Statistics for Master's students

Basics from Stochastics

Dirk Metzler

April 27, 2020

Contents

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

X is a **random variable** because it depends on random sampling.
Its **expectation 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 expectation 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 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.
The join

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

is the event that the event that X takes a value in A or in B (or both).
Sometimes the curly brackets are not written:

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

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

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

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

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

then

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

where $\emptyset$ is the (impossible) empty event (for which we use the same symbol as for the empty set).

If fact, events are (certain) subsets of a so-called sample space Ω. For example, if X is the result of rolling a dice, then

$$\Omega = \Big\{\{X = 1\}, \{X = 2\}, \{X = 3\}, \{X = 4\}, \{X = 5\}, \{X = 6\}\Big\}$$

- In case like this with a finite Ω, als subsets of Ω are also events, and their probabilities are just the sums of their elements.

- For infinite Ω things become more complicated:
  - Events can have non-zero probability even if all their elements have zero probability.
  - We cannot assume that all subsets of Ω are events (mathematical details are complicated).
- A probability distribution assigns to each event $U \subseteq \Omega$ a probability $\Pr(U)$ that the event takes place.

Example for an infinite 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

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

The state space Ω consists of all 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.

An important axiom for infinite state spaces

Countable additivity (also called "sigma additivity")

If $A_1, A_2, A_3, \ldots \subset \Omega$ 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

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

Back to finite state spaces:

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:

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

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 $A, B \subseteq \{1, 2, \ldots, 6\}$

If S is the sum of the results $S = W_1 + W_2$, what is the probability that $S = 5$, if dice 1 shows $W_1 = 2$?

$$\Pr(S = 5|W_1 = 2) \overset{!}{=} \Pr(W_2 = 3)$$
$$= \tfrac{1}{6} = \tfrac{1/36}{1/6} = \tfrac{\Pr(S=5,W_1=2)}{\Pr(W_1=2)}$$

What is the probability $S \in \{4, 5\}$ under the condition $W_1 \in \{1, 6\}$?

$$\begin{aligned}
&\Pr(S \in \{4, 5\}|W_1 \in \{1, 6\}) \\
&= \frac{\Pr(S \in \{4, 5\}, W_1 \in \{1, 6\})}{\Pr(W_1 \in \{1, 6\})} \\
&= \frac{\Pr(W_2 \in \{3, 4\}, W_1 = 1)}{\Pr(W_1 \in \{1, 6\})} \\
&= \frac{2/36}{2/6} = \frac{1}{6}
\end{aligned}$$

Calculation rules:

We consider events from a sample space Ω.

- $0 \leq \Pr(U) \leq 1$ for all events $U \in \Omega$
- Ω and the impossible event $\emptyset$ are events, and $\Pr(\Omega) = 1$ and $\Pr(\emptyset) = 0$.
- If $U, V \subset \Omega$ are disjoint, that is $U \cap V = \emptyset$, 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 $U \cap V \neq \emptyset$, then still $U \cup V$ is also an event and the inclusion-exclusion formula holds:

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

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

How to say

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

in words:

The probability that both U and V take place can be computed in two steps:

- For $U \cap V$, the event V must take place.
- Multiply the probability of V with the conditional probability of U, given that V is already known to take place. (Not relevant are the time points *when* it turns out that U or V take place.)

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:

- Tossing two dice: X = result dice 1, Y = result dice 2.

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

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 $f(x)$.

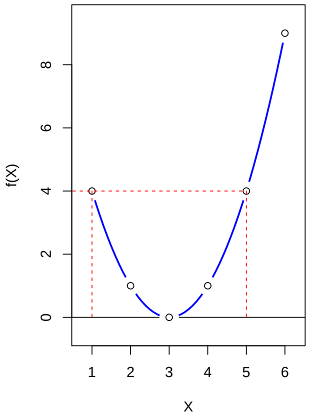
This implies:

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

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

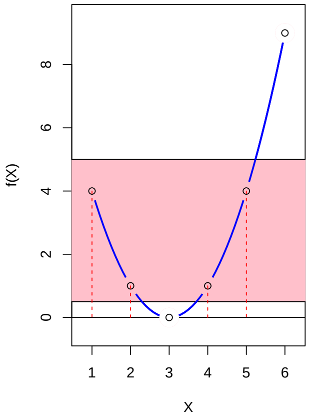
$$f^{-1}(U) = \{x \ : \ f(x) \in U\}$$

(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

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

2 Conditional Probabilities and the Bayes-Formula

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. [1].

- 8 of them answered: 90%
- 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:

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

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

Example: Let $W \in \{1,2,3,4,5,6\}$ be the result of rolling a dice. How probable is $W \geq 5$ if W is an even number?

$$\begin{aligned} A &:= \{W \geq 5\} \\ B &:= \{W \text{ is even }\} \\ A \cap B &= \{W \text{ is even and } \geq 5\} \end{aligned}$$

	A	**A^c**
B	⚅	⚁ ⚃
B^c	⚄	⚀ ⚂

[0.5cm]

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{1/6}{3/6} = \frac{1}{3}$$

$$\Pr(B|A) = \frac{\Pr(B) \cdot \Pr(A|B)}{\Pr(A)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{1/3} = \frac{1}{2}$$

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.[0.5cm] 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%. The right answer "approximately 10%" was only given by 4 of the 24 medical practitioners. Two of them gave an explanation that was so fallacious that we have to assume that they gave the right answer only by accident.

The Monty Hall problem

The Monty Hall problem (the goat problem)

- In the US-American TV-Show *Let's Make a Deal* the candidate can win a sports car at the end of the show if he or she selects the right one of three doors.
- Behind the two wrong doors there are goats.
- The candidate first selects one of the three doors, let's say door 1.
- The host of the show, Monty Hall, then says "I show you something" and opens one of the two other doors, let's say door 2. A goat is standing behind this door.
- The candidate can then stay with door 1 or switch to door 3.
- Should he switch to door 3?

A : The host opens door 2.

B : The car is behind door 3.

C : The car is behind door 1.

D : The car is behind door 2.

$\Pr(B) = 1/3 = \Pr(C) = \Pr(D)$ $\Pr(A|B) = 1$, $\Pr(A|C) = 1/2$, $\Pr(A|D) = 0$.

$$\begin{aligned}\Pr(B|A) &= \frac{\Pr(B)\cdot\Pr(A|B)}{\Pr(B)\cdot\Pr(A|B)+\Pr(C)\cdot\Pr(A|C)+\Pr(D)\cdot\Pr(A|D)}\\ &= \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot 1+\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot 0}\\ &= 2/3\end{aligned}$$

Thus, it is advisable to switch to door 3.

3 The binomial distribution

Bernoulli distribution

A Bernoulli experiment is an experiment with two possible oucomes “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.

