

Stochastic Independence

Definition 1 (stochastic independence) *Two events U, V are (stochastically) independent if*

$$\Pr(U, V) = \Pr(U) \cdot \Pr(V).$$

*Two random variables X and Y are (stochastically) independent, if **all** pairs of events of the form $(X \in A, Y \in B)$ for all possible A and B are stochastically independent.*

Example:

- Rolling a dice twice: X = first result, Y = second result.

$$\Pr(X = 2,\ Y = 5) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \Pr(X = 2) \cdot \Pr(Y = 5)$$

If X is a random variable with values in $\mathcal{S}$ and $f : \mathcal{S} \to \mathcal{R}$ is the function (or, more generally, a map), then $f(X)$ is a random variable that depends on X. If X takes the value x, $f(X)$ takes the value $f(x)$.

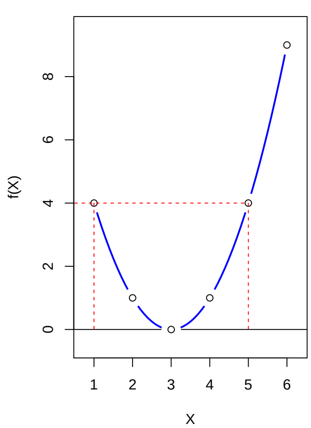
This implies:

$$\Pr(f(X) \in U) = \Pr(X \in f^{-1}(U)),$$

Where $f^{-1}(U)$ is the *inverse image* of U, that is the set of all x such that $f(x) \in U$, formally:

$$f^{-1}(U) = \{x \ : \ f(x) \in U\}$$

(Note the difference between $f^{-1}(\{y\})$ and $f^{-1}(y)$. The latter only exists if f invertible, and is then a number. The first is a set of numbers. Note also that $\{y\}$ is not a number but a set containing one number.)

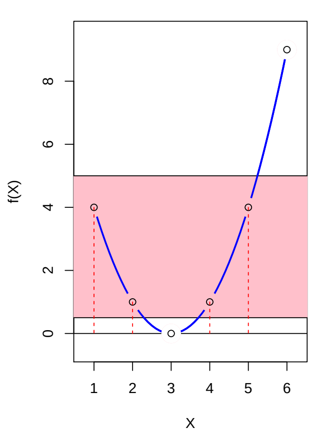


The function $f : x \mapsto (x - 3)^2$ for $x \in \{1, 2, 3, 4, 5, 6\})$ is not invertible. Thus, $f^{-1}(4)$ is not defined, and indeed $f(1) = 4 = f(5)$. However, in $f^{-1}(\{4\})$, the f^{-1} is not an inverse function but the **inverse image function**, which operates on sets:

$$f^{-1}(\{4\}) \ = \ \{x \ : \ f(x) \in \{4\}\} \ = \ \{1, 5\}$$

Or, e.g.:

$$f^{-1}([0.5, 5]) \ = \ \{x \ : \ f(x) \in [0.5, 5]\} \ = \ \{1, 2, 4, 5\}$$



Example: Let f be the function $f(x) = (x-3)^2$, and let X be the result of rolling a dice. (Imagine a game, in which you can move on $f(x)$ steps if the dice shows x pips).Then

$$f^{-1}(\{1\}) = \{2, 4\},$$

and therefore

$$\begin{aligned} \Pr(f(X) = 1) &= \Pr(f(X) \in \{1\}) \\ &= \Pr(X \in f^{-1}(\{1\})) \;=\; \Pr(X \in \{2,4\}) \;=\; \frac{1}{3}. \end{aligned}$$

A simple but important distribution with a continuous state space
Uniform distribution on $[0, 1]$

If U is one of the closed, half-open or open intervals $[a, b]$, $(a, b]$, $[a, b)$ or (a, b) with $0 \leq a \leq b \leq 1$, be

$$\Pr(X \in U) = b - a.$$

We consider events of the form $\{X \in V\}$, where V is a countable join of intervals.

Note that probabilities of "elementary events" $\{X = y\}$ do not help to define $\Pr(X \in V)$, because

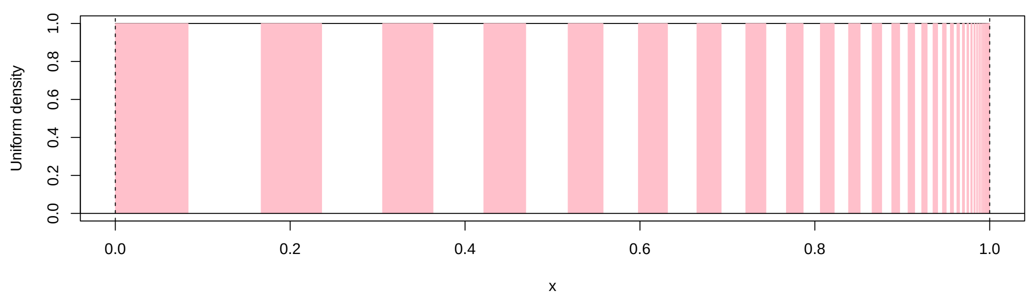
$$\Pr(X = y) \;=\; \Pr(X \in [y, y]) \;=\; y - y \;=\; 0$$

$\Pr(X \in V)$ is defined due to countable additivity, see below.

A rule that is relevant for infinite state spaces
Countable additivity (also called "sigma additivity")

If $A_1, A_2, A_3, \ldots \subset \mathcal{S}$ is a sequence of events such that $\Pr(A_i)$ is defined for each $i \in \{1, 2, 3, \ldots\}$ and $A_i \cap A_j = \emptyset$ holds for each pair (i, j) with $i \neq j$, then

$$\Pr(A_1 \cup A_2 \cup A_3 \cup \ldots) \;=\; \sum_{i=1}^{\infty} \Pr(A_i) \;=\; \lim_{n\to\infty} \sum_{i=1}^{n} \Pr(A_i).$$



3 The binomial distribution

Bernoulli distribution

A Bernoulli experiment is an experiment with two possible outcomes "success" and "fail", or 1 or 0.

Bernoulli random variable X: State space $\mathcal{S} = \{0, 1\}$. Distribution:

$$\Pr(X = 1) = p$$
$$\Pr(X = 0) = 1 - p$$

The parameter $p \in [0, 1]$ is the success probability.

Bernoulli distribution

Examples:

- Tossing a coin: Possible outcomes are "head" and "tail"
- Does the Drosophila have a mutation that causes white eyes? Possible outcomes are "yes" or "no".

Assume a Bernoulli experiment (for example tossing a coin) with success probability p is repeated n times *independently*.

What is the probability that it...

