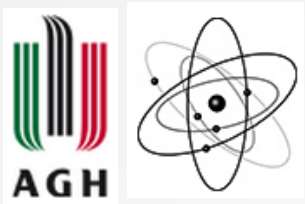


Introduction to probability, statistics and data handling

Tomasz Szumlak, Agnieszka Obłąkowska-Mucha

Faculty of Physics and Applied Computer Science

AGH UST Krakow

Random variable (I)

RANDOM VARIABLE - a `mapping' of the set of (elementary) events E onto the set of real numbers R.

For instance:

- height of a person met in the street;
- number of people in Krakow down with flu each day;
- number of meteorites falling each year per 1 km^2;
- number of minutes you wait every day for the street-car;
- number of accidents per months at a given street-intersection;
- strength of a climbing-rope;
- **a result of every measurement**.

Random variables (II)

- There is no surprise, we can have either discrete or continuous RV
- Now, let's have a discrete RV X that can assume the following values: $X = \{x_1, x_2, \dots, x_n\}$. Suppose, these values are assumed with certain probabilities:

$$p(X = x_i) = f(x_i), i = 1, 2, \dots, n$$

- We can introduce **probability function**, that we call **probability distribution** for RV X
- In general, any function can be a probability function if:
  - Its values are always positive: $\boldsymbol{f(x) \geq 0 \vee x \in X \subset \Omega}$
  - The sum taken over all possible x_i is: $\boldsymbol{\sum_X f(x) = 1}$
- It is easy to extend all of this to RVs that are continuous, so we will not do that here (in principle we should remember that the sum changes into the integral)

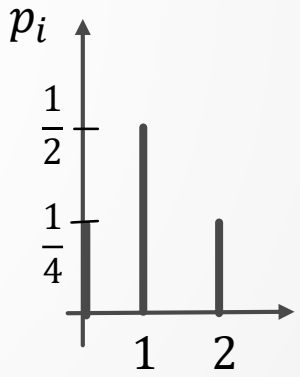
Random variables (III)

- **Ex. 2** Again, let's look at the double coin toss. How do we define the **probability distribution function** (P.D.F.)?
  - The sample space $\Omega = \{HH, HT, TH, TT\}$, each of these events has the same probability $p(HH) = p(HT) = \cdots = 1/4$
  - Using the Ex. 1 we can write:

$$p(X = 0) = p(TT) = 1/4$$

$$p(X = 1) = p(HT \cap TH) = 1/2$$

$$p(X = 2) = p(HH) = 1/4$$



x	0	1	2
$f(x)$	1/4	1/2	1/4

Distribution function (I)

- Closely related to P.D.F. is the **cumulative distribution function (CDF)**
- We define it as follow:

$$F(x) = p(X \leq x)$$

- The CDF has the following properties
  - $F(x)$ must be non-decreasing

    $F(x_i) \leq F(x_j) \rightarrow x_i \leq x_j$
  - Asymptotic behaviour

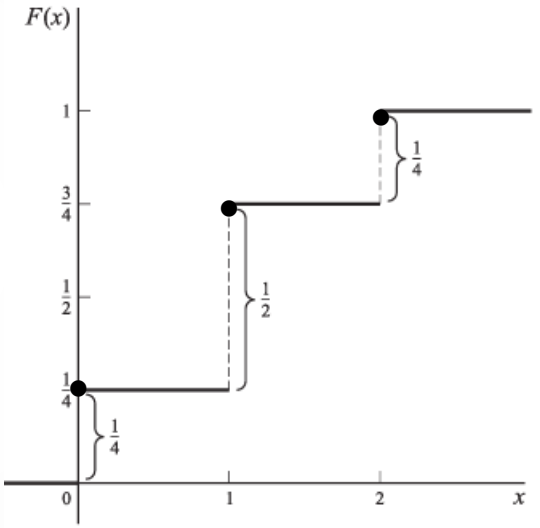
    $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$
  - DF is continuous from the right

    $\lim_{h \to 0^+} F(x+h) = F(x), \ \forall\, x \in \mathrm{X}$

Distribution function (II)

- **Ex. 4** Again, taking the two tosses example, we can work out the distribution function.