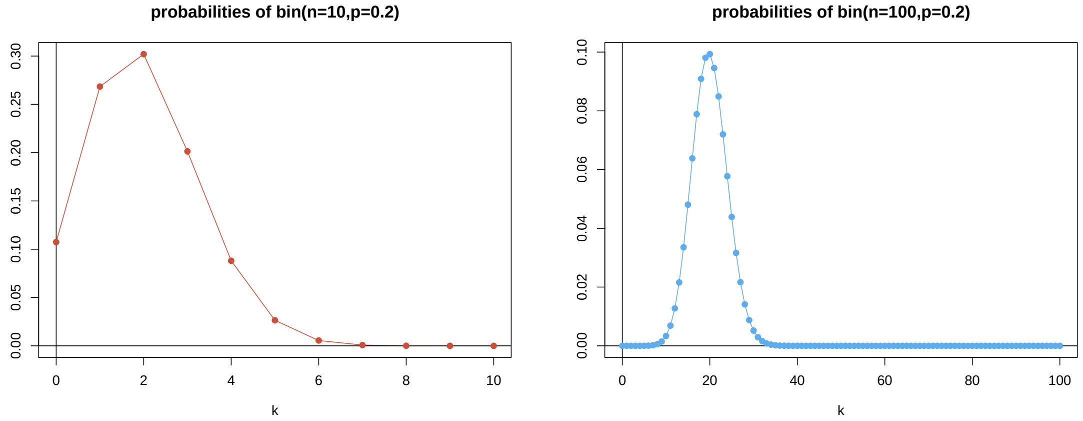
Binomial distribution

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 Expectation 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 **expectation value** of a random variable.

Definition 2 (Expectation value) *Let X be a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \dots\} \subseteq \mathbb{R}$. The expectation 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 (Expectation value) *If X is a random variable with finite or countable state space $\mathcal{S} = \{x_1, x_2, x_3 \dots\} \subseteq \mathbb{R}$, the expectation 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:

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

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.
Proof of $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$:

With $f(x) = a \cdot x$ we obtain from the previous Theorem:

$$\begin{aligned}\mathbb{E}(a \cdot X) = \mathbb{E}(f(X)) &= \sum_x f(x) \cdot \Pr(X = x) \\ &= \sum_x a \cdot x \cdot \Pr(X = x) \\ &= a \cdot \sum_x x \cdot \Pr(X = x) \\ &= a \cdot \mathbb{E}X\end{aligned}$$

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

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

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

Linearity of expectation implies

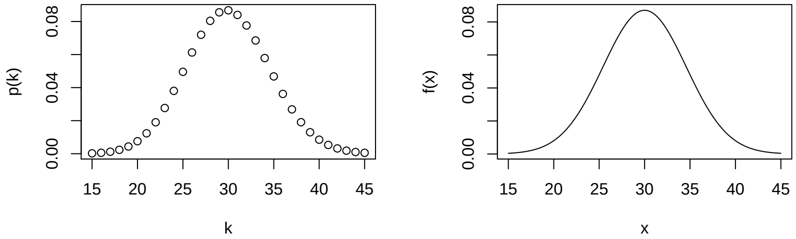
$$\begin{aligned}
\mathbb{E}X &= \mathbb{E}(Y_1+\cdots+Y_n) \\
&= \mathbb{E}Y_1+\cdots+\mathbb{E}Y_n \\
&= p+\cdots+p \;=\; np
\end{aligned}$$

Note:

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

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 expectation 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:

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

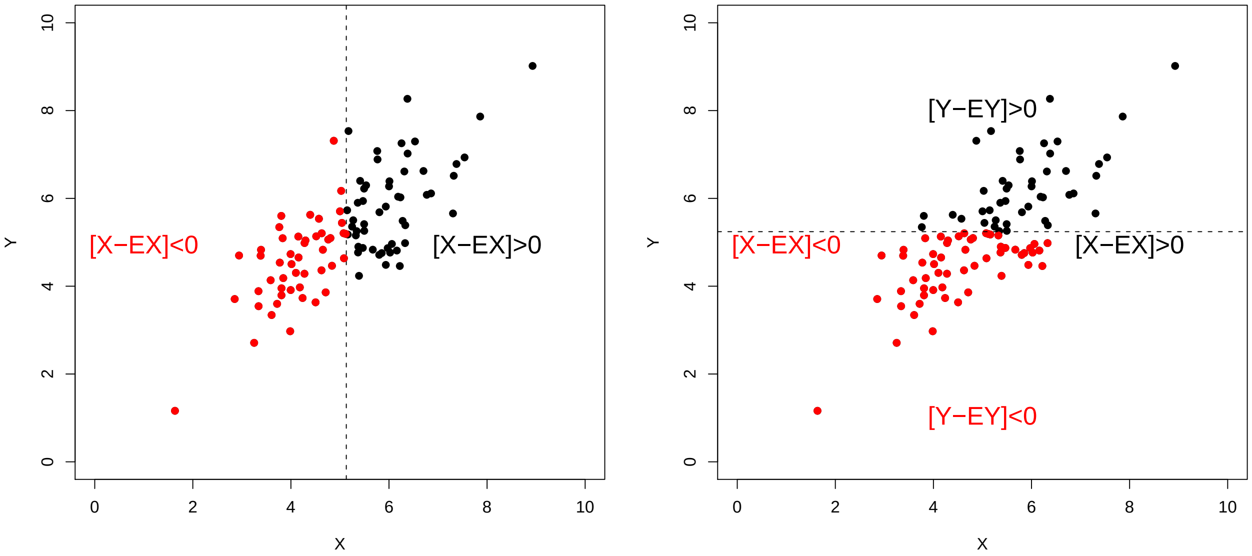
and

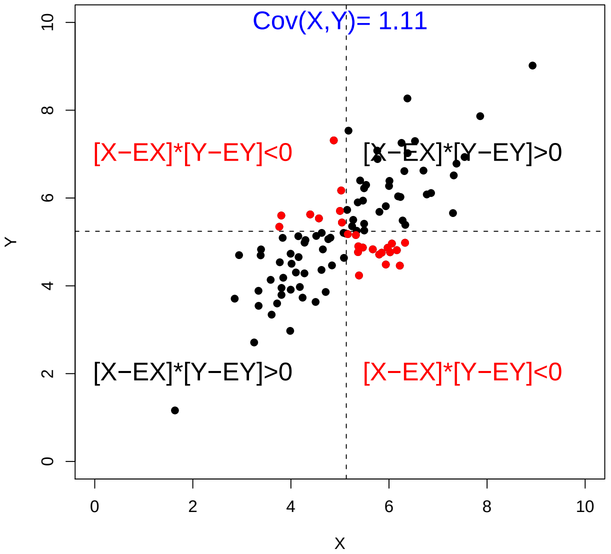
$$\mathrm{Var}\ X=\mathbb{E}\big[\big(X-\mathbb{E}X\big)^2\big]=\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})^2$$

