CORTEX fellows training course in biostatistics

Neuroscience Center Zurich

University of Zurich

Dr. C.J. Luchsinger

1 Probability

Nice to have read: Chapters 1, 2 and 3 in Stahel or Chapters 1 and 2 in Cartoon Guide

Further readings: Chapter 4 in Stahel or Chapter 3 in Cartoon Guide

In this chaper...

we formalize **Events**, introduce the **P**robability of an Event ($P[A]$) and get to know the manipulations we may make with Events and Probabilities. Then we introduce independence. Independence is in general an important assumption when analysing data.

1.1. What do you associate with "Probability" and/or "Statistics"?

1.4 Little bit of combinatorics

Factorial

$$n! := n(n-1)(n-2)\cdots 2 \cdot 1 \qquad \text{called ”Factorial } n\text{”}$$

By convention: $0! := 1$.

Travelling Salesman Problem: Coming from Europe, have to visit 50 Cities in the U.S. (once each) and return to the first one at the end. How many possibilities?

Binomial coefficient

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!} \qquad \text{called ”}n\text{ choose }k\text{”}$$

In a study with 10 mice (numbered 1 to 10), 6 die and 4 survive (I did *not* say number 1 to 6 and 7 to 10 respectively). How many possibilities are there for such an event (we are not yet looking at the *probability*, only how many possibilities)?

1.5 Probability P

For the next few steps, it is easiest to think about games, gambling (we will return to more serious problems later): Dice / Coin Flipping.

* Ω is called the **sample space**, set (German: Menge) of possible outcomes of an experiment. In each experiment we have precisely one **elementary outcome**: $\omega_1 \in \Omega$ or $\omega_2 \in \Omega$ etc.

* Dice: *we choose* $\Omega := \{1, 2, 3, 4, 5, 6\}$.

* Coin: *we choose* $\Omega := \{h, t\}$.

* Pairs of Dice, red and blue (Monopoly): *we choose* $\Omega := \{(1, 1), (1, 2), \ldots, (6, 5), (6, 6)\}$; the sample space has 36 elements in this experiment.

* Return to single Dice: $A := \{2, 4, 6\}$ set of even numbers, $B := \{4, 5, 6\}$ set of numbers larger than 3. A and B are called **events**. In general: events are subsets of Ω. If the elementary outcome of the experiment "throw a Dice" is 2, then event A occurred: we *have* an even number. Event B did not occur: $2 < 4$. By the way: elementary outcomes are events too: $\{2\} \subset \Omega$.

* Small exercise: Play Monopoly, need 8 to get to "Start". Describe C, the event that we have "8".

We use set theory to formalize events:

Symbol	**set theory / what does it mean for probability theory**
Ω	Set / Sample Space, all possible outcomes of an experiment
ω	Element of Ω / elementary outcome of experiment
A	Subset of Ω / Event; if $\omega \in A$, Event A occurred
A^c	Complement of A / no elementary outcome of A occurred
$A \cap B$	Intersection of A and B / elementary outcome occurred that lies in A **and** B
$A \cup B$	Union of A and B / elementary outcome lies in A **or** B (or in both, A and B).
$A \backslash B$	A but not B / elementary outcome lies in A but not in B
$A \subset B$	A is subset of B / if we have event A, we always have event B too.
ϕ	empty set / impossible event
Ω	entire set / certain event, something must happen

Important examples for

* A^c (if A very complex, A^c might be simple to describe)

* $A \subset B$ (Zurich is Part of Switzerland!)

We now want to introduce the "P", the **probability** of an event. If you've grown up in a "normal" family with "normal" siblings, you have numbers such as 0.5 (Coin), 1/6 (Dice) and 1/36 (Pair of Dice) in mind. But this must not be so. All we ask for is

Definition 1.1 [Probability P] *Probability P is a function from the subsets of Ω into the interval $[0, 1]$. P must satisfy:*

a) $0 \leq P[A] \leq 1$ *for all* A

b) $P[\Omega] = 1,\ P[\phi] = 0$

c) A_1, A_2 *disjoint [German: elementfremd] sets - that is* $P[A_1 \cap A_2] = \phi$*, then:*

$$P[A_1 \cup A_2] = P[A_1] + P[A_2].$$

Dice: $\Omega := \{1, 2, 3, 4, 5, 6\}$. A fair *choice* (this is a model) is $P[\{i\}] = 1/6$ for all $i \in \Omega$.