1. ...alway succeeds?
$$p \cdot p \cdot p \cdots p = p^n$$
2. ...always fails?
$$(1-p) \cdot (1-p) \cdots (1-p) = (1-p)^n$$
3. ...first succeeds k times and then fails $n-k$ times?
$$p^k \cdot (1-p)^{n-k}$$
4. ...succeeds in total k times and fails the other $n-k$ times?
$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Note

$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$ ("n choose k") is the number of possibilities to choose k successes in n trials.

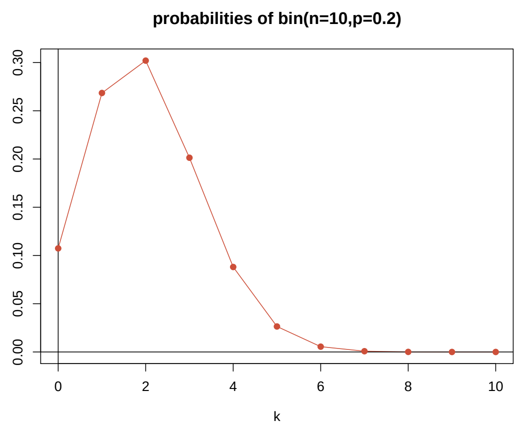
Binomial distribution

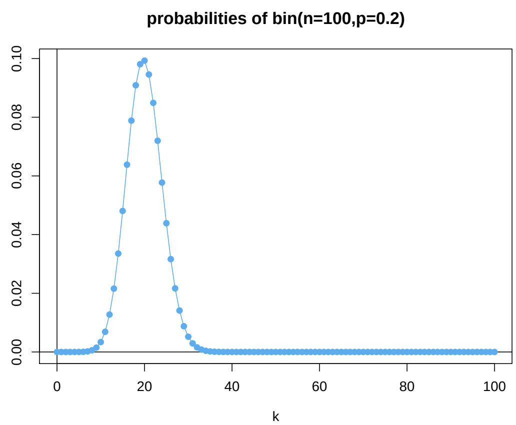
Let X be the number of successes in n independent trials with success probability of p each. Then,

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

holds for all $k \in \{0, 1, \ldots, n\}$ and X is said to be *binomially distributed*, for short:

$$X \sim \text{bin}(n, p).$$





With the binomial distribution we can treat our initial question

Assume in a small population of $n = 100$ individuals the neutral allele A has a frequency of 0.3.

How probable is it that X, the frequency of A in the next generarion is 0.32?

$$\Pr(X = 0.32) = ?$$

We can only answer this on the basis of a probabilistic model, and the answer will depend on how we model the population.

Modeling approach

We make a few simplifying assumptions:

- Discrete generations
- The population is haploid, that is, each individual has exactly one parent in the generation before.
- constant population size $n = 100$

$\Pr(X = 0.32)$ still depends on whether few individuals have many offspring or whether all individuals have similar offspring numbers. $\Pr(X = 0.32)$ is only defined with additional assumptions, e.g.:

- Each individual chooses its parent purely randomly in the generation before.

"purely randomly" means *independent of all others* and *all potential parents with the same probability.*

Our assumptions imply that each individuals of the next generations have a probability of 0.3 to obtain allele A, and they get their alleles independently of each other.

Therefore, the number K of individuels who get allele A is binomially distributed with $n = 100$ and $p = 0.3$:

$$K \sim \text{bin}(n = 100, p = 0.3)$$

For $X = K/n$ follows:

$$\Pr(X = 0.32) = \Pr(K = 32) = \binom{n}{32} \cdot p^{32} \cdot (1-p)^{100-32}$$

$$= \binom{100}{32} \cdot 0.3^{32} \cdot 0.7^{68} \approx 0.078$$

Some of the things you should be able to explain

- Concepts of events, random variables and probabilities, and their notations
- Inclusion-exclusion formula
- How to apply a function to a random variable
- Conditional probabilities
- Stochastic independence of events, and of random variables
- Bayes formula and how to apply it
- Binomial distribution and $\binom{n}{k}$

4 Expected value

Example: genetic and environmental effects

Example: In population on a continent, skin pigmentation S of an individual depends on

- genetic effects G
- environmental effects E (e.g. due to local amount of sunshine)
- random effects R

Simple Model:

$$S = G + E + R$$

S, G, E, R are random variables if they refer to a randomly chosen individual from the population.

Question

Is the population mean of S the sum of the population means of G, E and R?

We need to formalize what population mean means.

General concept: The **expected value** of a random variable.

Definition 2 (Expected value) *Let X be a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \ldots\} \subseteq \mathbb{R}$. The expected value of X is defined by*

$$\mathbb{E}X = \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x)$$

It is also common to write μ_X instead of $\mathbb{E}X$.

If we replace probabilities by relative frequencies in this definition, we get the formula for the mean value (of a sample).

Definition 3 (Expected value) *If X is a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \ldots\} \subseteq \mathbb{R}$, the expected value of X is defined by*

$$\mathbb{E}X = \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x)$$

Examples:

- Let X be Bernoulli distributed with success probability $p \in [0, 1]$. Then we get

$$\mathbb{E}X = 1 \cdot \Pr(X = 1) + 0 \cdot \Pr(X = 0) = \Pr(X = 1) = p$$

- Let W be the result of rolling a dice. Then we get

$$\begin{aligned}\mathbb{E}W &= 1 \cdot \Pr(W = 1) + 2 \cdot \Pr(W = 2) + \ldots + 6 \cdot \Pr(W = 6) \\ &= 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + \ldots + 6 \cdot \tfrac{1}{6} = 21\tfrac{1}{6} = 3.5\end{aligned}$$

Calculating with expectations

Theorem 1 (Linearity of Expectation) *If X and Y are random variables with values in $\mathbb{R}$ and if $a \in \mathbb{R}$, we get:*

- $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$
- $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$

Theorem 2 (Only if independent!) *If X and Y are* ***stochastically independent*** *random variables with values in $\mathbb{R}$, we get*

- $\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y$.