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{4} & 0 \le x < 1 \\ \frac{3}{4} & 1 \le x < 2 \\ 1 & 2 \le x < \infty \end{cases}$$



Some textbooks use a slightly different definition

$$\boxed{F_X(x) \equiv F(x) = \mathcal{P}(X < x)}$$

It has no any influence in the case of continuous RV; but for a discrete RV **it makes quite a difference**

Continuous RV

- There is a natural extension to continuous RV, however the exact definition is based on the properties of the distribution function
- **Def.1** We say, that a non-discrete random variable X is continous if its distribution function may be represented as:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du\,, (-\infty < x < \infty)$$

- We know already, that the function $f(x)$ should represent a P.D.F:

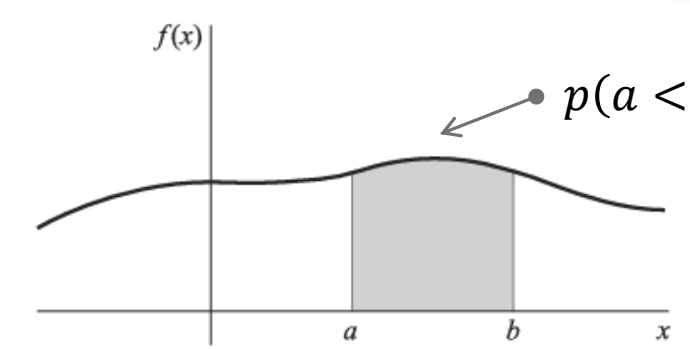
$$f(x) \geq 0 \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

- There are some interesting properties related to the CRV
  - The probability that X takes on any one particular value is zero!
  - The interval probability can be estimated as:

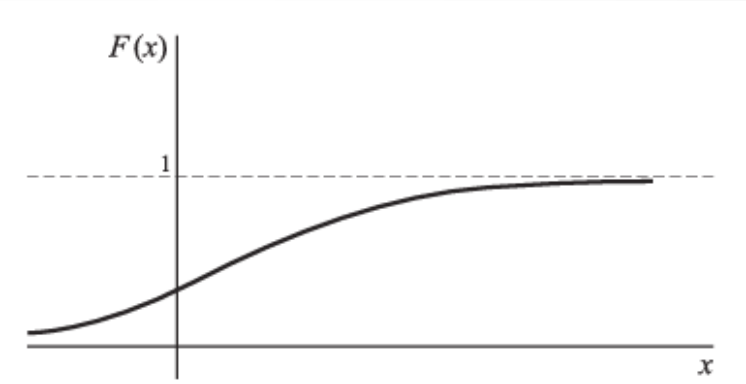
$$p(a < X < b) = \int_{a}^{b} f(x)dx$$

Graphical interpretation

- Let f(x) be the **density function** for a random variable X. This function can be represented as a graph by some curve



- The **distribution function** is, in turn, a monotonically increasing function which value goes from 0 to 1



Independent RVs

- We learned how to calculate probability of independent events:

$$p(\mathbb{A} \cap \mathbb{B}) = p(\mathbb{B}|\mathbb{A})p(\mathbb{A}) = p(\mathbb{B})p(\mathbb{A})$$

- This definition can also be used for probability functions. Say, X and Y are RVs. If the events $X = x$ and $Y = y$ are independent for all x and y, then we say that X and Y are independent RVs. We also have:

$$p(X = x, Y = y) = p(X = x) \cdot p(Y = y)$$

$$f(x, y) = f_1(x) f_2(y)$$

- Similarly, we say that X and Y are independent RVs if the events $X \leq x$ and $Y \leq y$ are independent for all x and y. We can write:

$$p(X \leq x, Y \leq y) = p(X \leq x)p(Y \leq y) \rightarrow F(x, y) = F_1(x)F_2(y)$$

Mathematical Expectation

- The **mathematical expectation** (or expected value) is one of the most important notions in statistics. Let's start (as usual…) from a DRV
- **Definition 1**. Assume that X is a DRV having the possible values as follow $\{x_1, x_2, \dots, x_n\}$, the **expectation** of X is defined as:

$$E[X] = x_1 \cdot p(X = x_1) + \cdots + x_n \cdot p(X = x_n) = \sum_{i/1}^{i/n} x_i \cdot p(X = x_i)$$

$$\boldsymbol{E[X] = x_1 \cdot f(x_1) + \cdots + x_n \cdot f(x_n) = \sum_{i/1}^{i/n} x_i \cdot f(x_i)}$$

where: $f(x_i)$ is the DRV's P.D.F.