Let us look at c), and then go back to a) and b).

Returning to your childhood and your big brother: we (or better he) *can* choose $P[\{1\}] = 0.2, P[\{2\}] = 0.2, P[\{3\}] = 0.2, P[\{4\}] = 0.2, P[\{5\}] = 0.1, P[\{6\}] = 0.1$ and there are such dice. But they are not fair.

Probability and proportions: 7.5 million inhabitants in Switzerland. Choose one with equal probability for each of the $7.5M$ inhabitants. Probability that I choose a millionaire(s) is the proportion of people of Switzerland who are millionaires. If proportion of millionaires is 10 %, then the probability to choose a millionaire(s) is 0.1.

Lemma 1.2 [useful stuff about P] *P satisfies*

a) $P[A] = 1 - P[A^c]$ *[in German: "Prinzip der Gegenwahrscheinlichkeit"]*

b) A_1, A_2 *sets, then:*

$$P[A_1 \cup A_2] \leq P[A_1] + P[A_2].$$

c) $A \subset B \Rightarrow P[B] = P[A] + P[B \backslash A]$

d) $A \subset B \Rightarrow P[A] \leq P[B]$

e) $P[A \cup B] = P[A] + P[B] - P[A \cap B].$

Exercises: p. 36-39 in the cartoon guide

1.7 Independent Events

Definition 1.4 [Independent Events] *Events A and B are independent of each other, if*

$$P[A \cap B] = P[A]P[B].$$

Compute probability that first 4 children born are boys:

Modelling of experiments: you must decide yourself, looking at your experiment, whether you want to and can assume, that events are independent of each other. For example, you have 10 mice. If first one dies, does that change the probability that the second dies too? This is not to be mixed up with low survival probability in general. If you give all mice a lethal dosis, they all die. But number 2 does *not* die because number 1 died. They die independently of one an other.

1.9 Exercises (more in course)

1.9.1 Combinatorics:

1.1 Verify:

$$n\binom{n+k}{k} = (k+1)\binom{n+k}{k+1}$$

1.2 Verify:

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

1.3 How many guest are present if pairs of two shake hands 253 times (assume all shake hand with everybody once).

1.4 If you have 30 strait lines, how many points of intersections can you have at the maximum?

1.9.2 Probability:

1.5 Probability for 40 y old man to die in next y is 0.002. 20 men of age 40 meet. How high is the probability that at least one of the 20 dies in the coming y?

1.7 Looking at families with 3 children, how high is the probability that we choose a family with 2 girls and 1 boy if we choose randomly with equal probabilities amongst families with 3 children?

1.8 There are 6 red, 4 green and 2 yellow balls in a box. You take two out of the box randomly. What is the probability that the 2 balls have the same color?

1.9.3 Independent Events:

1.9 Playing Monopoly with pair of dice (red and blue). C is the event that red comes up "1", D is the event that blue comes up "2". Show that C and D are independent of each other.

1.10 Playing Monopoly with pair of dice (red and blue). C is the event that red comes up "1", D is the event the sum is 3. Show that C and D are not independent of each other.

1.11 Given there are 300 students in a class. 200 are studying biology, 80 geography and 20 an other subject. You select a person randomly. How high is the probability that this person is studying biology and is born on a Sunday?

2 Random Variables

Further readings: Chapter 4 in Stahel or Chapter 4 in Cartoon Guide

In this chaper...

we introduce **Random Variables (RV)**: Bernoulli, Binomial, Uniform and Normal distributions are presented. The most difficult part in the whole course is the introduction of continuous RV's. Independence of RV's is strait forward from the independence of Events. The often used assumption of "iid" (**i**ndependent and **i**dentically **d**istributed) is introduced.