But in general $\mathbb{E}(X \cdot Y) \neq \mathbb{E}X \cdot \mathbb{E}Y$. Example:

$$\mathbb{E}(W \cdot W) = \tfrac{91}{6} = 15.167 > 12.25 = 3.5 \cdot 3.5 = \mathbb{E}W \cdot \mathbb{E}W$$

Note: from $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$ follows $\mathbb{E}(a + Y) = a + \mathbb{E}Y$.
Proof of $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$:

Let $\mathcal{S}$ be the state space of X and $\mathcal{R} = \{a \cdot x \mid x \in \mathcal{S}\} = \{y \mid y/a \in \mathcal{S}\}$ be the state space of $a \cdot X$.

$$\begin{aligned}\mathbb{E}(a \cdot X) &= \sum_{y \in \mathcal{R}} y \cdot \Pr(a \cdot X = y) \\ &= \sum_{y \in \mathcal{R}} y \cdot \Pr(X = y/a) \\ &= a \cdot \sum_{y \in \mathcal{R}} y/a \cdot \Pr(X = y/a) \\ &= a \cdot \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x) \\ &= a \cdot \mathbb{E}X\end{aligned}$$

Theorem 3 *If X is random variable with finite state space $\mathcal{S} \subset \mathbb{R}$, and if $f : \mathbb{R} \to \mathbb{R}$ is a function, we obtain*

$$\mathbb{E}(f(X)) = \sum_{x \in \mathcal{S}} f(x) \cdot \Pr(X = x)$$

Exercise: proof this.
With this, the proof of $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$ can be written as follows:

Let $\mathcal{S}$ be the state space of X and define $f(X) = a \cdot X$.

$$\begin{aligned} \mathbb{E}(a \cdot X) = \mathbb{E}(f(X)) &= \sum_{x \in \mathcal{S}} f(x) \cdot \Pr(X = x) \\ &= \sum_{x \in \mathcal{S}} a \cdot x \cdot \Pr(X = x) \\ &= a \cdot \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x) \\ &= a \cdot \mathbb{E}X \end{aligned}$$

If X and Y are random variables, and Y has a countable state space $\mathcal{S}$, then

$$\Pr(X = x) = \Pr(X = x, Y \in \mathcal{S}) = \sum_{y \in \mathcal{S}} \Pr(X = x, Y = y).$$

We will use this in the next proof.
Proof $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$: To simplify notation we assume that X and Y have the same state space $\mathcal{S}$.

We apply the same theorem as before, this time with $f(x, y) = x + y$, and obtain:

$$\begin{aligned} \mathbb{E}(X + Y) &= \mathbb{E}(f(X, Y)) = \sum_{(x,y) \in \mathcal{S}^2} f(x, y) \cdot \Pr((X, Y) = (x, y)) \\ &= \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} (x + y) \cdot \Pr(X = x, Y = y) \\ &= \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} x \cdot \Pr(X = x, Y = y) + \sum_{y \in \mathcal{S}} \sum_{x \in \mathcal{S}} y \cdot \Pr(X = x, Y = y) \\ &= \sum_{x \in \mathcal{S}} x \cdot \sum_{y \in \mathcal{S}} \Pr(X = x, Y = y) + \sum_{y \in \mathcal{S}} y \cdot \sum_{x \in \mathcal{S}} \Pr(X = x, Y = y) \\ &= \sum_{x \in \mathcal{S}} x \cdot \Pr(X = x) + \sum_{y \in \mathcal{S}} y \cdot \Pr(Y = y) \\ &= \mathbb{E}(X) + \mathbb{E}(Y) \end{aligned}$$

Proof of the product formula: Let $\mathcal{S}$ be the state space of X and Y, and let X and Y be (stochastically) independent.

$$\begin{aligned} &\mathbb{E}(X \cdot Y) \\ &= \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} (x \cdot y) \Pr(X = x, Y = y) \\ &= \sum_{x \in \mathcal{S}} \sum_{y \in \mathcal{S}} (x \cdot y) \Pr(X = x) \Pr(Y = y) \\ &= \sum_{x \in \mathcal{S}} x \Pr(X = x) \cdot \sum_{y \in \mathcal{S}} y \Pr(Y = y) \\ &= \mathbb{E}X \cdot \mathbb{E}Y \cdot \end{aligned}$$

Expectation of the binomial distribution
Let $Y_1, Y_2, \ldots, Y_n$ be the indicator variables of the n independent trials, that is:

$$Y_i = \begin{cases} 1 & \text{if trial } i \text{ succeeds} \\ 0 & \text{if trial } i \text{ fails} \end{cases}$$

Then the Y_i are Bernoulli distributed and $X = Y_1 + \cdots + Y_n$ is binomially distributed with parameters (n, p), where p is the success probability of the trials.

Linearity of expectation implies:

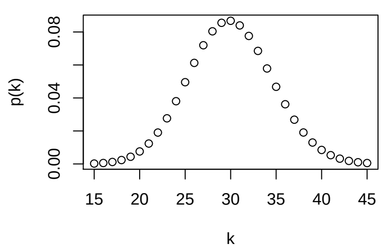
$$\begin{aligned} \mathbb{E}X &= \mathbb{E}(Y_1 + \cdots + Y_n) \\ &= \mathbb{E}Y_1 + \cdots + \mathbb{E}Y_n \\ &= p + \cdots + p \;=\; np \end{aligned}$$

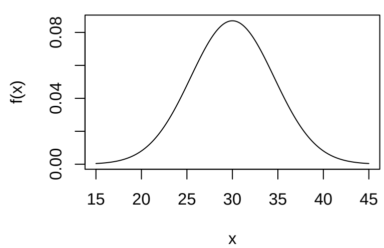
Thus, we obtain:

$$X \sim \text{bin}(n, p) \Rightarrow \mathbb{E}X = n \cdot p$$

Probability distributions on continuous ranges are defined by densities instead of probabilities of single values. Compare, e.g.:

$p(k) = \binom{100}{k} \cdot 0.3^k \cdot 0.7^{100-k}$ $\qquad$ $f(x) = \frac{e^{-(x-30)^2/42}}{42 \cdot \pi}$





In this case, the sum in the definition of $\mathbb{E}$ turns into an integral:

$$\mathbb{E}(K) = \sum_k k \cdot Pr(K = k) \qquad\qquad \mathbb{E}(X) = \int_x x \cdot f(x)\, dx$$

The calculation rules for $\mathbb{E}$ still apply in the continuous case.

5 Variance and Correlation

Question: (for skin pigmentation example)
How does the standard deviation of S depend on the standard deviations of G, E and R?

How to infer σ_S, σ_G, σ_E and σ_R?

σ_S can be estimated from indivduals sampled from the whole population (same probability for each individual).

σ_R can in principle be estimated with genetically identical individuals living in same environment.

But how to measure σ_G and σ_E?

Definition 4 (Variance, Covariance and Correlation) *The Variance of a $\mathbb{R}$-valued random variable X is*

$$VarX = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}X)^2\right].$$

$\sigma_X = \sqrt{Var\ X}$ *is the Standard Deviation.*

If Y is another $\mathbb{R}$-valued random variable,

$$Cov(X,Y) = \mathbb{E}\left[(X - \mathbb{E}X)\cdot(Y - \mathbb{E}Y)\right]$$

is the Covariance of X and Y.
The Correlation of X and Y is

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}.$$

The Variance

$$\mathrm{Var}X = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

is the average squared deviation from the expectation.

