Crash Course in Statistics for Neuroscience Center Zurich University of Zurich

Dr. C.J. Luchsinger

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. We get to know the rules associated with E and V. At the end we look at the **correlation** between two RV's.

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.

4. Expectation of $\mathcal{N}(\mu, \sigma^2)$-Random Variable. Density is:

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

We therefore have:

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{(x-\mu)+\mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{(x-\mu)}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx + \int_{-\infty}^{\infty} \frac{\mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} \frac{(x-\mu)}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx + \mu \\
&= 0 + \mu \\
&= \mu
\end{aligned}$$

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 $U[0,1]$:

7. What do you expect the variance of a $U[0,3]$ to be (To be continued)?

8. Variance of $\mathcal{N}(\mu, \sigma^2)$:

$$\begin{aligned}
V[X] := E[(X-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2} dy \\
&= \int_{-\infty}^{\infty} y(y \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2}) dy \\
&= -y \frac{\sigma^2}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2} |_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{\sigma^2}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}y^2} dy \\
&= \sigma^2
\end{aligned}$$

3.2 Important rules for E and V

When constructing Confidence Intervals (CI's) and statistical tests (Chapters 6 and 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 following 3 Lemmas will help us find them. Please read pages 68-72 in the Cartoon Guide if you want some more explanations/interpretations for these rules.

Lemma 3.3 [Rules for E]

a) $E[aX] = aE[X]$ *[in German: "Konstante herausziehen"]*

b) $E[X + Y] = E[X] + E[Y]$ *[in German: "Summen auseinanderziehen"]*

c) a) and b) combined: $E[aX + bY] = aE[X] + bE[Y]$ *(*$\Rightarrow E[b] = b$ *and* $E[0] = 0$*)*

From Example 1 we know that the expectation of a Be(p) is p. But Bin(n, p) is just a sum of n Be(p). Because of Rule b) in Lemma 3.3 the expectation of a Bin(n, p) must therefore be np.

Lemma 3.4 [Rule for V] $V[aX + b] = a^2 V[X]$ *[in German: "Konstante quadratisch raus"]*

Remarks to Lemma 3.4: 1. In Example 7 we asked: What do you expect the variance of a $U[0, 3]$ to be? Answer the question using Lemma 3.4.

2. For $a \in \mathbb{R}$ we have:

$$sd[aX] = |a| sd[X];$$

[in German: "Konstante herausziehen, aber mit Absolutbetrag!"].

Lemma 3.5 [Independence of Random Variables and V] *Let* $(X_i)_{i=1}^n$ *be a sequence of independent random variables (they do not need to have the same distribution). Then:*

$$V[\sum_{i=1}^n X_i] = \sum_{i=1}^n V[X_i]$$

[in German: "Varianz der Summe ist Summe der Varianzen"].

Remark to Lemma 3.5: Example 5 shows that Variance of Be(p) is $p(1-p)$. How can we now compute the variance of a $Bin(n,p)$?

Important preparation for Chapters 5, 6, and 7: Let $X_1, \ldots, X_n$ be a sequence of iid RV's,

$$\bar{X} := \frac{1}{n} \sum_{k=1}^n X_k$$

and let $\mu := E[X_1], \sigma^2 := V[X_1]$. Compute $E[\bar{X}], V[\bar{X}]$ and $sd[\bar{X}]$.

3.3 Intermezzo: Data and empirical mean/variance

* So far we only did probability theory.
* Elementary probability theory has no model risk. We start with phrases like: "let X be a $\mathcal{N}(0,1)$ random variable". You never have to worry and think: "but what, if that is not true?" - It's probability **THEORY**.
* Statistics: **All we have is data:** $d = (x_1, x_2, x_3, \dots, x_n)$**!!!**. We do not know, what the "true" distribution is.

So what do we do? Look at data. (See plot and output file for this chapter)

* sort data
* smallest, largest value, both together, difference between smallest and largest value (so called range)
* Median, what is that?
* Histogram
* arithmetic mean $\overline{x}$, empirical variance s^2 (what's that?)

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

and

$$s^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

show that

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0.$$

Looking at the definition of $\overline{x}$, it is obvious, that this is simply a mean as defined in Definition 3.1 if we give probability $1/n$ to each of the n data points. This is called the **empirical distribution**.

* Please don't worry about dividing by n or $(n-1)$ in the definition of s^2. We will treat this question in chapter 6. It's not so important - the holy grail was lost somewhere else!

* Let's summarize how we treat data from the standpoint of probability theory:

1. We assign probability $1/n$ to each data point

2. We assume that data points were generated independently according to Definition 2.4, if we have no explicit reason not to do so.

3. We assume that these data points $(x_1, x_2, \ldots, x_n)$ are actually

$$(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) = (x_1, x_2, \ldots, x_n),$$

with ω some elementary outcome.

We then must think about what we want to assume about the X's (choose a model).

3.4 Covariance and Correlation

Definition 3.6 [uncorrelated, covariance, coefficient of correlation] X, Y *random variables (technical: satisfying* $E[X^2] < \infty, E[Y^2] < \infty$*).*

a) $Cov(X, Y) := E[(X - E[X])(Y - E[Y])]$ *is called covariance of* X *and* Y*.*

b) If $V[X] > 0, V[Y] > 0$*, we define*

$$Cor(X, Y) := \frac{Cov(X, Y)}{\sqrt{V[X]V[Y]}}$$

to be the coefficient of correlation of X *and* Y*.*

Uncorrelated means $Cov(X, Y) = Cor(X, Y) = 0$. $Cor(X, Y) > 0$ means X and Y are positively correlated and $Cor(X, Y) < 0$ means that X and Y are negatively correlated.

Independence implies Uncorrelated, inverse result does *not* hold!

Lemma 3.8 [Coefficient of Correlation is a measure for *linear* association; better in German: Gleichläufigkeit] *X, Y random variables (satisfying $E[X^2] < \infty, E[Y^2] < \infty$). Then:*

a) $|Cor(X,Y)| \leq 1$

b) $|Cor(X,Y)| = 1$ *if and only if all data points* $(X(\omega), Y(\omega)), \omega \in \Omega$, *lie on 1 line.*

Examples of Plots and their Correlation Coefficients:

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:

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

3.5 [needs calculus] R.V. X with density $2x$ on interval $[0,1]$ (0 otherwise). Compute $E[X]$ and $V[X]$.

3.6 $(X_i)_{i=1}^n$ sequence of iid R.V. with distribution $P[X = 1] = P[X = -1] = 0.5$ (we call this a Bernoulli-R.V. too). Define

$$S_n := \sum_{i=1}^n X_i.$$

a) Compute $E[X_1]$ and $V[X_1]$.

b) Compute $E[S_n]$ and $V[S_n]$.

c) Find function $g(n)$ such that

$$\frac{S_n}{g(n)}$$

has same variance (or standard deviation) for all n.