2.1 Definition and basic properties

Definition 2.1 [Random Variable (R.V.) X**]** *A Random Variable is a function* $X : \Omega \to \mathbb{R}$.

Examples:

1. Dice: $\Omega = \{1, 2, 3, 4, 5, 6\}$. We want X to be a random variable describing the number that shows up. We choose $X(i) = i$ for all $1 \leq i \leq 6$.

2. Coin tossing: I win 1.- if there is a head (h) and loose 1.- if there is a tail (t). X_1 is my account after the *first* coin tossed: $X_1(h) = 1, X_1(t) = -1$.

2'. Multiple coin tossing: Continue and throw same coin 10 times, independently of earlier outcomes (see 2.4 for a precise definition): X_i is amount gained/lossed in i'th throw.

$$Y := \sum_{i=1}^{10} X_i$$

is my account at time 10. By the way: good choice is

$$\Omega := \{(h,h,h,h,h,h,h,h,h,h),(h,h,h,h,h,h,h,h,h,t),\ldots,(t,t,t,t,t,t,t,t,t,t)\}$$

with $2^{10} = 1024$ elements! Let $\omega_1 := (h,h,h,h,h,h,h,h,h,h)$, then $Y(\omega_1) = X_1(\omega_1) + X_2(\omega_1) + \ldots + X_{10}(\omega_1) = 1 + 1 + \ldots + 1 = 10$.

3. Experiment on mouse: $\Omega := \{S, D\}$ ("survives, dies"). $U_1(S) = 1$ if mouse survives, $U_1(D) = 0$ if mouse dies. U_1 is number of mice survived after first experiment with one mouse.

3'. Experiments on mice: Continue experiments with 5 mice, independently of earlier outcomes (see again 2.4 for a precise definition): U_i is the number of mice that survive in i'th experiment (either 0 or 1).

$$W := \sum_{i=1}^{5} U_i$$

is the number of surviving mice after experiments with all 5 mice.

We now combine this with our first chapter, using the "P". Given the survival probability is 10 %, we are able to compute things like

$$P[U_1 = 1]$$

in example 3; the result is:

Mathematically correct, this is $P[U_1 = 1] := P[\{\omega | U_1(\omega) = 1\}] = P[\{S\}] = 0.1$, because P operates on subsets of Ω. But we usually do not need this.

Let us look at a simply more complicated example. Again, given the survival probability is 10 %, how do we compute

$$P[W = 4]$$

in example 3'; the result is: 0.00045; it is

$$\binom{5}{4} 0.1^4 0.9^1.$$

What the hell is this?

Which distributions are known (more in chapter 4)? (Name, Probabilities or Density Functions, what is it used for?)

discrete:

* Bernoulli:

* Binomial:

* Uniform (discrete):

continuous:

* Uniform (continuous):

* Normal:

2.2 Cumulative Distribution Function - useful technicality

Definition 2.2 [Cumulative Distribution Function F**]** *The Cumulative Distribution Function* F *of a random variable* X *is defined by*

$$F(a) := P[X \leq a] := P[\{\omega | X(\omega) \leq a\}].$$

Often written F_X *instead of* F *for better identification.*

Useful for: reading statistical tables, deriving densities, various computations.

a) Bernoulli Be(p) & Bin(n, p)

b) Try: Uniform on $[-1, 0.5]$

c) and N(0, 1); Normal with mean 0 and variance 1 (we do not yet know "mean" and "variance").

Obvious: $\lim_{a \to -\infty} F(a) = 0$ und $\lim_{a \to \infty} F(a) = 1$; $F(a)$ increases monotonously as a increases.

2.3 Discrete and continuous Random Variables

Don't panic if you don't know what an Integral is. We won't use it a lot and if we do use it, then as a guided tour - and I am your tour guide.

