Crash Course in Statistics for Neuroscience Center Zurich University of Zurich

Dr. C.J. Luchsinger

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 (more to come in Chapter 4). 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 $1 \le i \le 6$.

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 for people who love Maths: 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 slightly 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 $a < b$:

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

4. If F is differentiable:

$$F'(x) = f(x).$$

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). Again, we can write this as*

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

which can be interpreted easier.

For us: **Independence just means that X_1 and X_2 are generated by independent mechanisms.** For example, the number that shows up in the first dice is independent of the number in the second dice: we are not going to compute all possible sets B_1, B_2, but just some to show what that means:

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 $P[Z^2 \in [30, 35]]$.

2.4 [needs calculus] Distributions must some up to one (or the integral of the density must be 1). Verify if the following is a density:

$$f(x) = 2x$$

for $x \in [0, 1]$ and 0 otherwise.

2.5 [needs calculus] Compute $F(a)$ for RV from 2.4.

2.6 Compute $P[X \in [0.5, 0.75]]$ for RV from 2.4. Use 2.5.