The Correlation

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

is always between in the range from -1 to 1. The random variables X and Y are

- positively correlated, if X and Y tend to be both above average or both below average.
- negatively correlated, if X and Y tend to deviate from their expected values in opposite ways.

If X and Y are independent, they are also uncorrelated, that is $\mathrm{Cor}(X,Y) = 0$.

Example: rolling dice

Variance of result from rolling a dice W:

$$\begin{aligned}
\mathrm{Var}(W) &= \mathbb{E}\big[\big(W - \mathbb{E}W\big)^2\big] \\
&= \mathbb{E}\big[\big(W - 3.5\big)^2\big] \\
&= (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + \ldots + (6-3.5)^2 \cdot \frac{1}{6} \\
&= \frac{17.5}{6} \\
&= 2.91667
\end{aligned}$$

Example: Empirical Distribution

If $x_1, \ldots, x_n \in \mathbb{R}$ are data and if X is the result of a random draw from the data (such that $\Pr(X = x) = \frac{n_x}{n}$, where n_x is the number of x_i that are equal to x, formally $n_x = |\{i \ : \ x_i = x\}|$), we get:

$$\mathbb{E}X = \sum_x x \cdot \frac{n_x}{n} = \frac{1}{n}\sum_x x \cdot n_x = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

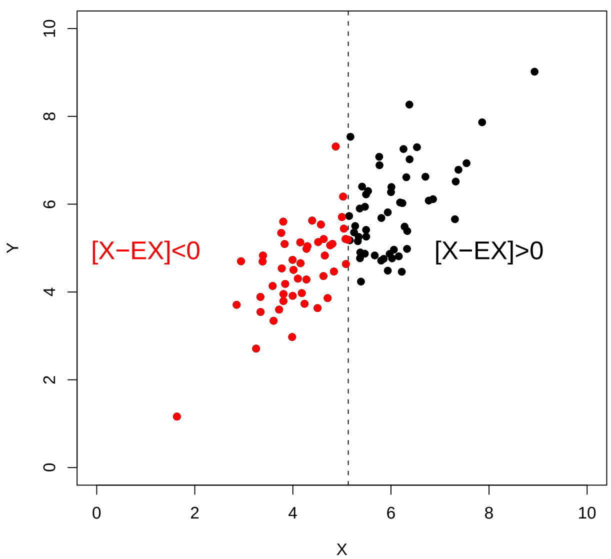
and

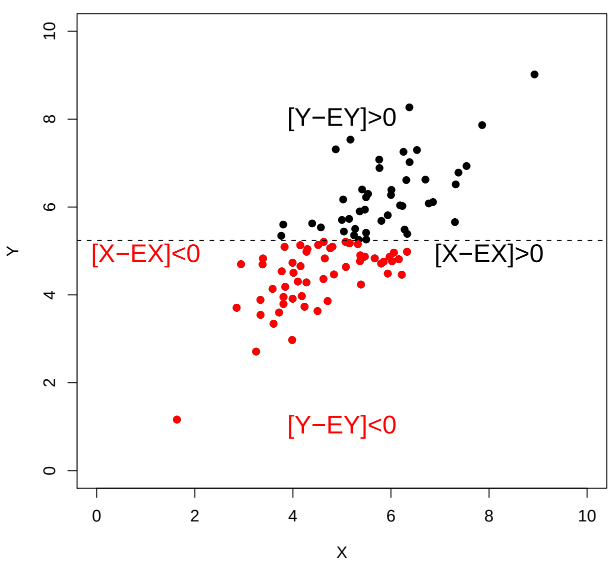
$$\mathrm{Var}\ X = \mathbb{E}\big[\big(X - \mathbb{E}X\big)^2\big] = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

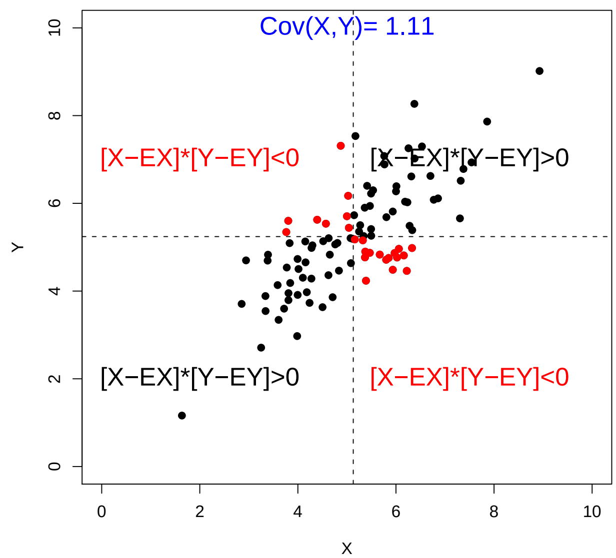
If $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R} \times \mathbb{R}$ are data if (X,Y) are drawn from the data such that $\Pr((X,Y) = (x,y)) = \frac{|\{i \ : \ (x_i, y_i) = (x,y)\}|}{n}$, we get

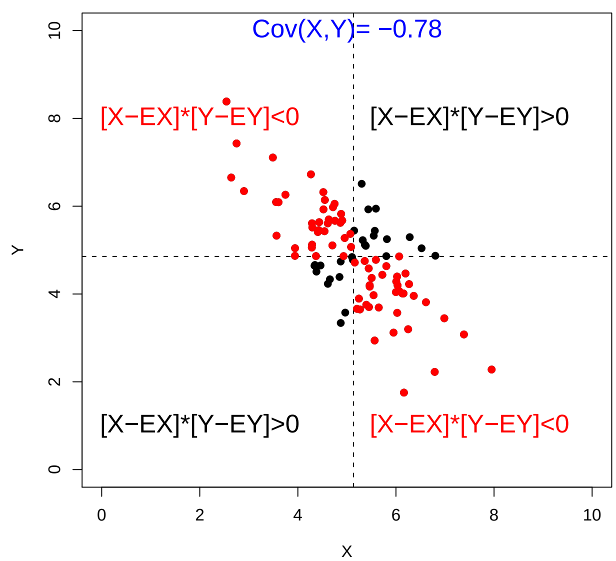
$$\mathrm{Cov}\ (X,Y) = \mathbb{E}\big[\big(X - \mathbb{E}X\big)\big(Y - \mathbb{E}Y\big)\big] = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