Definition 2.3 [Discrete and continuous Random Variables] *If the outcomes of a random variable X are isolated (discrete) points in $\mathbb{R}$, we call X discrete (math. exact definition is "countable set"). Random variable Y is said to be continuous, if F_Y can be written as an integral in the form*

$$F_Y(a) := P[Y \leq a] = \int_{-\infty}^{a} f(u)du,$$

with f a nonnegative function on $\mathbb{R}$ such that

$$\int_{-\infty}^{\infty} f(u)du = 1.$$

We call f density or density function of Y (sometimes denoted as f_Y for better identification). En bref: discrete is isolated points and continuous is on entire intervals.

Continuous Random Variables satisfy (see Normal Distribution as an illustration):

1. For $a < b$ we have:

$$P[a < X \leq b] = P[X \leq b] - P[X \leq a] = F(b) - F(a) = \int_a^b f(x)dx.$$

2. $P[X = x_0] = \int_{x_0}^{x_0} f(x)dx = 0.$

3. Because of remark 2 we have for $a < b$:

$$\begin{aligned} P[a \leq X \leq b] &= P[a < X \leq b] = P[a < X < b] = P[a \leq X < b] \\ &= P[X \leq b] - P[X \leq a] = F(b) - F(a) = \int_a^b f(x)dx. \end{aligned}$$

2.4 Independence of Random Variables

We use the concept of independence of events to define independence of random variables:

Definition 2.4 [Independence of Random Variables] *Random variables X_1, X_2 are independent of each other, if*

$$P[X_1 \in B_1, X_2 \in B_2] = P[X_1 \in B_1]P[X_2 \in B_2]$$

for all subsets B_1, B_2 of $\mathbb{R}$ (mathematically not quite correct, but OK for this course).

For us: **Independence just means that X_1 and X_2 are generated by independent mechanisms.** Jargon: independent and identically distributed is abbreviated: **iid**

2.5 Exercises (more in course)

2.1 Assume probability that a person has blood group 0+ is 36 %. Compute the probability that among 8 randomly selected people less than 4 have blood group 0+.

2.2 Probability of getting a boy is 51.3 %. Look at families with 4 children. How high is the probability that they have exactly 2 boys (and therefore also exactly 2 girls)?

2.3 Let Z be a $U[5, 9]$-R.V.. Find $P[Z \in [8, 8.7]]$ and $P[Z^2 \in [30, 35]]$.

3 Expectations (Measures for Location (Mean) and Scale (Variance))

Further readings: Chapters 5 and 6 in Stahel or Chapter 4 in Cartoon Guide

In this chaper...

we introduce **Expectation** (other words are mean, average) (E[X]) and **Variance** (V[X]) of a RV.

3.1 Expectation and Variance of discrete and continuous random variables

If we have data $x_1, x_2, \ldots, x_n$ (more on data later on in this chapter), we can define a so called **sample mean**:

$$\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{3.1}$$

It is well possible, that some of the x_i have equal values. Just think of example 2' in chapter 2 with multiple coin tossing: the only possible values are -1 and 1. So we could try to rewrite (3.1) by summing over all possible values of x and defining n_x to be the number of data points with value x. We then have

$$\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{\text{all } x} x n_x = \sum_{\text{all } x} x\frac{n_x}{n}. \tag{3.2}$$

Why did we do that? Because the

$$\frac{n_x}{n}$$

is the relative frequency: how often did we have x in n data points, divided by n. We will see in chapter 5 that this converges to the true probability for such a value $P[X = x]$ as $n \to \infty$. This allows us to define:

Definition 3.1 [Expectation (mean) of discrete and continuous random variables] *Expectation $E[X]$ of discrete random variable X is defined as*

$$E[X] := \sum_x xP[X = x].$$

We sum over all possible values of X. By analogy: Expectation $E[Y]$ of continuous random variable Y is defined as

$$E[Y] := \int_{-\infty}^{\infty} u f_Y(u) du.$$

$f_Y(u)$ is the density function.

Examples I (Mean as "Center of Gravity", in German: "Schwerpunkt"):

1. "Reality differs from Expectations!"; see following example: Be(p), $p \in (0, 1)$ and $P[X = 1] = p, P[X = 0] = 1 - p$: $E[X] =$?