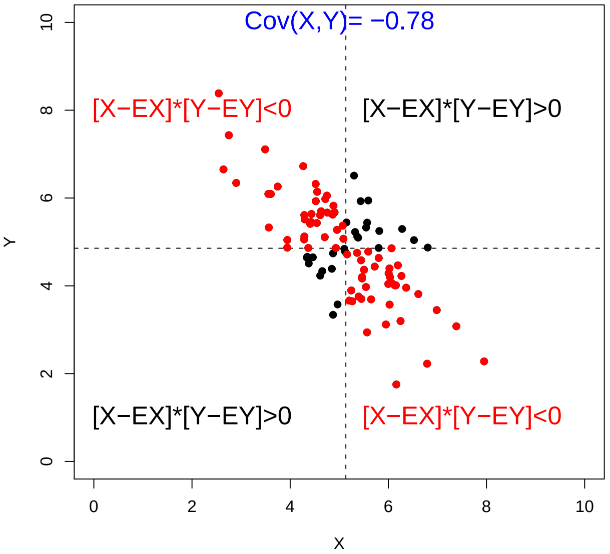
If $(x_1,y_1),\ldots,(x_n,y_n)\in\mathbb{R}\times\mathbb{R}$ are data if (X,Y) are drawn from the data such that $\Pr((X,Y)=(x,y))=\frac{|\{i\ :\ (x_i,y_i)=(x,y)\}|}{n}$, we get

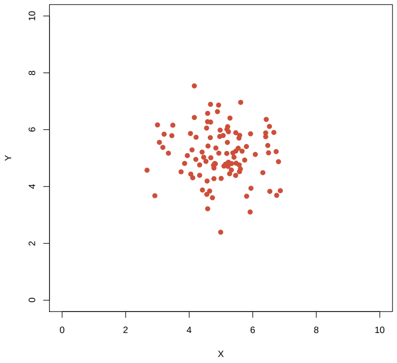
$$\mathrm{Cov}\ (X,Y)=\mathbb{E}\big[\big(X-\mathbb{E}X\big)\big(Y-\mathbb{E}Y\big)\big]=\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})(y_i-\overline{y})$$

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

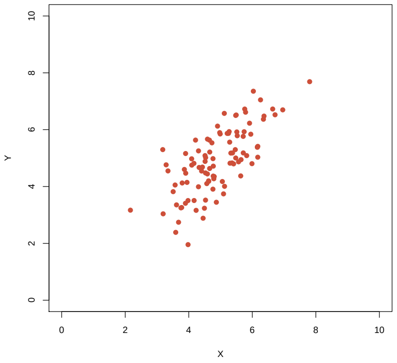




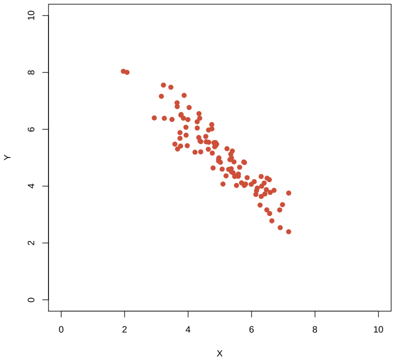




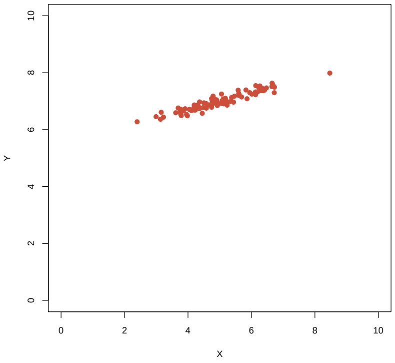
$\sigma_X = 0.95$, $\sigma_Y = 0.92$
$\text{Cov}(X, Y) = -0.06$
$\text{Cor}(X, Y) = -0.069$



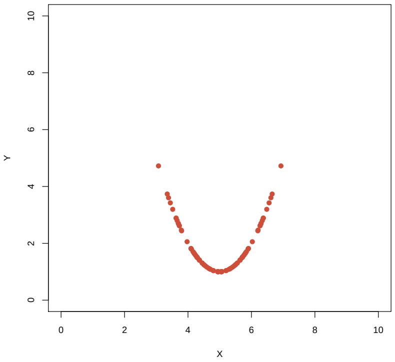
$\sigma_X = 1.14$, $\sigma_Y = 0.78$
$\text{Cov}(X, Y) = 0.78$
$\text{Cor}(X, Y) = 0.71$



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



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



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

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!)
- $\mathrm{Var}(a \cdot X) = a^2 \cdot \mathrm{Var}X$
- $\mathrm{Var}(X+Y) = \mathrm{Var}X + \mathrm{Var}Y + 2 \cdot \mathrm{Cov}(X,Y)$
- $\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \mathrm{Var}\big(X_i\big) + 2 \cdot \sum_{j=1}^n \sum_{i=1}^{j-1} \mathrm{Cov}(X_i, X_j)$
- 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

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

With these rules we can proof:

Theorem 4 *If $X_1, X_2, \dots X_n$ are independent $\mathbb{R}$-valued random variables with expectation value μ and variance σ^2, we get for $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$:*

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

and

$$Var\ \overline{X} = \frac{1}{n}\sigma^2,$$

that is,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

In particular: The standard error $\frac{s}{\sqrt{n}}$ is a estimator for the standard deviation of the $\sigma_{\overline{X}}$ sample mean $\overline{X}$ of $(X_1, X_2, \dots, X_n)$.
The sample standard deviation s is an estimator of the standard deviation σ in the entire population.

Proof: Linearity of the expectation value implies

$$\begin{aligned}\mathbb{E}\overline{X} &= \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big(X_i\big) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mu = \mu.\end{aligned}$$

The independce of X_i helps to simplify the variance:

$$\begin{aligned}\mathrm{Var}\ \overline{X} &= \mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\big(X_i\big) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}\sigma^2\end{aligned}$$

Bernoulli distribution

A Bernoulli distributed random variable Y with success probability $p \in [0,1]$ has expectation 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 5 (Expectation 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

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

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:

$$\mathrm{Var}(X) = \mathrm{Var}(K/n) = \mathrm{Var}(K)/n^2 = n \cdot p \cdot (1 - p)/n^2$$

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

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 is 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

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

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 Applications in Quantitative Genetics

Quantitative Traits

continuous traits: weight, size, growth rate...

discrete traits: number of offspring, bristle number,...

traits with quantitative thresholds: environment and genes determine whether a character is expressed

Quantitative Genetics

- natural selection needs phenotypic variation to operate
- many traits are influenced by few major and many minor genes
- Q.G. has been successfully applied in animal an plant breeding
- application to evolutionary and ecological processes not trivial
- no exact knowledge of genetic mechanisms, rather statistical approach
- QTL analysis to search for genomic regions that influence a trait