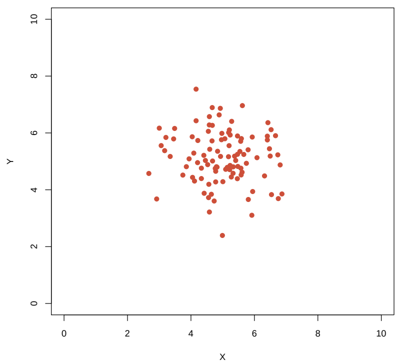
Why Cov$(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$**?**



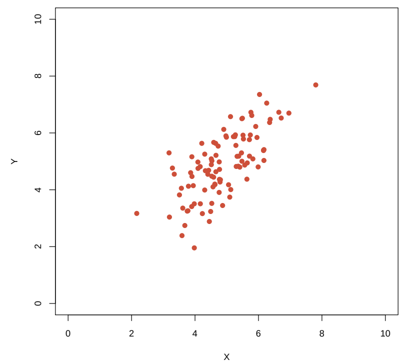



Cov(X,Y)= 1.11
[X−EX]*[Y−EY]<0
[X−EX]*[Y−EY]>0
[X−EX]*[Y−EY]>0
[X−EX]*[Y−EY]<0
0
2
4
6
8
10
0
2
4
6
8
10
X
Y

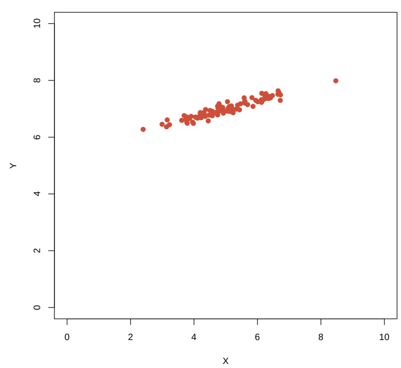

Cov(X,Y)= −0.78
[X−EX]*[Y−EY]<0
[X−EX]*[Y−EY]>0
[X−EX]*[Y−EY]>0
[X−EX]*[Y−EY]<0
0
2
4
6
8
10
0
2
4
6
8
10
X
Y



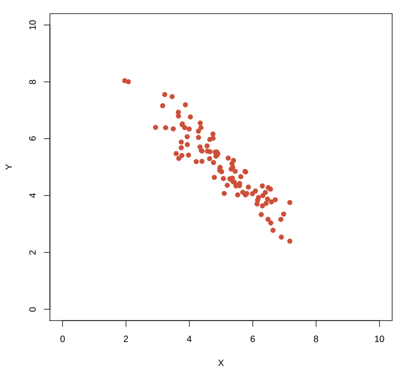
$\sigma_X = 0.95$, $\sigma_Y = 0.92$
$\mathrm{Cov}(X, Y) = -0.06$
$\mathrm{Cor}(X, Y) = -0.069$



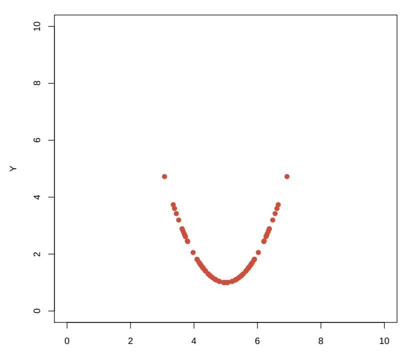
$\sigma_X = 1.14$, $\sigma_Y = 0.78$
$\mathrm{Cov}(X, Y) = 0.78$
$\mathrm{Cor}(X, Y) = 0.71$



$\sigma_X = 1.03$, $\sigma_Y = 0.32$
$\mathrm{Cov}(X, Y) = 0.32$
$\mathrm{Cor}(X, Y) = 0.95$



$\sigma_X = 1.13$, $\sigma_Y = 1.2$
$\mathrm{Cov}(X, Y) = -1.26$
$\mathrm{Cor}(X, Y) = -0.92$



$\sigma_X = 0.91$, $\sigma_Y = 0.88$
$\mathrm{Cov}(X, Y) = 0$
$\mathrm{Cor}(X, Y) = 0$

Calculation rules for Covariances

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X) \cdot (Y - \mathbb{E}Y)]$$

- If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$ (but not the other way around!)
- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- $\mathrm{Cov}(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$ (Exercise!)
- $\mathrm{Cov}(a \cdot X, Y) = a \cdot \mathrm{Cov}(X,Y) = \mathrm{Cov}(X, a \cdot Y)$
- $\mathrm{Cov}(X + Z, Y) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(Z,Y)$
- $\mathrm{Cov}(X, Z + Y) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(X,Y)$

The last three rules describe the bilinearity of covariance.

Calculation rules for Correlations

$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$

- $-1 \leq \mathrm{Cor}(X,Y) \leq 1$
- $\mathrm{Cor}(X,Y) = \mathrm{Cor}(Y,X)$
- $\mathrm{Cor}(X,Y) = \mathrm{Cov}(X/\sigma_X, Y/\sigma_Y)$
- $\mathrm{Cor}(X,Y) = 1$ if and only if Y is an increasing, affine-linear function of X, that is, if $Y = a \cdot X + b$ for appropriate $a > 0$ and $b \in \mathbb{R}$.
- $\mathrm{Cor}(X,Y) = -1$ if and only if Y is an decreasing, affine-linear function of X, that is, if $Y = a \cdot X + b$ for appropriate $a < 0$ and $b \in \mathbb{R}$.

Calculation rules for variances

$\mathrm{Var}X = \mathbb{E}[(X - \mathbb{E}X)^2]$

- $\mathrm{Var}X = \mathrm{Cov}(X,X)$
- $\mathrm{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ (Exercise!)
- $\mathrm{Var}(a \cdot X) = a^2 \cdot \mathrm{Var}X$
- $\mathrm{Var}(X + Y) = \mathrm{Var}X + \mathrm{Var}Y + 2 \cdot \mathrm{Cov}(X,Y)$ $(\Rightarrow \mathrm{Var}(X + a) = \mathrm{Var}(X))$
- $\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{Var}\big(X_i\big) + 2 \cdot \sum_{j=1}^{n} \sum_{i=1}^{j-1} \mathrm{Cov}(X_i, X_j)$
- If (X,Y) stochastically independent we get:

$$\mathrm{Var}(X + Y) = \mathrm{Var}X + \mathrm{Var}Y$$

Question for skin pigmentation example: How does the standard deviation of S depend on the standard deviations of G, E and R?