- NOTE. When the respective probabilities for events x_i are all equal, we have (**arithmetic mean**):

$$E[X] = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_i x_i$$

Mathematical Expectation

- For a CRV X having P.D.F. $f(x)$, the expectation of X is defined as follow (we always silently assume that the integral converges absolutely):

$$\boldsymbol{E[X] = \int_{-\infty}^{\infty} x f(x) dx}$$

- If we know completely the P.D.F. of RV X then we call $\boldsymbol{E[X]}$ the **mean value** of **X** and denote it by $\boldsymbol{\mu_X}$ **(true mean value)**
- Various notation: $E(X), \hat{x}, \mu, m$ (estimated mean value)
- Since, the mean gives a single value that represents the values of RV X we call it a **measure of central tendency** (remember we loose something here – data reduction)

Mathematical Expectation

- NOTE something quite here – the mean value is a **single number** – it is not a RV! We call it a parameter of a RV X's P.D.F.
- The crucial point is that we **assumed that we know** and **understand** the P.D.F. of a RV X – we are going to learn soon that this is usually **not the case**! All we can know is a **sample** drawn from a **population** that is described by an **unknown** P.D.F. **This is the core of statistical reasoning**!

Functions of RV

- Functions of RV are of great importance for statistics
- Interestingly such function is also a RV itself! For instance, you calculate how much you earn when you sell x items, each 10E (X is a RV):

$$\boldsymbol{Y = K(X)}$$

- Now, we can write formulas for the expectation value in a similar manner to defined in previous slides:

$$E[K(X)] = \sum_{i/1}^{i/n} K(x_i) \cdot f(x_i) = \sum_{i/1}^{i/n} K(x_i) \cdot p(X = x_i)$$

$$\boldsymbol{E[K(X)] = \int_{-\infty}^{\infty} K(X) f(x) dx}$$

- In particular we can pick a special function: $K(X) = \boldsymbol{(X - \alpha)^l}$, and its expectation values are called l-th moments about point α (constant)

$$m_l = E[(X - \alpha)^l]$$

Moments

- The **rth moment of a RV X about the mean μ**, also called the **rth central moment**, is defined as follow:

$$\mu_r = E[(X-\mu)^r], r = 0, 1, 2, \ldots$$

$$\mu_0 = 1, \mu_1 = 0, \mu_2 = \sigma^2, \ldots$$

- Assuming absolute convergence we write explicitly for both DRV and CRV:

$$\mu_r = \sum (x-\mu)^r f(x)\,, \mu_r = \int_{-\infty}^{\infty} (x-\mu)^r f(x) dx$$

- The **rth moment of a RV X about the origin**, also called the **rth raw moment**, is defined as:

$$\mu'_r = E[X^r]$$

$$\mu'_0 = \mu_0 = 1, \mu'_1 = \mu$$

Moments

Let's consider the case of a continuous variable:

$$\mu_0 = \int_{-\infty}^{\infty} (x-\hat{x})^0 f(x)dx = 1$$

$$\mu_1 = \int_{-\infty}^{\infty} (x-\hat{x})^1 f(x)dx = 0$$

$$\mu_2 = \int_{-\infty}^{\infty} (x-\hat{x})^2 f(x)dx \stackrel{\text{def}}{=} VAR(X) = \sigma^2(X) = \text{VARIANCE}$$

$$\mu_3 = \int_{-\infty}^{\infty} (x-\hat{x})^3 f(x)dx = \text{SKEWNESS}$$

$$\mu_4 = \int_{-\infty}^{\infty} (x-\hat{x})^4 f(x)dx = \text{KURTOSIS}$$

- VARIANCE — a measure of the spread (dispersion) (always > 0)
- SKEWNESS — a measure of asymmetry
- KURTOSIS— a measure of the spread as compared with a special type of distribution – normal distribution

Moments

- A general formula that relates the both types of moments can be written as follow:

$$\mu_r = \mu'_r - \binom{r}{1}\mu'_{r-1}\mu + \cdots + (-1)^j \binom{r}{j}\mu'_{r-j}\mu^j + \cdots + (-1)^r \mu'_0 \mu^r$$

$$\mu_2 = \mu'_2 - \mu^2$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu + 2\mu^3$$

Note, by using moments we can describe any probability distribution function. This is not trivial! Sometimes we do not know (even in principle) what is the P.D.F. of a RV X

- Assuming some concrete function may lead to completely wrong results of statistical analysis. However, we still can calculate the moments using a sample data taken experimentally
- We need, in principle, infinite number of moments to describe a given P.D.F.

