Written by Yarden Nativ, mail: nativyardenac@gmail.com

Self-learning recitation

Probability and Statistics

Probability and statistics:

"In experimental science you plan and carry out experiments, and then analyse and interpret the results. To do this you use statistical arguments and calculations. Like any other tool, you can use statistics delicately or clumsily, skilfully or ineptly. The more you know about it and understand how it works, the better you will be able to use it and the more useful it will be." (Statistics, R.J. Barrow).

We'll cover two basic topics in statistics:

Quantitative data: data is quantitative or numeric if it can be written using number (otherwise it is a qualitative data). Quantitative data can be divided into two types, continuous data and discrete. Every recording of continuous signal makes it discrete by using either averaging or a delta function (as will be explained later). In our course, we will deal with discrete quantitative data.

Written by Yarden Nativ, mail: nativyardenac@gmail.com

Random variable (RV):

The probability theory is a mathematical framework to study the patterns emerged from random events repeated. The probability of an event ω is marked by the function P(ω) and it defines the chance that this event will happen. This function is always positive and the sum of probabilities of event space $\omega_i \in \Omega$ is always $P(\Omega) = 1$. If the events are independent then $A \cap B = \phi \to P(A \cup B) = P(A) + P(B)$.

To map the connection between the events and a real number we use the function RV (random variable), which is a measure function that allow to talk about events using probability theory.

Self-Exercise:

For the following events, calculate how many there are in a deck of 52 playing cards with no jokers, and what is the probability of each one. Which of the following are independent events?

- "Red and black at the same time without being a joker"
- "The 5 of Hearts"
- "A King"
- "A Face card"
- "A Spade"
- "A Face card or a red suit"
- "A card"

Solution:

- "Red and black at the same time without being a joker" (0 elements, p=0)
- "The 5 of Hearts" (1 element, p=1/52)
- "A King" (4 elements, p=4/52)
- "A Face card" (12 elements, p=12/52)
- "A Spade" (13 elements, p=13/52)
- "A Face card or a red suit" (32 elements, p=32/52)
- "A card" (52 elements, p=1)

A spade and the 5 of hearts are independent events.

Cumulative distribution function (CDF), Probability density function (PDF):

CDF is a distribution function describing a range of possible outcomes, for example the probability of having a number smaller than 5 when throwing a cube. It is defined by:

$$F_X(x) = P(X \leq x),\ \ F(x) = \int_{-\infty}^{x} f(t)dt$$

Some of its features:

$$F_X(x = -\infty) = 0$$

$$F_X(x = \infty) = 1$$

$$x_1 < x_2 \rightarrow\ F_X(x_1) \leq F_x(x_2)$$

$$P(X > x_1) = 1 - F_X(x_1)$$

$$P(x_1 < x < x_2) = F_X(x_2) - F_x(x_1)$$

Using the last feature, we can define PDF by the limit $x_1 \rightarrow x_2$, practically write the derivative of CDF:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$$F_X(x_2) - F_x(x_1) = \int_{x_1}^{x_2} f_X(x) \cdot dx$$

We should note that only this function fulfils:

$$\int_{-\infty}^{\infty} f_X(x) \cdot dx = 1$$

Commonly we use PDF not in its continuous form but rather in its discrete one:

$$P(x \leq x_0) = F_X(x_0)$$

$$p(x) = f_X(x)$$

$$p(\Omega) = \sum_{i=1}^{n} f_X(x_i) = 1$$

Average, Variance and standard deviation (STD): Given N measurements of process x (meaning the vector $\{x_1, x_2, \dots, x_N\}$), the basic quantities to describe them are average and variance. The statistical equivalent to average is expectation and it represents the same ideas, just using probability functions instead of counting. Expectation and Variance are also considered as the first two moments of data.