Answer: $\sigma_S = \sqrt{\mathrm{Var}(S)}$, and

$$\begin{aligned}\mathrm{Var}(S) &= \mathrm{Var}(G) + \mathrm{Var}(E) + \mathrm{Var}(R) + 2 \cdot \mathrm{Cov}(G,E) + \\ &\quad +2 \cdot \mathrm{Cov}(G,R) + 2 \cdot \mathrm{Cov}(E,R)\end{aligned}$$

Perhaps we may assume $\mathrm{Cov}(G,R) = \mathrm{Cov}(E,R) = 0$, but $\mathrm{Cov}(G,E) > 0$ is plausible as individuals who live in more sunny areas may have genes for darker pigmentation.

So, how to measure σ_G and σ_E?
(at least in principle)

Var(R)**:** infer from genetically identically individuals in same environment

Var$(G+R)$**:** infer from individuals sampled from whole population but exposed to same environment

Var$(E+R)$**:** infer from genetically identically individuals exposed to random environments

If $\mathrm{Cov}(G,R) = \mathrm{Cov}(E,R) = 0$, then

$$\begin{aligned}\sigma_G &= \sqrt{\mathrm{Var}(G+R) - \mathrm{Var}(R)} \qquad \text{and} \\ \sigma_E &= \sqrt{\mathrm{Var}(E+R) - \mathrm{Var}(R)}.\end{aligned}$$

Bernoulli distribution

A Bernoulli distributed random variable Y with success probability $p \in [0,1]$ has expected value

$$\mathbb{E}Y = p$$

and variance

$$\mathrm{Var}\ Y = p \cdot (1-p)$$

Proof: From $\Pr(Y=1) = p$ and $\Pr(Y=0) = (1-p)$ follows

$$\mathbb{E}Y = 1 \cdot p + 0 \cdot (1-p) = p.$$

variance:

$$\begin{aligned}\mathrm{Var}\ Y &= \mathbb{E}\big(Y^2\big) - \big(\mathbb{E}Y\big)^2 \\ &= 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p \cdot (1-p)\end{aligned}$$

Binomial distribution

Let $Y_1, \cdots, Y_n$ be independent Bernoulli distributed with success probability p. Then follows

$$\sum_{i=1}^{n} Y_i =: X \sim \mathrm{bin}(n,p)$$

and we get:

$$\mathrm{Var}\ X = \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \mathrm{Var}\ Y_i = n \cdot p \cdot (1-p)$$

Binomial distribution

Theorem 4 (Expected value and variance of the binomial distribution) *If X is binomially distributed with parameters (n,p), we get:*

$$\mathbb{E}X = n \cdot p$$

und

$$Var\ X = n \cdot p \cdot (1-p)$$

Example: Genetic Drift

In a haploid population of n individuals, let p be the frequency of some allele A. We assume that (due to some simplifying assumptions?) the absolute frequency K of A in the next generation is (n,p)-binomially distributed.

For $X = K/n$, the relative frequency in the next generation follows:

$$\begin{aligned}\mathrm{Var}(X) = \mathrm{Var}(K/n) = \mathrm{Var}(K)/n^2 &= n \cdot p \cdot (1-p)/n^2 \\ &= \frac{p \cdot (1-p)}{n}\end{aligned}$$

Example: Genetic Drift

If we consider the change of allele frequencies over m generations, the variances add up. If m is a small number, such that p will not change much over m generations, the variance of change of allele frequencies is approximately

$$m \cdot \mathrm{Var}(X) = \frac{m \cdot p \cdot (1-p)}{n}$$

(because the changes per generation are independent of each other) and thus, the standard deviation is about

$$\sqrt{\frac{m}{n} \cdot p \cdot (1-p)}$$

Remember what genetic drift is and what it is not

Random events that contribute to genetic drift:

- random number of offspring of individuals
- which copy of a chromosome is passed on
- random death of a juvenile individual

What is **not** considered as part of genetic drift:

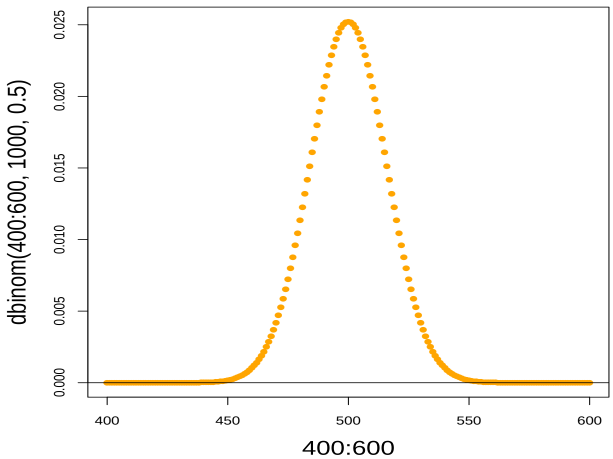
- mutations
- selection and fitness differences
- catastrophic events in which a large part of the population dies

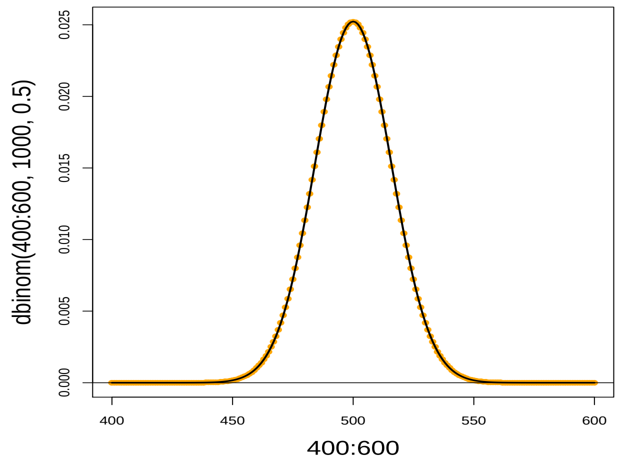
Some of the things you should be able to explain

- Definitions of $\mathbb{E}$, Var, Cov, Cor for random variables
- Calculation rules for $\mathbb{E}$, Var, Cov, Cor and how to use them
- Difference between correlation and stochastic dependence
- $\mathbb{E}$ and Var (and SD) of the binomial distribution
- how genetic drift depends on population size and allele frequency
- basic principles and ideas of the proofs in this section

6 Normal distribution

A binomial distribution with large n looks like a normal distribution:

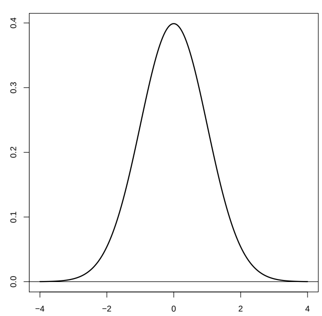




Density of the standard normal distribution

A random variable Z with the density

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



"Gaussian bell-curve"

for short:
$Z \sim \mathcal{N}(0,1)$

$$\mathbb{E}Z = 0$$
$$\text{Var } Z = 1$$

is called *standard-normally distributed*.