Aims for now

- use formulas for Var and Cov to understand how
  - natural variation and
  - correlation of a trait with fitness
  - heritability of the trait

  influence the effect of selection
- based on the theoretical considerations how to predict effect of selection based on data?
- Results will be summarized in
  - Robertson-Price identity
  - breeder's equation

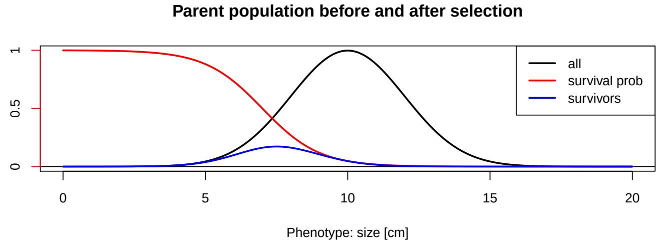
Recommended Books

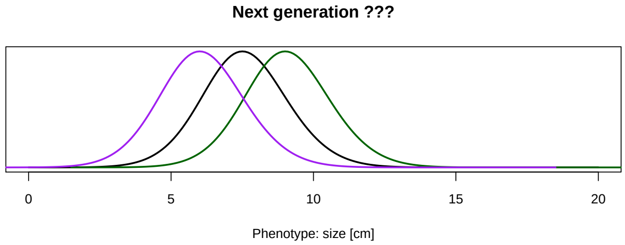
References

[LW98] M. Lynch, B. Walsh (1998) *Genetics and Analysis of Quantitative Traits* Sinauer Associates, Inc., Sunderland, MA, USA

[BB+07] N.H. Barton, D.E.G. Briggs, J.A. Eisen, D.B. Goldstein, N.H. Patel (2007) *Evolution* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA

Selection on quantitative trait

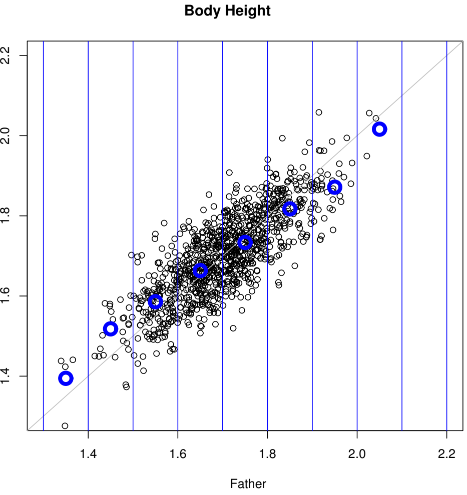

Parent population before and after selection

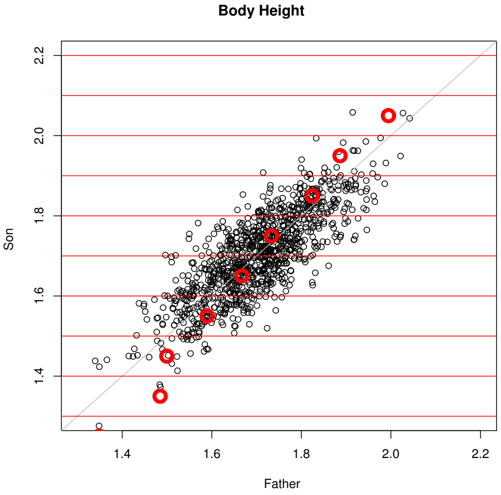

Next generation ???

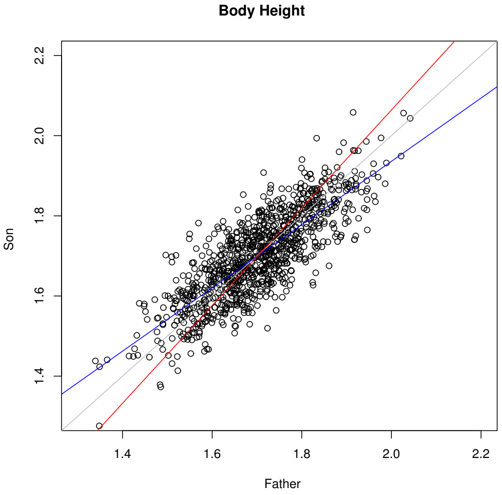
Origin of the word "Regression"

Sir Francis Galton (1822–1911): Regression toward the mean.

Tall fathers tend to have sons that are slightly smaller than the fathers.
Sons of small fathers are on average larger than their fathers.