Average: $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

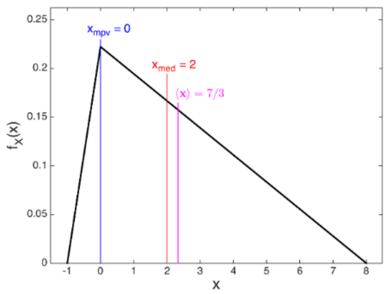
Expectation: discrete form $\langle x \rangle = \sum_{i=1}^{m} x_i P(x_i)$ **continuous form** $\langle x \rangle = \int_a^b x f_X(x) dx$

Variance: $V(x) = \langle x^2 \rangle - \langle x \rangle^2 = \langle (x - \langle x \rangle)^2 \rangle$ **STD:** $\sigma^2 = V(x)$

discrete form $V(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$ **continuous form** $V(x) = \int_a^b (x - \langle x \rangle)^2 f_X(x) dx$

Self-exercise:

A sensory neuron has the following probability function (PDF) for number of spikes every second. x represents the time axis, where 0 is the time of the stimulus presentation. Use the PDF $f_X(x) = \begin{cases} \frac{2}{9}(1+x) & -1 \le x \le 0 \\ \frac{2}{9}\left(1 - \frac{x}{8}\right) & 0 \le x \le 8 \end{cases}$



(Adopted from the open university physics practicum course, statistics introduction)

Calculate the expectation:

$$\langle x \rangle = \int_a^b x f_X(x) dx = \int_{-1}^{0} \frac{2}{9}(1+x)x dx + \int_0^8 \frac{2}{9}\left(1 - \frac{x}{8}\right) x dx$$

$$\int_{-1}^{0} \frac{2}{9}(1+x)x dx = \frac{2}{9}\int_{-1}^{0}(x + x^2)dx = \frac{2}{9}\left[\frac{1}{2}x^2 + \frac{1}{3}x^3\right]_{-1}^{0} = \frac{2}{9}\left(-\frac{1}{2} + \frac{1}{3}\right) = -\frac{1}{27}$$

$$\int_0^8 \frac{2}{9}\left(1 - \frac{x}{8}\right) x dx = \frac{2}{9}\int_0^8 (x - \frac{x^2}{8})dx = \frac{2}{9}\left[\frac{1}{2}x^2 - \frac{1}{24}x^3\right]_0^8 = \frac{2}{9}\left[\frac{1}{2} \cdot 8^2 - \frac{1}{24} \cdot 8^3\right] = \frac{64}{27}$$

$$\langle x \rangle = \int_{-1}^{0} \frac{2}{9}(1+x)x dx + \int_1^8 \frac{2}{9}\left(\frac{1-x}{8}\right) x dx = -\frac{1}{27} + \frac{64}{27} = \frac{7}{3}$$

Written by Yarden Nativ, mail: nativyardenac@gmail.com

Covariance, Pearson correlation coefficient:

Commonly we measure more than one item (for example two neurons). If we have N measures from each neuron, we can write them in pairs $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$. Alongside the averages and variances of each one of them, we can ask if these measurements are independent. This neurons can be dependent, for example neuron 1 sends inhibition signals to neuron 2 and by so lowers the probability to spikes in the second one. To describe the level of influence of one item on the probability of the other, we use covariance which is the shared variance of the neurons:

$$Cov(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x}\cdot\bar{y}$$

If the two items are fully dependent than $Cov(x,y) = V(x)\cdot a$ while a is some number. If they are independent than $Cov(x,y) \to 0$. Using this relationship, we can define a normalized correlation coefficient:

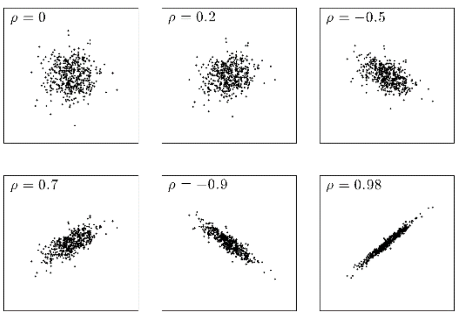
$$\rho = \frac{Cov(x,y)}{\sigma_x \cdot \sigma_y}$$

ρ is between -1 and 1 and can mean three things:

1. $\rho = 0$, than x and y are not correlated and the relationship between them is random.
2. $\rho = 1$, than x and y are in full correspondent and are considered correlated.
3. $\rho = -1$, than x and y are the opposite to each other and are considered anti-correlated but still correspondent.