We never have $X = E[X]$, because X is either 0 or 1.

2. Compute expected number that shows up with a fair dice. Think before hand what the result should be.

3. Compute the Expectation of a $U[-2, 1]$-Random Variable. Think before hand what the result should be.

Compute the expected value of a Binomial random variable with $n = 3$ and $p = 0.5$. What is the result for general n and p?

Expectation of Bin(100, 0.6), typical values (preparation for chapter 7)?

Definition 3.2 [Variance of discrete and continuous random variables] $\mu_X := E[X]$, *then we define variance* $V[X]$ *of a discrete random variable* X *as*

$$V[X] := E[(X - \mu_X)^2] = \sum_x (x - \mu_X)^2 P[X = x].$$

Sum is over all possible values of X*. By analogy:* $\mu_Y := E[Y]$*, then we define variance* $V[Y]$ *of a continuous random variable* Y *as*

$$V[Y] := E[(Y - \mu_Y)^2] = \int_{-\infty}^{\infty} (y - \mu_Y)^2 f_Y(y) dy$$

where $f_Y(y)$ *is the density function of* Y*. Standard deviation is (for both, discrete and continuous):*

$$sd[X] := \sqrt{V[X]}.$$

Remark to 3.2: Variance and Standard Deviation are *two possible* measures for the deviation around the mean. Other choice is

$$E[|X - \mu_X|].$$

This is the mean ("E") absolute ("$|$") deviation ("$X - \mu_X$")

Examples II

5. Compute Variance of a Be(p), $p \in (0, 1)$:

6. Compute Variance of a $U[0, 1]$:

3.5 Exercises (more in course)

3.1 Suppose that in 60 items delivered there are on average 1 out of 10 with defects. You now take a sample of 4 and count the number of defects in that sample. Compute expectation and variance of the number of defects in your sample of 4.

3.2 You throw a dice until you have a 6. If this does not happen within 4 times, you stop. Let W be the number of times you throw until you stop (either because you have a 6 or because you did it 4 times). Give the distribution of W in a table and compute the expectation.

3.3 Draw the graphs of the density function and the cumulative distribution functions in the following cases: $\mathcal{N}(0,1), \mathcal{N}(2,1), \mathcal{N}(0,2), \mathcal{N}(-5,1), \mathcal{N}(10,0.1)$.

3.4 Find a) $E[X]$ and b) $E[X^2]$ for X having distribution as follows:

$P[X=8]=1/8; P[X=12]=1/6; P[X=16]=3/8; P[X=20]=1/4; P[X=24]=1/12.$

4 Selected Probability Distributions - all in 1 chapter so that you find them again

* discrete: Bernoulli, Binomial
* continuous: Uniform, Normal
* possible values, Distribution (Probability function / Density), E, V

4.1 Discrete

4.1.1 Bernoulli Be(p)

Two possible values 0 and 1 (alternatively -1 and $+1$). $P[X=1]=p$ (Success) and $P[X=0]=1-p$ (Failure). $E[X]=p$ and $V[X]=p(1-p)$.

$$P[X=x]=p^x(1-p)^{1-x}$$

4.1.2 Binomial Bin(n,p)

X_i, $1 \leq i \leq n$, n iid Be(p)-R.V. $Y := \sum_{i=1}^{n} X_i$. Y so called Binomial RV with parameters n and p; Bin(n,p). $E[Y] = np$ and $V[Y] = np(1-p)$. Possible values: natural numbers $0 \leq y \leq n$:

$$P[Y = y] = \binom{n}{y} p^y (1-p)^{n-y}.$$

"How many successes in n trials?"

4.2 Continuous

4.2.1 Uniform $U[a, b]$

U is distributed uniformly on $[a, b]$, if U has following density function:

$$f(u) = (b-a)^{-1},$$

$a \leq u \leq b$. For $u \notin [a, b]$ density is 0. $E[U] = (a+b)/2$ and $V[U] = (b-a)^2/12$.

