Introduction to probability, statistics and data handling

Tomasz Szumlak, Agnieszka Obłąkowska-Mucha
Faculty of Physics and Applied Computer Science

AGH UST Krakow

Joint distributions (I)

- If X and Y are two discrete RVs, we can define P.D.F. of X and Y as follow:

$$p(X = x, Y = y) = f(x, y)$$

$$f(x, y) \geq 0$$

$$\sum_x \sum_y f(x, y) = 1$$

- If we assume that: $X = \{x_1, x_2, \dots, x_m\}$ and $Y = \{y_1, y_2, \dots, y_n\}$, then the probability of the event that $X = x_i$ and $Y = y_j$ is given by:

$$p\left(X = x_i, Y = y_j\right) = f\left(x_i, y_j\right)$$

- Respective probabilities for $X = x_i$ and $Y = y_j$ are given by:

$$p(X = x_i) = f_1(x_i) = f_x(x_i) = \sum_k f(x_i, y_k)$$

$$p\left(Y = y_j\right) = f_2\left(y_j\right) = f_y\left(y_j\right) = \sum_l f\left(x_l, y_j\right)$$

Joint distributions (II)

X \ Y	y_1	y_2	$\cdots$	y_n	Totals ↓	
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\cdots$	$f(x_1, y_n)$	$f_1(x_1)$	
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\cdots$	$f(x_2, y_n)$	$f_1(x_2)$	
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	$\cdots$	$f(x_m, y_n)$	$f_1(x_m)$	
Totals →	$f_2(y_1)$	$f_2(y_2)$	$\cdots$	$f_2(y_n)$	1	← Grand Total

Rigt Margin

Bottom Margin

- Because the respective probabilities: $p(X = x_i)$ and $p(Y = y_j)$ are found on the margins of the joint probability table, we call both functions $f_1(x_i)$ and $f_2(y_j)$ the **marginal probability functions** of X and Y

Example: Students in a class of 100 were classified according to gender (G) and smoking (S) as follows:

		S			
		s	q	n	
G	male	20	32	8	60
	female	10	5	25	40
		30	37	33	100

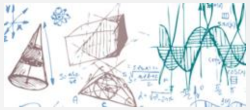
s: “now smokes”,
q: “did smoke but quit”
n: “never smoked”.

Identify:

- joint distribution function
- marginal distributions

Find the probability that a randomly selected student

1. is a male;
2. is a male smoker;
3. is either a smoker or did smoke but quit;
4. is a female who is a smoker or did smoke but quit.



Joint distributions (III)

- It is essential to note that for both marginal density functions we have:

$$\sum_i f_1(x_i) = 1, \sum_j f_2(y_j) = 1$$

- The two above are equivalent of:

$$\sum_i \sum_j f(x_i, y_j) = 1$$

- Since all of these functions represent P.D.F. they must be normalised, or in other words the probability of all entries is 1
- The **joint distribution function** of RVs X and Y is given by

$$F(x, y) = p(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v)$$

- So, to get a value of $F(x, y)$ for a given pair (x, y) we need to sum-up all the entries for which $x_i \leq x$ and $y_j \leq y$

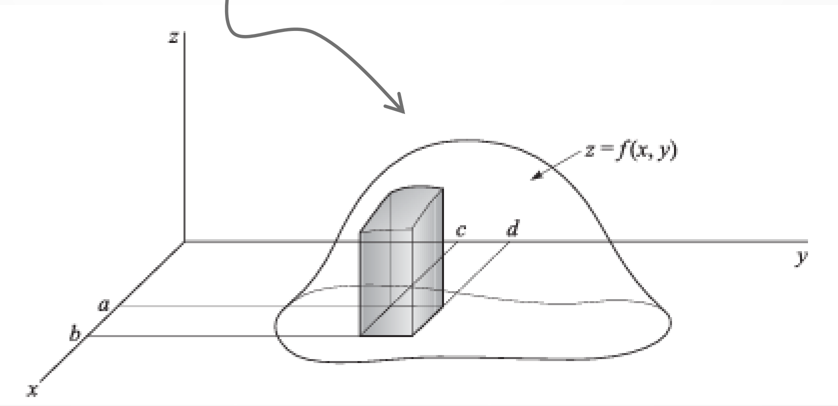
Joint distributions (IV)

- Again, by analogy we can easily obtain the joint probability function for continuous RVs X and Y:

$$f(x,y) \geq 0, \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$$

- The probability can be estimated using the joint P.D.F. as follow:

$$p(a < X < b, c < Y < d) = \int_{a}^{b}\int_{c}^{d} f(x,y)dxdy$$



Joint distributions (V)