If Z is $\mathcal{N}(0,1)$ distributed, then $X = \sigma \cdot Z + \mu$ is normally distributed with mean μ and variance σ^2, for short:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

X has the density

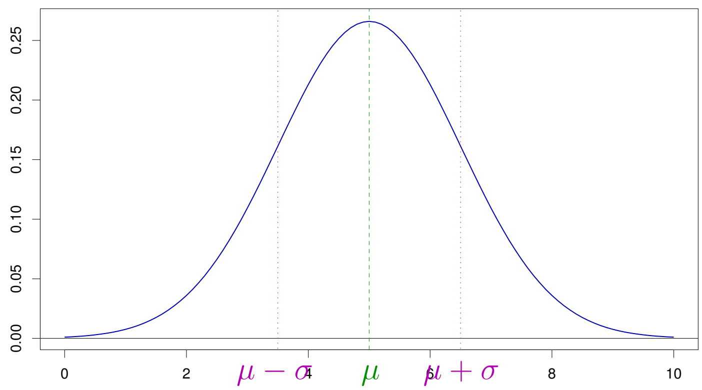
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Always have in mind:

If $Z \sim \mathcal{N}(\mu, \sigma^2)$, we get:

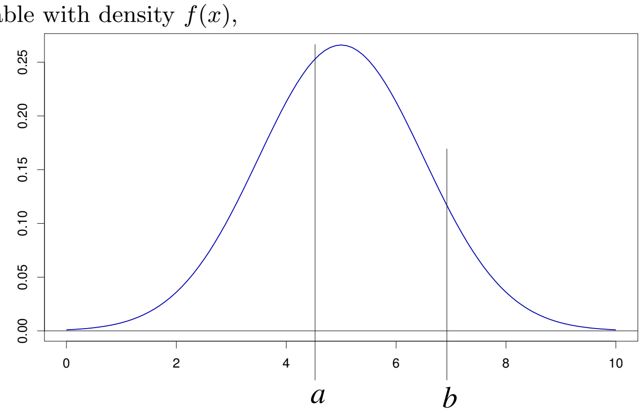
- $\Pr(|Z - \mu| > \sigma) \approx 33\%$
- $\Pr(|Z - \mu| > 1.96 \cdot \sigma) \approx 5\%$
- $\Pr(|Z - \mu| > 3 \cdot \sigma) \approx 0.3\%$

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



Densities need Integrals

If Z is a random variable with density $f(x)$,



we get

$$\Pr(Z \in [a,b]) = \int_a^b f(x)\mathsf{d}x.$$

Note: the probability density f **is not** the probability distribution of Z, but the probability distribution

$$A \mapsto \Pr(Z \in A)$$

can be calculated from the probability density:

$$A \mapsto \Pr(Z \in A) = \int_A f(x)dx$$

Question: How to compute $\Pr(Z = 5)$?

Answer: For each $x \in \mathbb{R}$ we have $\Pr(Z = x) = 0$ (Area of width 0)

What happens with $\mathbb{E}Z = \sum_{x \in \mathcal{S}} x \cdot \Pr(Z = x)$?

For a continuous random variable with density f we define:

$$\mathbb{E}Z := \int_{-\infty}^{\infty} x \cdot f(x)\mathsf{d}x$$

The $\mathbb{E}$-based defintions of Var, Cov, Cor still apply, e.g.:

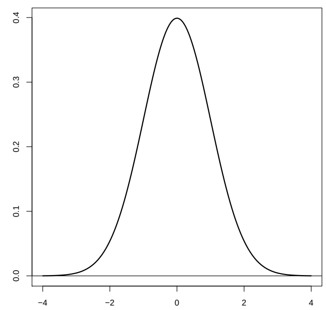
$$\mathrm{Var}(Z) = \mathbb{E}(Z - \mathbb{E}Z)^2 = \mathbb{E}Z^2 - (\mathbb{E}Z)^2$$

The normal distribution in *R*

dnorm(): **d**ensity of the normal distribution
rnorm(): drawing a **r**andom sample
pnorm(): **p**robability function of the normal distribution
qnorm(): **q**uantile function of the normal distribution

example: density of the standard normal distribution:

```
> plot(dnorm,from=-4,to=4)
```




```
> dnorm(0) [1] 0.3989423 > dnorm(0,mean=1,sd=2) [1] 0.1760327
```

example: drawing a sample

draw a sample of length 6 from standard normal:

```
> rnorm(6) [1] -1.24777899 0.03288728 0.19222813 0.81642692 -0.62607324 -1.09273888
```

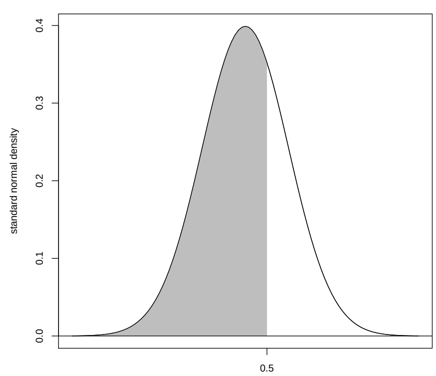
draw a sample of length 7 from standard normal with expected value 5 and standard deviation 3:

```
> rnorm(7,mean=5,sd=3) [1] 2.7618897 6.3224503 10.8453280 -0.9829688 5.6143127 0.6431437 8.123570
```

example: Computing probabilities: Let $Z \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$ be standard normally distributed

$\Pr(Z < a)$ can be computed in R by `pnorm(a)`

```
> pnorm(0.5) [1] 0.6914625
```

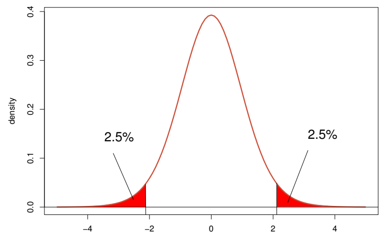


example: Computing probabilities: Let $Z \sim \mathcal{N}(\mu = 5, \sigma^2 = 2.25)$.

Computing $\Pr(Z \in [3, 4])$:

$$\Pr(Z \in [3, 4]) = \Pr(Z < 4) - \Pr(Z < 3)$$

`> pnorm(4,mean=5,sd=1.5)-pnorm(3,mean=5,sd=1.5) [1] 0.1612813` **example**: Computing quantiles: Let $Z \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$ be standard normally distributed. For which value z holds $\Pr(|Z| > z) = 5\%$?