4.2.2 Normal N(μ, σ^2)

Very important. Density:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

$E[X] = \mu$ and $V[X] = \sigma^2$.

When constructing statistical tests (Chapter 7), we often have to sum (normal) random variables:

$$X_1 + X_2 + \ldots + X_n.$$

We then need to know the expectation and variance of such sums and mostly even their distribution. The first 2 results to follow will help us find them.

DANGER: FOLLOWING RESULTS DO *NOT* HOLD FOR MANY DISTRIBUTIONS - BUT FOR THE NORMAL DISTRIBUTION IT'S OK!

1. Sum of 2 independent and normally distributed RV's is again normally distributed. Let X be $\mathcal{N}(\mu_1, \sigma_1^2)$ and Y be $\mathcal{N}(\mu_2, \sigma_2^2)$, X independent of Y. Then $X + Y$ has a

$$\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

distribution.

2. "Z-Transform": If X has a $\mathcal{N}(\mu, \sigma^2)$ distribution, then

$$\frac{X - \mu}{\sigma}$$

has a $\mathcal{N}(0, 1)$-distribution (which can be found in tables in books and libraries of statistical packages). $\mathcal{N}(0, 1)$ is called "standard" normal.

Compute $P[X > 13]$ when X is a $\mathcal{N}(10, 4)$ using tables. See 4.3 for further exercises.

3. approximate results: for any σ: let X be $\mathcal{N}(0, \sigma^2)$ RV, then

$$P[|X| > \sigma] \doteq 33\%$$

and

$$P[|X| > 2\sigma] \doteq 5\%.$$

Generally: Normal distribution has 66 % of it's probability within 1 standard deviation away from the mean and even 95 % within 2 standard deviations away from the mean:

4.3 Exercises (more in course)

4.1 X a $\mathcal{N}(4, 25)$-RV. Find $P[-1 < X < 5]$.

4.2 X a $\mathcal{N}(20, 16)$-RV. Find $P[X > 16]$.

4.4 Length of some animal is modelled with a normal distribution with $\mu = 3$ mm and $\sigma^2 = 2$. You look at 240 animals. How many of them do you expect to be less than 2.5 mm in length - how many between 2.8 and 3.1 mm?

7 Test theory

Further readings: Chapter 8, 10, 11 & 12 in Stahel or Chapter 8 in Cartoon Guide

Some remarks: There are very many tests in various situations and a Crash Course in Statistics can not cover all of them - especially not in depth. I suggest further readings especially in Stahel where many more tests are presented and the typical question "when should I use which test" is treated on several occasions more deeply. I further strongly suggest that you use the consulting services offered by Universities (usually free of charge) with large expertise. Having followed this course you can follow their suggestions much better.

7.3 Notation

There are 2 *possible* hypothesis (fair coin vs unfair coin). We are going to call them Null Hypothesis $\mathcal{H}_0$ and Alternative Hypothesis $\mathcal{H}_1$. *Either* $\mathcal{H}_0$ *or* $\mathcal{H}_1$ is true, but we don't know which is true. Then we get data (how many heads in 100 coin tossing). Depending on which hypothesis is true, the random variable that generated this data will have a different distribution (for example a $\text{Bin}(100, 0.5)$ vs $\text{Bin}(100, p)$, where $p \neq 0.5$). We then have to make a decision: say whether we are in $\mathcal{H}_0$ or $\mathcal{H}_1$. Before we get that data, we have some time to discuss a strategy. All we have is the following situation:

We have to either say: "we are in situation $\mathcal{H}_0$" or "we are in situation $\mathcal{H}_1$". 4 things can happen:

1. We are in fact in $\mathcal{H}_0$ and say: "we are in $\mathcal{H}_0$" [NO ERROR]
2. We are in fact in $\mathcal{H}_0$ and say: "we are in $\mathcal{H}_1$" [TYPE 1 ERROR]
3. We are in fact in $\mathcal{H}_1$ and say: "we are in $\mathcal{H}_1$" [NO ERROR]
4. We are in fact in $\mathcal{H}_1$ and say: "we are in $\mathcal{H}_0$" [TYPE 2 ERROR]