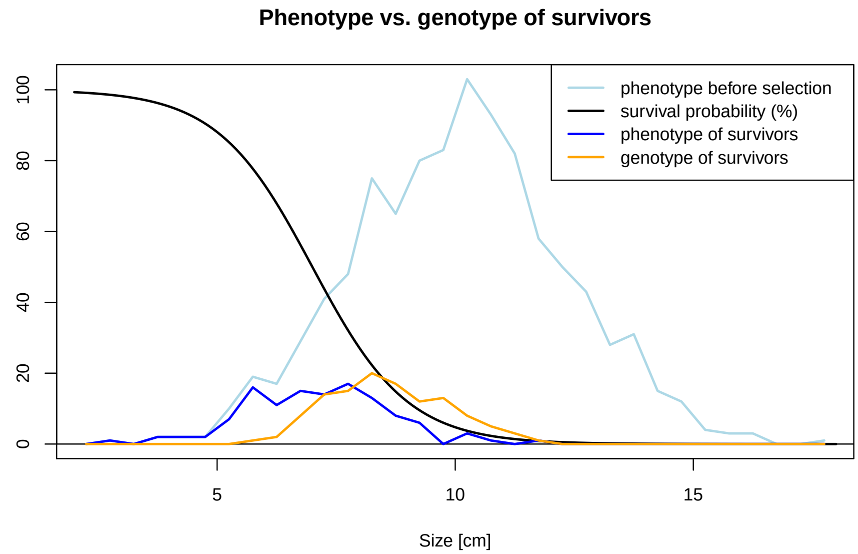

Body Height


Body Height

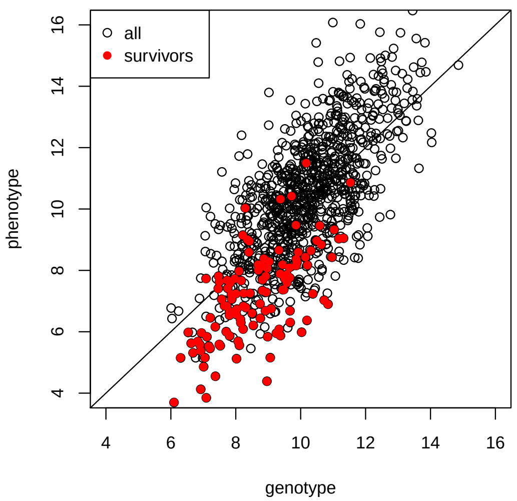

Body Height

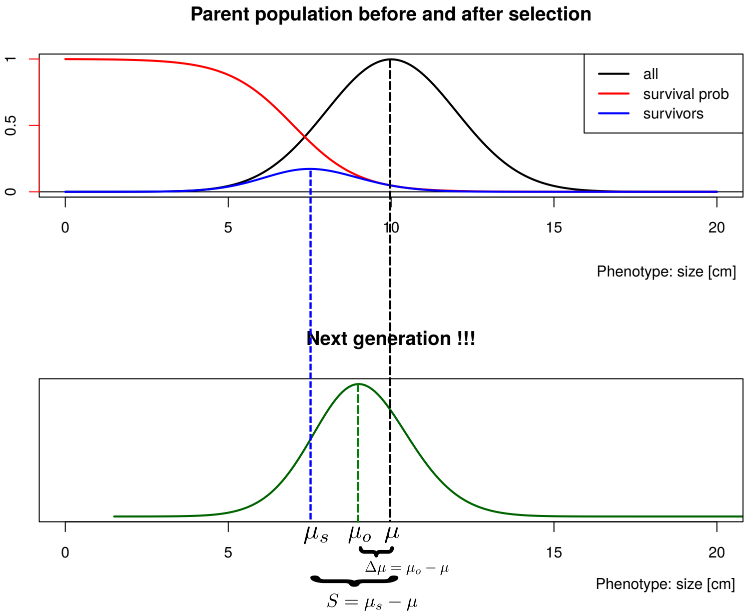
Similar effects

- In sports: The champion of the season will tend to fail the high expectations in the next year.
- In school: If the worst 10% of the students get extra lessons and are not the worst 10% in the next year, then this does not proof that the extra lessons are useful.



```
genotype <- rnorm(1000,10,1.5)
environment <- rnorm(1000,0,1.5)
phenotype <- genotype + environment

hist(phenotype,col="lightblue",breaks=4:36/2)

survival.prob <- function(x) {
  1-1/(1+exp(-x+7))
}

lines(20:180/10,survival.prob(20:180/10)*100,lwd=2)

survivors <- rbinom(1000,size=1,prob=(survival.prob(phenotype)))

hist(phenotype[survivors==1],add=TRUE,col="blue",breaks=4:36/2)
hist(genotype[survivors==1],add=TRUE,col="orange",breaks=4:36/2)
```





Classical estimated of heritabilities after Falconer (1981) *Introduction to quantitative gentics*

Species	*Trait*	*Heritability*
humans	stature	0.65
	serum immonuglobulin	0.45
cattle	body weight	0.65
	milk yield	0.35
poultry	body weight	0.40
	egg production	0.10

We will now derive two equations:

(Robertson-)Price-equation: How selection shifts the mean phenotype (in the same generation)

breeders' equation:

- Predict change from one generation to the next
- Account for selection and heritability
- use a measure of heritability that can be estimated from parent-offspring comparisons

μ mean phenotype before selection

μ_s mean phenotype after selection but before reproduction

$S = \mu_s - \mu$ **directional selection differential**

μ_o mean phenotype in offspring generation

$\Delta\mu = \mu_o - \mu$

$W(z)$ **individual fitness**: probability that individual with phenotype z will survive to reproduce

$p(z)$ density of phenotype z before selection

$\overline{W} = \int W(z) \cdot p(z)\ dz$ mean individual fitness

$w(z) = W(z)/\overline{W}$ relative individual fitness

$p_s(z) = w(z)p(z)$ density of phenotype z after selection but before reproduction (density in a stochastic sense, i.e. integrates to 1)

Let Z be the phenotype of an individual drawn randomly from the parent population before selection.

$$\begin{aligned} \mu &= \mathbb{E}Z \qquad \mathbb{E}(w(Z)) \;=\; 1 \\ \mu_S &= \int_z z \cdot p_S(z)\ dz \;=\; \int_z z \cdot w(z) \cdot p(z)\ dz \;=\; \mathbb{E}(Z \cdot w(Z)) \\ &\Rightarrow S = \mu_s - \mu = \mathbb{E}(Z \cdot w(Z)) - \mathbb{E}(Z) \cdot \mathbb{E}(w(Z)) = \mathrm{Cov}(Z, w(Z)) \end{aligned}$$

Thus, we obtain:

Theorem 6 (Robertson-Price identity; Robertson 1966; Price 1970/72)

$$S = Cov(Z, w(Z))$$

Assume we can partition the phenotypic value Z into a genotypic value G and an environmental (or random) deviation E:

$$Z = G + E$$

Then,

$$\mathrm{Cov}(Z, G) = \mathrm{Cov}(G + E, G) = \mathrm{Var}(G) + \mathrm{Cov}(E, G)$$

and

$$\mathrm{Cor}(Z, G) = \frac{\mathrm{Var}(G) + \mathrm{Cov}(G, E)}{\sigma_G \cdot \sigma_Z}.$$

In the special case of $\mathrm{Cov}(G, E) = 0$, we obtain for the genetic contribution of the phenotypic variance

$$\mathrm{Cor}^2(G, Z) = \frac{\mathrm{Var}(G)}{\mathrm{Var}(Z)}.$$

(Note that if $\mathrm{Cov}(G, E) = 0$, then Var(Z)=Var(G)+Var(E))

Note that if E is really due to environmental effects, $\mathrm{Cov}(G, E)$ may not be 0 if the population is genetically and spatially structured (and for many other possible reasons).

In any case,

$$\frac{\mathrm{Var}(G)}{\mathrm{Var}(Z)} =: H^2$$

is called **heritability in the broad sense**.

Problem: $\mathrm{Var}(G)$ and thus also H^2 are parameters that are hard to estimate.

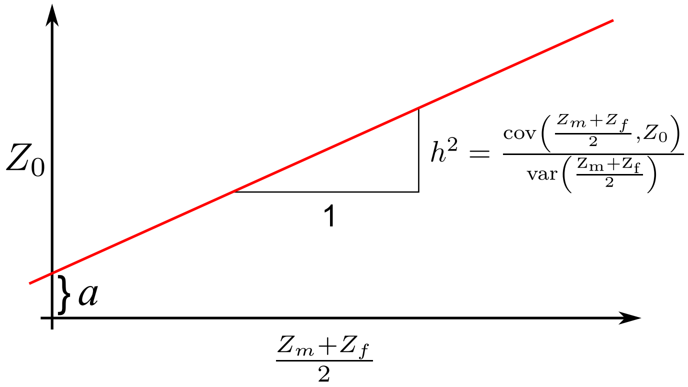
narrow-sense heritability

Let Z_m, Z_f, Z_o be a the phenotype sampled from a triplet of mother, father and an offspring, sampled from the population. The narrow-sense heritability h^2 is defined by

$$h^2 := \frac{\mathrm{Cov}\left(\frac{Z_m+Z_f}{2}, Z_o\right)}{\mathrm{Var}\left(\frac{Z_m+Z_f}{2}\right)}.$$

It is the slope of the regression line to predict Z_0 from the mid-parental phenotype $\frac{Z_m+Z_f}{2}$ and can be estimated from a sample of many parent-offspring triples.