Mathematical Expectation

- **Example 1**. Say, that we play a game where we toss a single die (assumed fair). A player wins if she/he has 2 (20$) or 4 (40$), looses if a 6 turns up. Find the expected amount of money to be won:

$$E[X] = (0\$) \cdot \left(\frac{1}{6}\right)_1 + (20\$) \cdot \left(\frac{1}{6}\right)_2 + (0\$) \cdot \left(\frac{1}{6}\right)_3 + (40\$) \cdot \left(\frac{1}{6}\right)_4$$
$$+(0\$) \cdot \left(\frac{1}{6}\right)_5 + (-30\$) \cdot \left(\frac{1}{6}\right)_6 = 5$$

x_j	0	+20	0	+40	0	−30
$f(x_j)$	1/6	1/6	1/6	1/6	1/6	1/6

- A player is expected to win 5$. So, for the game to be fair she/he is expected to pay 5$ in order to play the game…
- For fun – you can check if „Euro Millions" is a fair game…

Mathematical Expectation

- **Example 2.** Let's have a look how this works for a CRV. Say, the density function of a CRV X is given by:

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & otherwise \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{2} x\left(\frac{1}{2}x\right)dx = \int_{0}^{2} \frac{x^2}{2}dx =$$

$$= \frac{x^3}{6}\Big|_0^2 = \frac{4}{3}$$

Theorems on $E[X]$

- **Theorem 1**. If c is any constant, then:

$$\boxed{E[cX] = cE[X]}$$

$$E[cX] = \sum_i cx_i f(x_i) = cx_1 f(x_1) + \cdots + cx_n f(x_n) =$$

$$= c\big(x_1 f(x_1) + \cdots + x_n f(x_n)\big) = c\sum_i x_i f(x_i) = cE[X]$$

- **Theorem 2.** If X and Y are any RVs, then:

$$\boxed{E[X+Y] = E[X] + E[Y]}$$

$$E[X+Y] = \sum_i \sum_j (x_i + y_j) f(x_i, y_j) = \sum_i \sum_j x_i f(x_i, y_j) +$$

$$+ \sum_i \sum_j y_j f(x_i, y_j) = E[X] + E[Y]$$

Theorems on $E[X]$

- **Theorem 3**. If X and Y are independent RVs, then

$$\boxed{E[XY] = E[X]E[Y]}$$

If X and Y are independent their joint P.D.F. can be factorised:

$$f(x,y) = f_1(x)f_2(y)$$

$$E[XY] = \sum_i \sum_j x_i y_j f(x_i, y_j) = \sum_i \sum_j x_i y_j f_1(x_i) f_2(y_j) =$$

$$= \sum_i \left[x_i f_1(x_i) \sum_j y_j f_2(y_j) \right] = \sum_i [x_i f_1(x_i) E[Y]] =$$

$$= E[Y] \sum_i x_i f_1(x_i) = E[X]E[Y]$$

Variance

- Another important metric used in statistics is **variance**

$$V[X] = E[(X - \mu)^2]$$

- The variance **cannot** be a **negative number**, the positive square root of the variance is called the **standard deviation**

$$\sigma_X = \sqrt{V[X]} = \sqrt{E[(X - \mu)^2]}$$

- Writing explicitly we have:

$$V[X] = \sigma_X^2 = \sum_{i/1}^{i/n} (x_i - \mu)^2 f(x_i)$$

- If the probabilities are all equal, we have

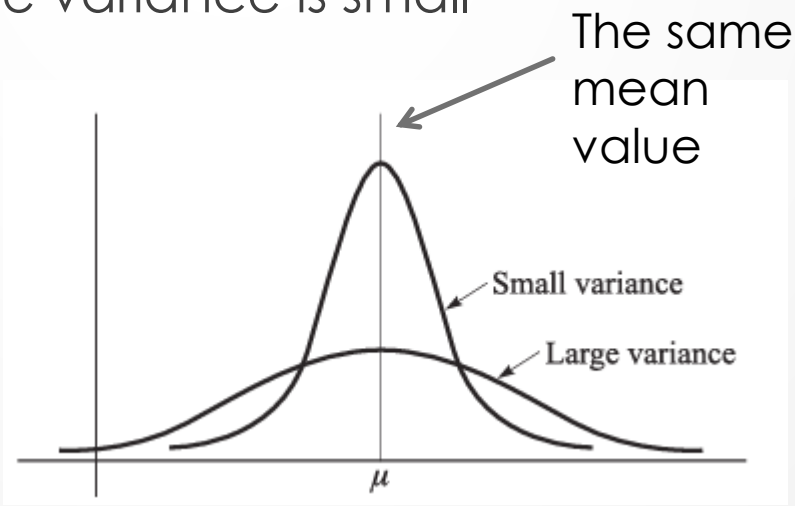
$$\sigma^2 = \frac{1}{n}[(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_n - \mu)^2]$$