In German: Fehler 1. bzw. 2. Art

Examples:

* $\mathcal{H}_0$: fair dice (1/6) vs $\mathcal{H}_1$: at least two (!) probabilities are not 1/6.
* $\mathcal{H}_0$: fair coin (1/2) vs $\mathcal{H}_1$: probability not equal to 1/2.
* $\mathcal{H}_0$: drug does not change blood pressure vs $\mathcal{H}_1$: drug does lead to lower blood pressure

Critical Region: where we reject the Null Hypothesis $\mathcal{H}_0$; **critical values** are at the edge. [In German: **Ablehnungsbereich**, wir lehnen dort die $\mathcal{H}_0$-Hypothese ab, Grenzen des Ablehnungsbereichs sind die **kritischen Werte**].

Probability for type 1 error is $\alpha \in [0, 1]$; usually 10, 5, 2.5, 1, 0.1, 0.01, 0.001 %, also called size of test [in German: Grösse des Tests, Risiko 1. Art]

Joke for German speaking people: What if you make a type 1 error? "Wir hatten aber ein α-Pech".

Probability for type 2 error is $\beta \in [0, 1]$. It is usually not pre-specified (in German: vorgegeben) as the α above. We want it to be small (see power just below too).

Power is $(1 - \beta)$ (in German: Macht). We want the power to be large: "The hypothesis testing design problem is: Choose a test to maximize power subject to a pre-specified size." (in German: Wir suchen zu einem vorgegebenen α einen Test, welcher die Macht maximiert - und damit das Risiko 2. Art minimiert).

Time to solve exercise 7.1.

Test Statistic: function of the data $x_1, \ldots, x_n$ which we use to solve our test problem: for example $\overline{x}$ or $\sum_i x_i$.

P-Value: Under $\mathcal{H}_0$ (if $\mathcal{H}_0$ is true) and we have data: what is the probability of observing a test statistic **at least as extreme (or even more extreme)** as with our data: Example with $n = 1$:

Why $\alpha = 0.1$ or $\alpha = 0.05$? Why not $\alpha = 0.5$ or better: $\alpha = \beta$ (would be fair)? $\mathcal{H}_0$ is usually the common, current knowledge in research or a careful assumption:

* "In dubio pro reo"
* "Old drug is better" (we know side effects of the old drug)
* "Drug does not alter blood pressure" (pharmaceutical company must prove they are better and not the FDA)

We are therefore very conservative such that we do not want results which are simply a result of pure chance and not of a systematic effect to become common opinion in science!

7.4 General Recipe

We illustrate it with following situation: We know (or assume) data $x_1, x_2, \ldots, x_9$ comes from a $\mathcal{N}(0,1)$ or a $\mathcal{N}(2,1)$ RV. We don't know whether the mean is 0 or 2.

Steps in Hypothesis Testing	**Example**
Set up both hypothesis	$\mathcal{H}_0 : \mu = 0$ vs $\mathcal{H}_1 : \mu = 2$
Set α and (if possible) n	$\alpha = 0.05, n = 9$
Choose a good test statistic	$\frac{1}{9}\sum_{i=1}^{9} X_i\ (= \overline{X})$
Find distribution of test statistic under $\mathcal{H}_0$	$\mathcal{N}(0, 1/9)$
Critical value	$1.64/3 \doteq 0.547$
Get data	$x_1, \ldots, x_9$
Reject or accept $\mathcal{H}_0$	$\overline{x} < 0.547$: accept $\mathcal{H}_0$, otherwise reject $\mathcal{H}_0$

Alternatively to above recipe (from Step 5 onwards): Compute P-Value and compare with α.