We will see later in this semester: The line that predicts Y from X has slope $\mathrm{Cov}(X,Y)/\mathrm{Var}(X)$.


Z_0
$h^2 = \frac{\mathrm{cov}\left(\frac{Z_m+Z_f}{2}, Z_0\right)}{\mathrm{var}\left(\frac{Z_m+Z_f}{2}\right)}$
1
$\}\, a$
$\frac{Z_m+Z_f}{2}$

If there was selection, then:

$$\mu_o = \mathbb{E}Z_o = \mathbb{E}\left(a + h^2 \cdot \frac{Z_m+Z_f}{2}\right) = a + h^2 \cdot \mathbb{E}\left(\frac{Z_m+Z_f}{2}\right) = a + h^2 \cdot \mu_S$$

If the values $\widetilde{Z}_m, \widetilde{Z}_f, \widetilde{Z}_o$ stem from a population with no selection, we assume that the mean phenotype is the same in the two generations:

$$\mu = \mathbb{E}\widetilde{Z}_o = a + h^2 \cdot \mathbb{E}\left(\frac{\widetilde{Z}_m + \widetilde{Z}_f}{2}\right) = a + h^2 \cdot \mu$$

This implies: $\Delta\mu = \mu_o - \mu = (a + h^2 \cdot \mu_S) - (a + h^2 \cdot \mu) = h^2 \cdot (\mu_S - \mu) = h^2 \cdot S$

Theorem 7 (breeders' equation)

$$\Delta\mu = h^2 S$$

Equivalent definition of h^2

Assume that that Z_m and Z_f are independent and have the same distribution as Z. Then follows

$$\mathrm{Var}\left(\frac{Z_m+Z_f}{2}\right) \;=\; \frac{1}{4}\mathrm{Var}\left(Z_m+Z_f\right) \;=\; \frac{1}{4}\left(\mathrm{Var}\left(Z_m\right)+\mathrm{Var}\left(Z_f\right)\right) \;=\; \frac{1}{2}\mathrm{Var}\left(Z\right),$$

and

$$\mathrm{Cov}\left(\frac{Z_m+Z_f}{2}, Z_0\right) \;=\; \frac{1}{2}\mathrm{Cov}\left(Z_m+Z_f, Z_0\right) \;=\; \frac{\mathrm{Cov}\left(Z_m, Z_0\right)+\mathrm{Cov}\left(Z_f, Z_0\right)}{2}.$$

And thus

$$h^2 = \frac{\mathrm{Cov}\left(\frac{Z_m+Z_f}{2}, Z_0\right)}{\mathrm{Var}\left(\frac{Z_m+Z_f}{2}\right)} = \frac{\mathrm{Cov}\left(Z_m, Z_0\right)+\mathrm{Cov}\left(Z_f, Z_0\right)}{\mathrm{Var}\left(Z\right)}$$

Equivalent definition of h^2 under certain assumptions

Let G_m and G_f be the phenotypic effects of the genes transmitted by the mother and the father to the offspring.

If mating is so random and if there are no correlations (between parental genotypes and environmental effects etc.), we obtain

$$\text{Cov}\left(\frac{Z_m + Z_f}{2}, Z_o\right) = \text{Cov}\left(\frac{G_m + G_f}{2}, G_m + G_f\right) = \frac{\text{Var}G_m + \text{Var}G_f}{2},$$

and thus

$$h^2 = \frac{\text{Cov}\left(\frac{Z_m+Z_f}{2}, Z_o\right)}{\frac{1}{2}\text{Var}\left(Z\right)} = \frac{\text{Var}G_m + \text{Var}G_f}{\text{Var}\left(Z\right)}$$

Example

References

[1] Galen (1996) Rates of floral evolution: adaptation to bumblebee pollination in an alpine wildflower, *Polemonium viscosum Evolution* 50(1): 120–125

- long-term experiment, trait is corolla flare
- S was measured as
  - 7% when estimated from number of seeds
  - 17% when estimated from number of surviving offspring after 6 years
- $h^2 \approx 1$
- Change of trait 9% in one generation

Some of the things you should be able to explain

- Robertson-Price identity
- Breeder's equation
- Why is the narrow-sense heritability and not the broad-sense heritability used in the breeder's equation

7 Application Example about Codon Bias

In

References

[1] E.N. Moriyama (2003) Codon Usage *Encyclopedia of the human genome*, Macmillan Publishers Ltd.

examines 9497 human Genes for Codon Bias.
In these genes the amino acid proline is coded 16710 times by the codon CCT and 18895 times by the codon CCC.
Does it only depend on pure randomness which codon is used?

Then the number X of the CCC would be binomially distributed with $p = \frac{1}{2}$ and $n = 16710 + 18895 = 35605$. Assume the number X $(= 18895)$ of CCC is binomially distributed with $p = \frac{1}{2}$ and $n = 16710 + 18895 = 35605$.

$$\mathbb{E}X = n \cdot p = 17802.5$$
$$\sigma_X = \sqrt{n \cdot p \cdot (1-p)} \approx 94.34$$
$$18895 - 17802.5 = 1092.5 \approx 11.6 \cdot \sigma_X$$

Does this look like purely random?

The question is:

How small is the probability of a deviation from the expectation of at least $\approx 11.6 \cdot \sigma_X$, if it is all purely random?

We have to calculate

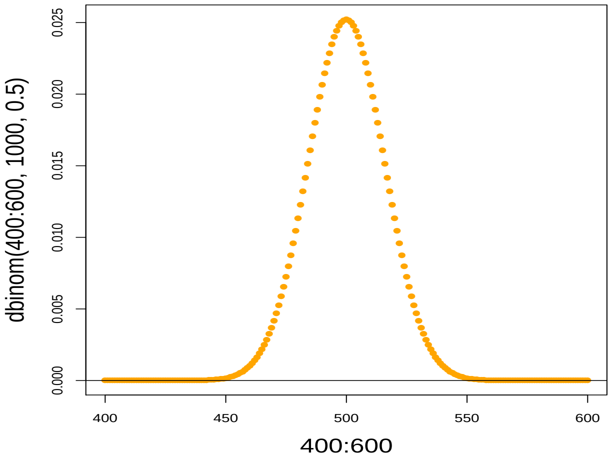
$$\Pr\big(|X - \mathbb{E}X| \geq 11.6\sigma_X\big).$$

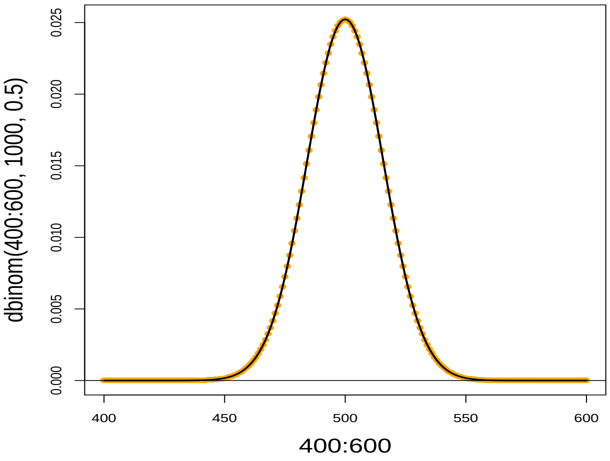
A problem with the binomial distribution is: Calculating $\binom{n}{k}$ precisely is slow for large n. Therefore:

The binomial distribution can be approximated by other distributions.