Variance

- In case when X is a CRV, we can write the variance as:

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

- We say that the variance is a measure of **the dispersion** (scatter) of the values of the RV **about the mean value** μ. For instance, if the values tend to be concentrated close to the mean, the variance is small



Theorems regarding variance

- **Theorem 4**. Let X be any RV:

$$\sigma^2 = E[(X-\mu)^2] = E[X^2] - \mu^2 = E[X^2] - E^2[X]$$

- **Theorem 5**. If c is any constant, we have:

$$V[cX] = c^2 V[X]$$

- **Theorem 6**. The quantity $E[(X-a)^2]$ is a minimum when $a = E[X]$

- **Theorem 7**. If X and Y are independent RVs,

$$V[X+Y] = V[X] + V[Y], \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

$$V[X-Y] = V[X] + V[Y], \sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

Change of variables (I)

- Let's assume we know distribution functions of one or more RVs. In practice, we are often interested in finding distributions of other RVs that depend on them (here we focus on CRV)
- **Theorem 1.** Let X be a CRV with P.D.F. given by $f(x)$. Next, define RV $U = \varphi(X)$, where $X = \omega(U)$. The P.D.F. of U is given by $g(u)$ where:

$$g(y)|dy| = f(x)|dx|$$

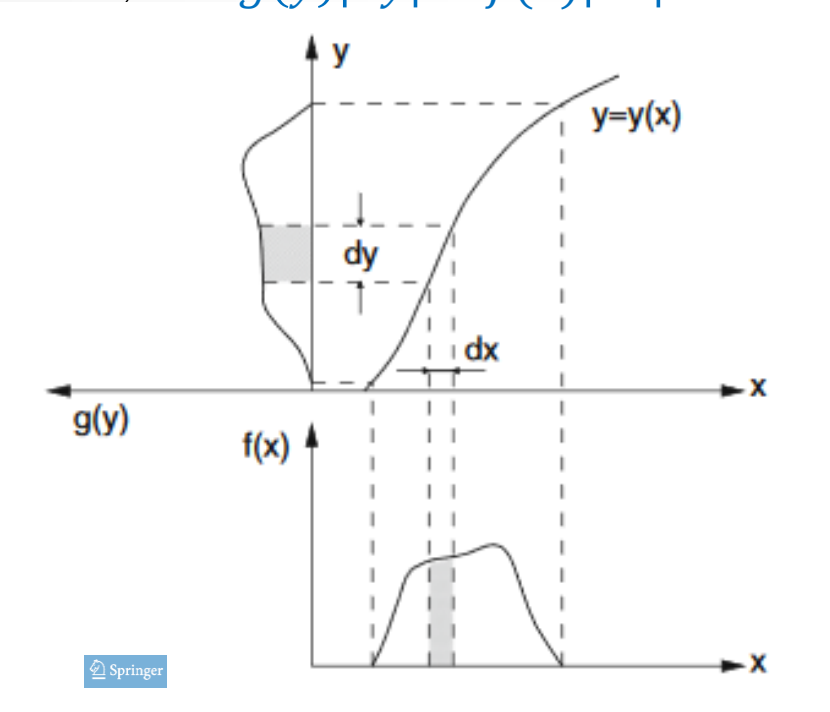
$$g(y) = f(x)\left|\frac{dx}{dy}\right| = f\big(\omega(y)\big)\omega'(y)$$

- $\frac{dx}{dy}$ is the derivative of X with respect to Y.

Change of variables (II)

For a non-monotonic $y = y(x)$ dependence one must take into account that different regions of the X variable may be mapped into one (the same) region of the Y variable. The $g(y)$ pdf in such a region will be a sum of f(x) pdf's multiplied by $\left|\frac{dx}{dy}\right|$ over all the regions of X which have been mapped into the given region of Y .

$$g(y)|dy| = f(x)|dx|$$



$$g(y) = f(x)\left|\frac{dx}{dy}\right|$$