7.5 Some classical (parametrical) tests

7.5.2: $x_1, \ldots, x_n$ from $\mathcal{N}(\mu, \sigma^2)$ with σ^2 known: is μ equal to some μ_0 (for example equal to 0)? ["z"-TEST]

Example: $x_1, \ldots, x_n$ being length of nerves that have grown under some particular circumstances. Generally known that these can be modelled with a $\mathcal{N}(10, 4)$-RV. We slightly changed setting and are now interested, whether mean is larger or not:

7.5.3: $x_1, \ldots, x_n$ from $\mathcal{N}(\mu, \sigma^2)$ with σ^2 unknown: is μ equal to some μ_0 (for example equal to 0)? [1 SAMPLE T-TEST]

Motivation/Example same as in 7.5.2, but realistically you don't know the variance.

SPSS: Analyze > Compare Means > One-Sample T Test

Solve Exercise 7.2 now.

7.9 Interpretation of Results - famous mistakes

Warning: Often heard, stays wrong never the less: "$\mathcal{H}_0$ is true with 95 % probability!"
True view of problem: *Either* $\mathcal{H}_0$ *or* $\mathcal{H}_1$ is true, 0 or 100 %. We never know (exceptions are trivial cases) which hypothesis is true. But we allow us to make a type 1 error in 5 % of the cases where $\mathcal{H}_0$ is true. In other words: given $\mathcal{H}_0$ is true, the test statistic will lie out of the critical region with 95 % probability. If not we reject $\mathcal{H}_0$ - although it is true.

See introductory example: **Wrong:** "Probability for 58 in a Bin(100, 0.5) is so small that we reject $\mathcal{H}_0$." **True view of problem:** Probability for 50 (the mean!) in a Bin(100, 0.5) is not much larger... . You must take the **whole tale of the distribution** (one or two-sided). You look at the **probability for such an extreme value or a value which is even more extreme.**

If you reject the $\mathcal{H}_0$, it is possible that $\mathcal{H}_1$ is true. But it is also possible that you reject $\mathcal{H}_0$ by chance ("α-Pech"), although it is true. You can make α smaller to make this more unlikely.

It is also possible that assumptions you made are wrong. For example data may not be normally distributed and you use a t-test with small sample.

To end this part, some correct statements:

The test statistic gave a value which is in the critical region. We therefore reject $\mathcal{H}_0$.

The test statistic gave a value which is outside of the critical region. We therefore accept $\mathcal{H}_0$.

P-Value (Probability for such an event or an even more extreme event under $\mathcal{H}_0$) is smaller than α. We therefore reject $\mathcal{H}_0$.

P-Value is larger than α. We therefore accept $\mathcal{H}_0$.

P-Value is 3.4 %. This means that if $\mathcal{H}_0$ is true, such values as we observe - or even more extreme values - only happen with probability 0.034. If we have an α larger than 0.034, we reject $\mathcal{H}_0$, if we have an α smaller than 0.034, we accept $\mathcal{H}_0$.

7.10 Exercises

7.1 Let $\mathcal{H}_0$ be $\mathcal{N}(0,1)$. We have $n = 1$ to keep the maths easy. Choose $\alpha = 0.05$ and test against

a) $\mathcal{H}_1$ being $\mathcal{N}(1,1)$ and compute the β too.

b) $\mathcal{H}_1$ being $\mathcal{N}(2,1)$ and compute the β too.

c) $\mathcal{H}_1$ being $\mathcal{N}(3,1)$ and compute the β too.

d) $\mathcal{H}_1$ being $\mathcal{N}(4,1)$ and compute the β too.

e) Summarize results from a)-d). Does it make sense?

Obviously: 1.64 and 1.96 are very important numbers for statisticians!

7.2 Medical treatment: You have 51 patients, measure blood pressure before treatment and after treatment. Data at hand is difference: $x_1, \ldots, x_{51}$. Pharmaceutical company claims, blood pressure is lower with treatment than without. σ has been estimated to be 8.4; $\overline{x} = -2.3$. Make a statistical test.

7.5 Tabletts are weighted. We got the following weight in grams:

$$1.19, 1.23, 1.18, 1.21, 1.27, 1.17, 1.15, 11.4.$$

a) Test, whether the average weight is 1.2 g (two-sided) at 5 %

b) Test, whether the average weight is less than 1.2 g (one-sided) at 5 %

Give precisely the two hypothesis.