From the symmetry around the y-axis follows

$$\Pr(|Z| > z) = \Pr(Z < -z) + \Pr(Z > z) = 2 \cdot \Pr(Z < -z)$$

So find $z > 0$, such that $\Pr(Z < -z) = 2.5\%$. `> qnorm(0.025,mean=0,sd=1) [1] -1.959964` Answer: $z \approx 1.96$, just below 2 standard deviations.

7 Normal approximation

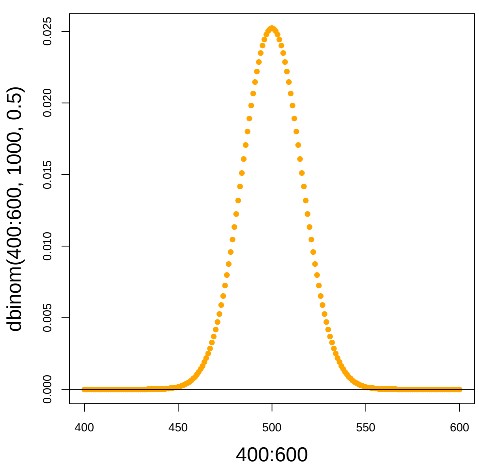
Normal approximation

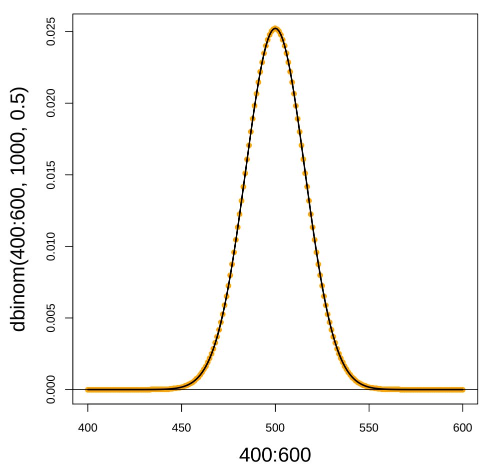
For large n and p which are not too close to 0 or 1, we can approximate the binomial distribution by a normal distribution with the corresponding mean and variance.

If $X \sim \text{bin}(n, p)$ and $Z \sim \mathcal{N}(\mu = n \cdot p, \sigma^2 = n \cdot p \cdot (1-p))$, we get

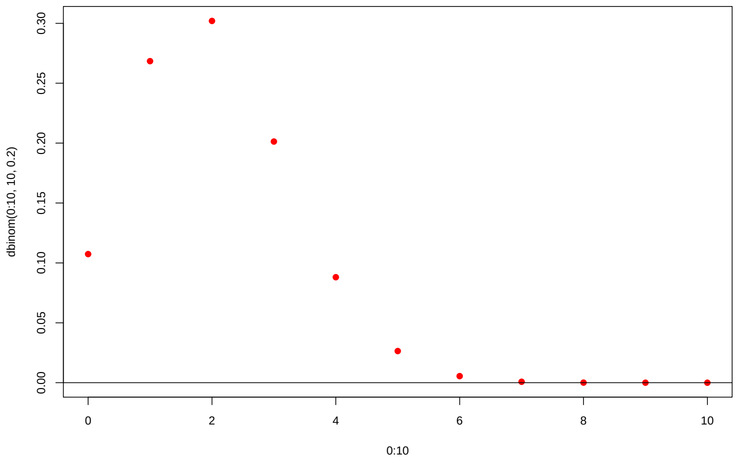
$$\Pr(X \in [a, b]) \approx \Pr(Z \in [a, b])$$

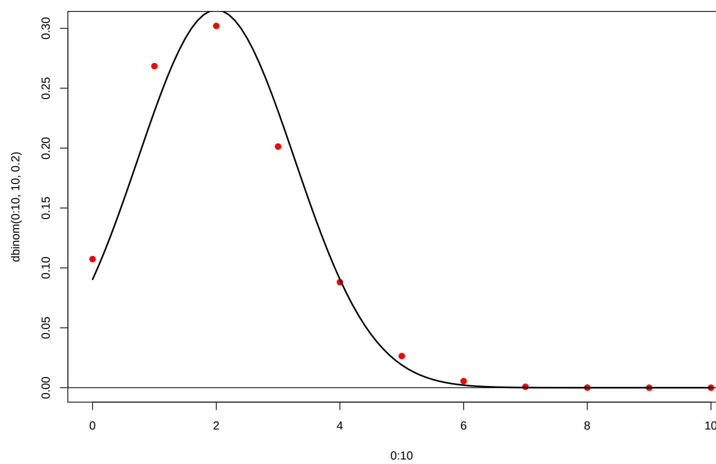
(rule of thumb: Usually okay if $n \cdot p \cdot (1-p) \geq 9$)
$n = 1000$, $p = 0.5$, $n \cdot p \cdot (1-p) = 250$





$n = 10$, $p = 0.2$, $n \cdot p \cdot (1-p) = 1.6$





Theorem 5 (Central Limit Law) *If the $\mathbb{R}$-valued random variables $X_1, X_2, \ldots$ are independent and identically distributed with finite variance $0 < Var\ X_i < \infty$ and if*

$$Z_n := X_1 + X_2 + \cdots + X_n$$

is the sum of the first n variables, then the centered and rescaled sum is in the limit $n \to \infty$ standard-normally distributed:

$$\frac{Z_n - \mathbb{E}Z_n}{\sqrt{Var\ Z_n}} \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$$

for $n \to \infty$. Formally: For all $-\infty \leq a < b \leq \infty$ holds

$$\lim_{n\to\infty} \Pr\left(a \leq \frac{Z_n - \mathbb{E}Z_n}{\sqrt{Var\ Z_n}} \leq b\right) = \Pr(a \leq Z \leq b),$$

where Z is a standard-normally distributed random variable.

In other words: For large n holds:

$$Z_n \sim \mathcal{N}\big(\mu = \mathbb{E}Z_n, \sigma^2 = \text{Var } Z_n\big)$$

The requirements "independent" and "identically distributed" can be diluted.

Usually holds:

If Y is the sum of many small contributions, most of which are independent of each other, then Y is approximately normally distributed.

that is

$$Y \sim \mathcal{N}\big(\mu = \mathbb{E}Y, \sigma^2 = \text{Var } Y\big)$$

Some of the things you should be able to explain

- Probability densities and how to get probability distributions from them
- when and how to approximate binomial by normal distribution
- Properties of the normal distribution (μ,σ^2, important quantiles,...)
- normal distribution of $a \cdot X + b$ if X is normally distributed

- meaning of the central limit law
- R commands to deal with probability distributions

Note also the lists on pages 8 and 18.