8 Normal distribution

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

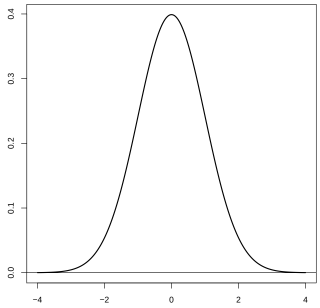




Density of the standard normal distribution

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

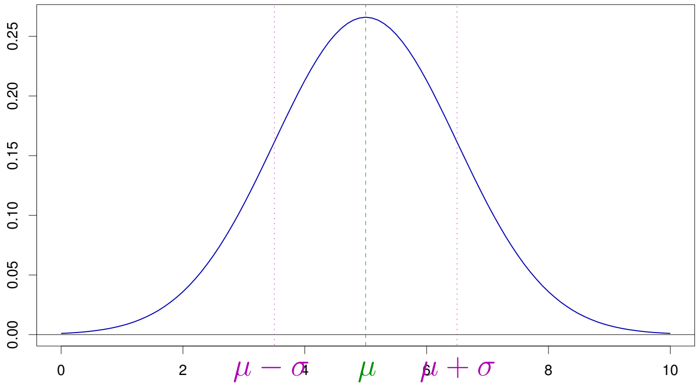
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Always have in mind:

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

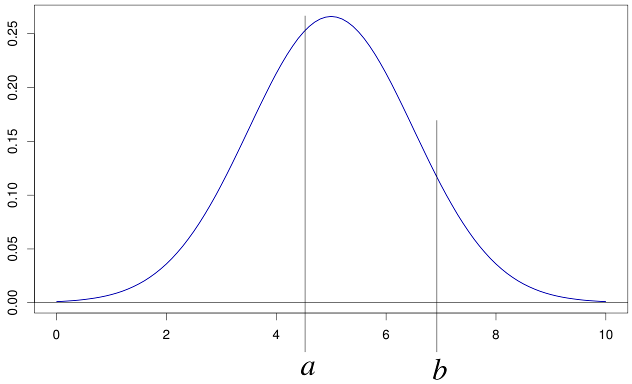
- $\Pr(|Z - \mu| > \sigma) \approx 33\%$
- $\Pr(|Z - \mu| > 1.96 \cdot \sigma) \approx 5\%$
- $\Pr(|Z - \mu| > 3 \cdot \sigma) \approx 0.3\%$

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



Densities need Integrals

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



we get

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

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:

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

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

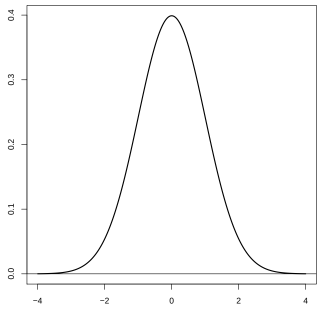
$$\mathrm{Var}(Z) = \mathbb{E}(Z - \mathbb{E}Z)^2 = \mathbb{E}Z^2 - (\mathbb{E}Z)^2$$

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
```

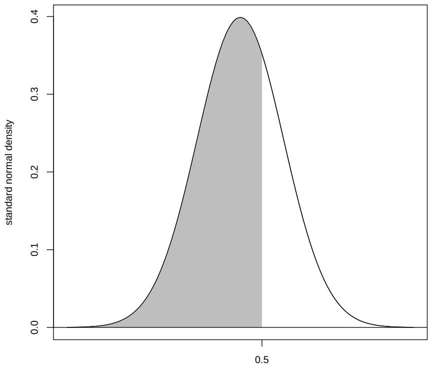
draw a sample of length 6 from standard normal with expectation value 5 and standard deviation 3:

```
> 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
```



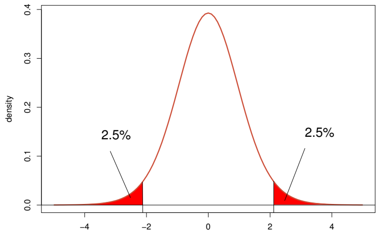
example: Computing probabilities: Let $Z \sim \mathcal{N}(\mu = 5, \sigma^2 = 2.25)$.

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. Fo 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.

9 Normal approximation

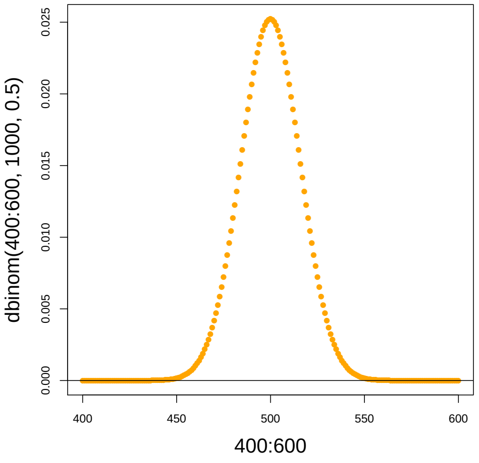
Normal approximation

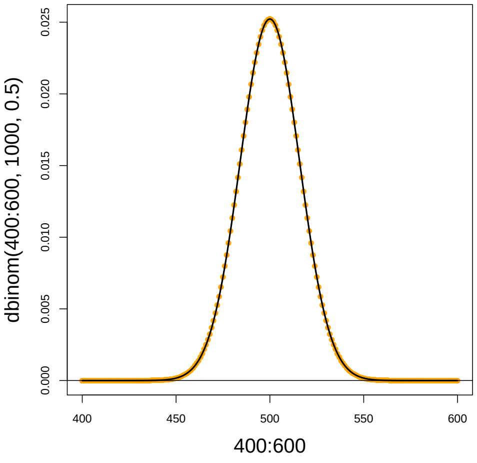
For large n and p which are not too close to 0 or 1, we can approximate the binomial distribution by a normal distribution with the corresponding mean and variance.

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

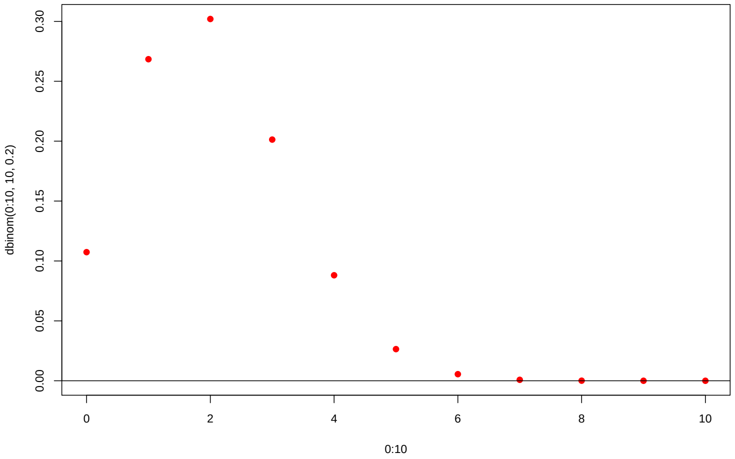
$$\Pr(X \in [a, b]) \approx \Pr(Z \in [a, b])$$

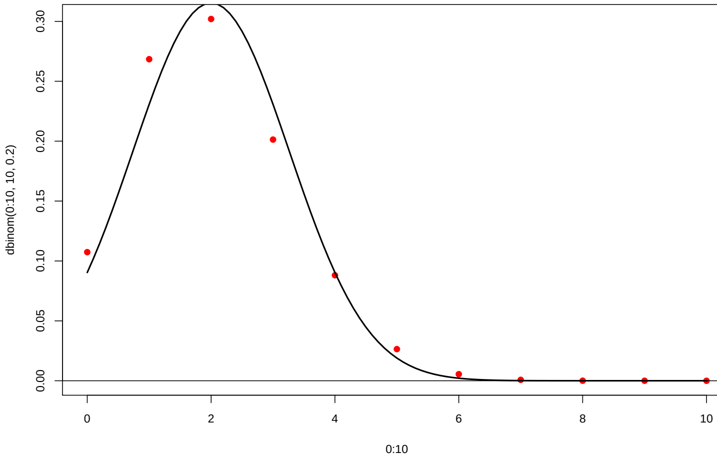
(rule of thumb: Usually okay if $n \cdot p \cdot (1 - p) \geq 9$)
$n = 1000$, $p = 0.5$, $n \cdot p \cdot (1 - p) = 250$





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





Theorem 8 (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:

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

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

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

10 The z-Test

Back to the example with the proline codons in the human genome.

CCT is used $k = 16710$ times CCC is used $n - k = 18895$ times

Do this look like purely random?
We say: No.

Doubters may say: Just random

The Hypothesis
Purely random No difference
is called the Nullhypothesis.
To convince the doubters, we have to find arguments against the null hypothesis. We show: If the null hypothesis is true the observation is very improbable.

CCT is used $k = 16710$ times CCC is used $n - k = 18895$ times Under the null hypothesis “just random” the number X of CCT is $\mathrm{bin}(n, p)$-distributed with $n = 35605$ and $p = 0.5$.

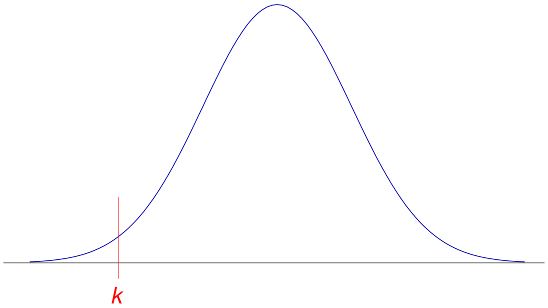
Normal approximation: X is approximately $\mathcal{N}(\mu, \sigma^2)$-distributed with

$$\mu = n \cdot p = 17802.5 \approx 17800$$

and

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)} = 94.34 \approx 95$$

Question: Is it plausible, that a random variable X that has taken the value $k = 18895$ is approximately $\mathcal{N}(17800, 95^2)$-distributed?

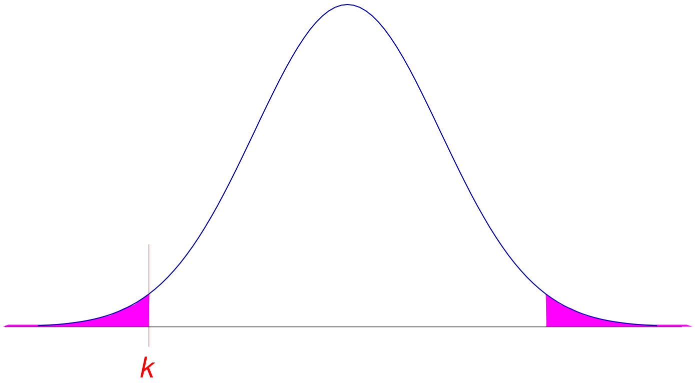


If the null hypothesis H_0 holds, then

$$\Pr(X = 17800) = 0$$

But that does not imply anything, because $\Pr(X = k) = 0$ holds for every value k!

Relevant is the probability (assuming the null hypothesis H_0) that X takes a value at least as extreme as k:



$$\Pr(|X - \mu| \geq |k - \mu|) = \Pr(|X - \mu| \geq 1092.5) \approx \Pr(|X - \mu| \geq 11.6 \cdot \sigma)$$

We have memorized already:

$$Pr(|X - \mu| \geq 3 \cdot \sigma) \approx 0.003$$

This means that $\Pr(|X - \mu| \geq 11.6 \cdot \sigma)$ must be extremely small.
Indeed:

```
> 2 * pnorm(18895,mean=17800,sd=95,lower.tail=FALSE) [1] 9.721555e-31
```

Without normal approximation:

```
> pbinom(16710,size=35605,p=0.5) + + pbinom(18895-1,size=35605,p=0.5,lower.tail=FALSE) [1] 5.329252e-31
```

We can argue that such a deviation from the expectation value could only be explaind by extreme random.

Thus we can reject the null hypothese "just random" and search for alternative explanations, as for example differences in the efficiency of CCC and CCT or in the availability of C and T.

Summary z-Test

Null hypothesis H_0 (what we want to reject): the observed value x comes from a normal distribution with mean μ and **known** Variance σ^2.

p-value $=\Pr(|X - \mu| \geq |x - \mu|)$, where $X \sim \mathcal{N}(\mu, \sigma^2)$, the probability of a deviation that is *at least* as large as the observed one.

Significance level α : usually 0.05. If the p-value is smaller than α, we reject the null hypothesis on the significance level α and look for an alternative explanation.

Limitations of the z-Test

The z-Test can only be applied if the variance of the normal distribution is known or at least assumed to be known according to the null hypothesis.

This is usually not the case when the normal distribution is applied in statistical tests.

Ususally the variance has to be estimated from the data. In this case we have to apply the famous

t-Test

instead of the z-Test.

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
- z-test, H_0, p-value, significance level α

Note also the lists on pages 10, 19 and 26.