- Now, we can define the joint distribution function:

$$F(x,y) = p(X \leq x, Y \leq y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv$$

- We also have the following:

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y)$$

- By analogy, the respective marginal functions can be defined for both density, $f(x,y)$, and distribution, $F(x,y)$ functions

$$f_1(x) = \int_{-\infty}^{\infty} f(x,v)dv \,, f_2(y) = \int_{-\infty}^{\infty} f(u,y)du$$

$$F_1(x) = p(X \leq x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,v)dudv$$

$$F_2(y) = p(Y \leq y) = \int_{-\infty}^{\infty}\int_{-\infty}^{y} f(u,v)dudv$$

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:

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

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

- 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:

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

Conditional P.D.F.s

- Let's assume that X and Y are CRVs. We define the conditional density function of Y given X, as:

$$f(y|x) = \frac{f(x,y)}{f_1(x)}$$

$$f(x|y) = \frac{f(x,y)}{f_2(y)}$$

- So, to define the conditional P.D.F. we need a joint P.D.F. and a marginal one to calculate an appropriate probability we do:

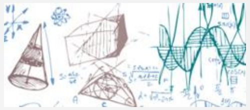
$$p(c < Y < d|x < X < x + dx) = \int_c^d f(y|x)dy$$

Example: Students in a class of 100 were classified according to gender (G) and smoking (S) as follows:

		S			
		s	q	n	
G	male	20	32	8	60
	female	10	5	25	40
		30	37	33	100

Calculate the probability that a randomly selected student is

1. a smoker given that he is a male;
2. female, given that the student smokes.



Covariance

- Next step, as usual, lead to more RVs. Let's see what's new if we consider two RVs X and Y with joint density function $f(x,y)$:

$$\mu_X = E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy$$

$$\mu_Y = E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy$$

$$\sigma_X^2 = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)^2 f(x,y)dxdy$$

$$\sigma_Y^2 = E[(Y-\mu_Y)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\mu_Y)^2 f(x,y)dxdy$$

- And what about the mixed terms? Analysis leads to the **covariance**

$$\boxed{\sigma_{XY} = Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]}$$

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f(x,y)dxdy$$

Theorems regarding covariance

- **Theorem 1**. For any RVs the following is true:

$$\sigma_{XY} = Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - E[X]E[Y]$$

- **Theorem 2**. In case the RVs X and Y are independent:

$$Cov[X,Y] = 0$$

- **Theorem 3**. For any two RVs we have:

$$V[X \pm Y] = V[X] + V[Y] \pm 2Cov[X,Y]$$

- **Theorem 4**. For any two RVs we have:

$$|\sigma_{XY}| \le \sigma_X \sigma_Y$$

- NOTE, that the converse of Theorem 9 is not necessarily true!

Correlation coefficient

- The covariance gives us a strong hint on how to measure **the dependence** of RVs. If X and Y are independent, then:

$$Cov[X,Y] = \sigma_{XY} = 0$$

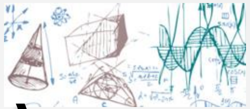
- On the other hand, if they are completely dependent (e.g., $X = Y$), then:

$$Cov[X,Y] = \sigma_{XY} = \sigma_X \sigma_Y$$

- So, one can use the following to measure the dependence of RVs:

$$\boxed{\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}}$$

- We call it the correlation coefficient and it is easy to note that its values vary between $[-1,1]$
- In case where the CC is equal zero, we call the RVs linearly uncorrelated. In general, however, the variables may or may not be independent.



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:

$$g(u)|du| = f(x)|dx|$$

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

- For more than one variable things getting a bit more difficult…
- **Theorem 2.** Let X and Y be CRVs having joint P.D.F. $f(x, y)$. Let's define new variables $U = \varphi_1(X, Y)$ and $V = \varphi_2(X, Y)$, where $X = \omega_1(U, V)$ and $Y = \omega_2(U, V)$. Then the joint density function of U and V is given as:

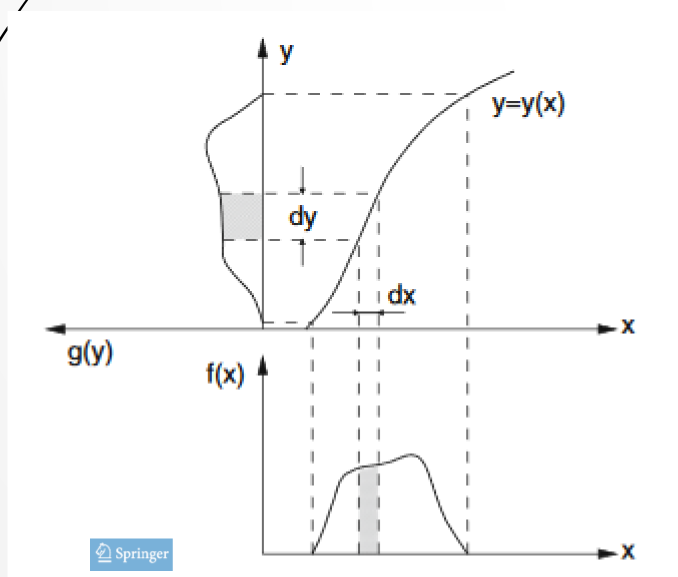
$$g(u, v)|dudv| = f(x, y)|dxdy|$$

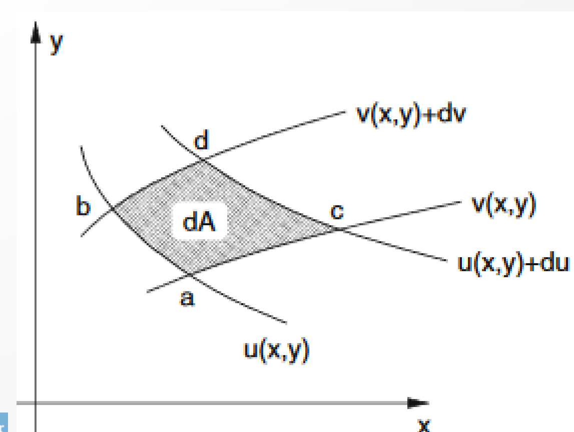
Change of variables (II)

$$g(u,v) = f(x,y)\left|\frac{\partial(x,y)}{\partial(u,v)}\right| = f(\omega_1(u,v),\omega_1(u,v))|J|$$

- For multi-dimensional case we have something brand new – Jacobian determinant or Jacobian

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$





Springer

Standarised RVs

- Let X be a RV with the mean value μ and standard deviation σ. We can define an associated RV that is called **standardised random variable**:

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

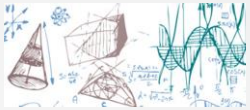
- Note, that X^* has a mean of zero and a variance of 1 – this is why we call it standardised in the first place!

$$E[Z] = 0, V[Z] = 1$$

$$E[Z] = E\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma} E[(X - \mu)] = \frac{1}{\sigma}(E[X] - \mu) = 0$$

$$V[Z] = V\left[\frac{X - \mu}{\sigma}\right] = \frac{1}{\sigma^2} E[(X - \mu)^2] = 1$$

- We will be using the SRV all the time – it makes comparison of different distributions possible.



More than μ and σ

- Sometimes we are more interested in the most probable value of RV instead of the mean (especially important for asymmetric distributions)
- The MPV, also called the mode is the value of the random variable that corresponds to the highest probability:

$$\mathcal{P}(X = x_m) = max$$

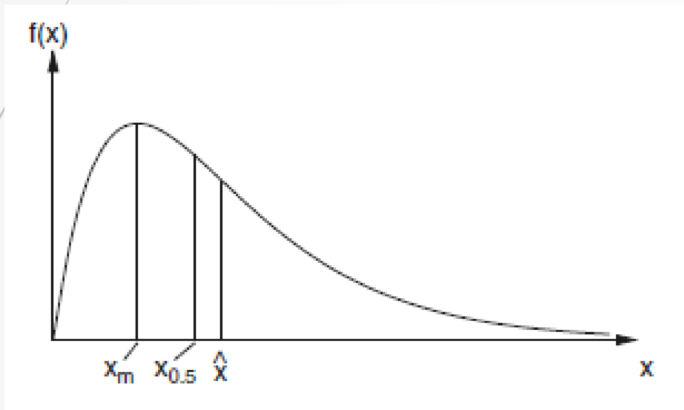
- In case we have a regular function representing the P.D.F. then the mode can be easily found:

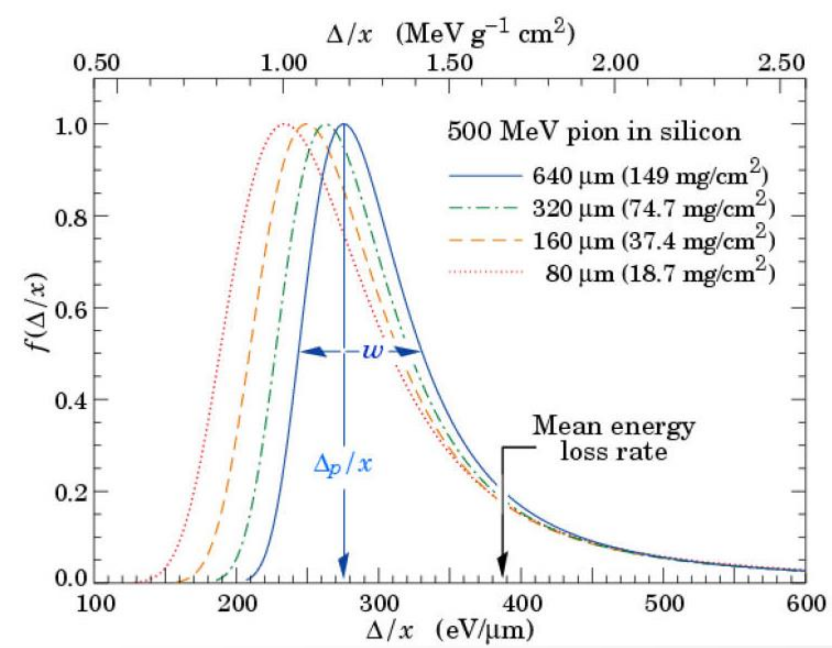
$$\frac{d}{dx} f(x) = 0, \frac{d^2}{dx^2} f(x) < 0$$

- If a given P.D.F has just one maximum, we call it a unimodal.
- The median can also be defined using the distribution function:

$$F(x_{0.5}) = \mathcal{P}(X < x_{0.5}) = 0.5$$

More than μ and σ



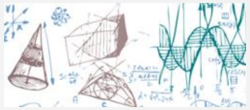


- It is also useful to define **quantiles**:

$$F(x_{0.25}) = 0.25, F(x_{0.75}) = 0.75$$

- And deciles…

$$F(x_q) = \int_{-\infty}^{x_q} f(x)dx = q$$



Chebyshev's Inequality

- There is an extraordinary theorem related to the fundamental properties of RV (both discrete and continuous). We just need both the expectation value and variance to be finite.
- **Theorem 5**. Suppose that X is a random variable. Let the mean and variance of this RV be μ and σ^2 respectively. If we assume that they are both finite, then if ϵ is any positive number:

$$p(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

$$\epsilon = k\sigma \rightarrow p(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- For instance, let $k = 2$:

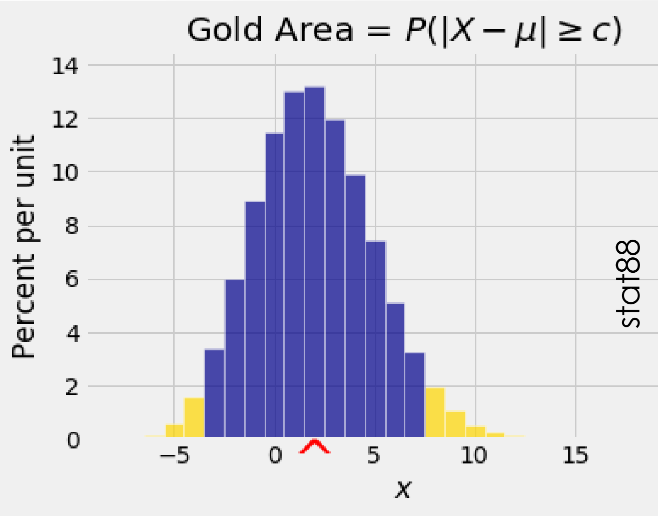
$$p(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \rightarrow p(|X-\mu| \geq 2\sigma) \leq \frac{1}{4}$$

$$p(|X-\mu| < 2\sigma) \geq \frac{3}{4}$$

Chebyshev's Inequality

- This simple rule is actually quite incredible – **without** making any assumptions regarding the probability distribution!

  - For any RV the probability of X being different from its mean value by more that **two** „standard deviations" is less than $1/4$ (25%).
  - … at least 3σ is less than $1/9$,
  - … at least 4σ is less than $1/16$,
  - …..



- Chebyshev's inequality is a probability theory that guarantees only a definite fraction of values will be found within a specific distance from the mean of a distribution.
- The fraction for which no more than a certain number of values can exceed is represented by $1/k^2$.

Chebyshev's Inequality

- Example:

Suppose that we extract an individual at random from a population whose members have an average income of \$10 000, with a standard deviation of \$3 000.

- What is the probability of extracting an individual whose income is either less than \$5 000 or greater than \$15 000?
- In the absence of more information about the distribution of income, we cannot compute this probability exactly. However, we can use Chebyshev's inequality to compute an upper bound to it.

if RV X denotes income, $\mu = 10\,000,\ \sigma = 3\,000,\ \ |X - 3\,000| \geq k$ and k=5 000

$$P(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2} = \frac{3\,000^2}{5\,000^2} = \frac{9}{25} = 36\%$$