Exercise:

For the following two variables, write if they have strong correspondent or not:



Answer: The correspondent between the variables is stronger in the last three examples, although one of them shows higher anti-correlation.

Written by Yarden Nativ, mail: nativyardenac@gmail.com

Distributions

Uniform distribution: a distribution of events in which each one has an equal probability. In the discrete case, for k events the probability of each one is $\frac{1}{k}$, while in the continuous case $f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & else \end{cases}$. Uniform distribution example: Decent cube.

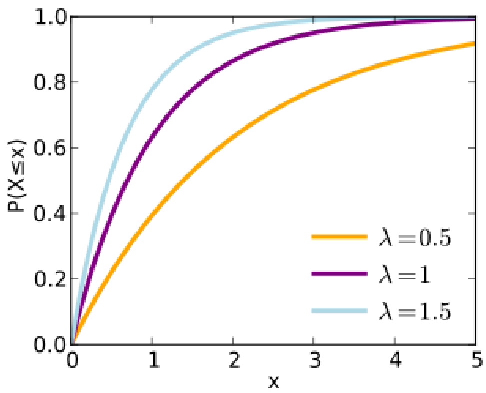
Binomial distribution: commonly used to described number of successes in a set of trials. Defined only for the discrete case: $f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in [a,b] \\ 0 & else \end{cases}$

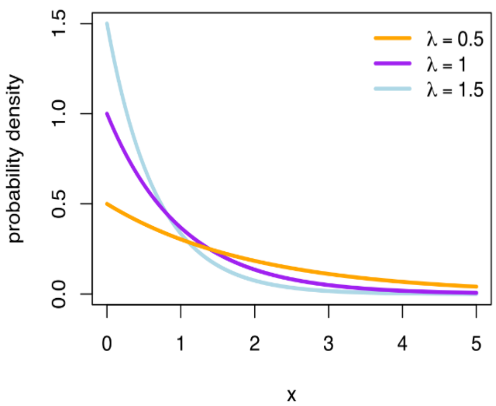
Exponential distribution: commonly used in neuroscience to describe the distribution of intervals between spikes of the common neuron model (Poissonian model). We'll study this subject in more depth in the future, but meanwhile these are the probability functions of this distribution:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (\lambda > 0)$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (\lambda > 0)$$

CDF ***PDF***

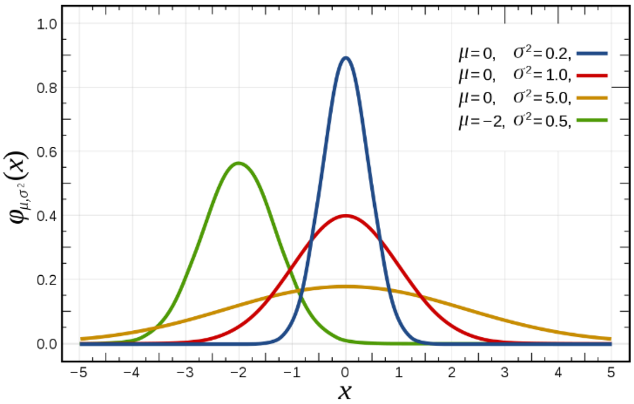




Written by Yarden Nativ, mail: nativyardenac@gmail.com

Normal distribution: Based on the central limit theorem, most distributions can be evaluated as normal distributions when the number of samples becomes bigger. The most standard way to describe a normal distribution is by using the PDF $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

PDF:



Written by Yarden Nativ, mail: nativyardenac@gmail.com

Higher moments: skewness (third moment) and curtosis (fourth moment).

Moments are defined as estimators for observations, expectation of powers of measurements.

$$m_n = \bar{X}^n = E(X^n), \quad n \geq 1$$

The first two moments are expectation and variance, but sometimes we are interested also in the third and fourth moments. The 3rd moment, skewness, is corresponding to the symmetry level of the distribution so that for normal distribution it is given by 0. The 4th moment, kurtosis, describes how “peaked” is the distribution and in normal distribution is given by 3. Illustrations: