National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”
Institute of Physics and Technology

Lecture 2
Review of Probability Theory

Dmytro Progonov,
PhD, Associate Professor,
Department Of Physics and Information Security Systems

Content

- Historical background;
- Definition of probability;
- Basic rules for probabilities
- Parameters of distributions.

History of probability theory

Abu Yūsuf Yaʻqūb ibn ʼIsḥāq aṣ-Ṣabbāḥ al-Kindī
(803-873)

Gerolamo Cardano
(1501-1576)

Ancient and Medieval time

- Law of evidence.

Renaissance time

- Risk theory foundation.

18th century

- Mathematical foundation of classic probability theory.

Jacob Bernoulli
(1655-1705)

19th century

- Theory of errors;
- statistical mechanics.

Carl Friedrich Gauss
(1777-1855)

20th century

- Hypothesis testing;
- Markov process theory.

Andrey Kolmogorov
(1903-1987)

Definitions of probability (1/2)

Pierre-Simon Laplace

(1749-1827)

Probability theory is nothing but common sense reduced to calculation (**Pierre Laplace**, 1812)

Definitions of probability (2/2)

For ***discrete random variable*** $X \in \mathcal{X}$ *probability mass function* of the event that $X = x$ is:

$$p(X = x) = p(x)$$

such as

$$\sum_{x \in \mathcal{X}} p(x) = 1\,; p(x) \in [0; 1].$$

For ***continuous random variable*** $X \in \mathcal{X}$ probability that X lies in any interval $a < X \leq b$ can be computed as:

$$p(a < X \leq b) = F(b) - F(a),$$

where $F(q) \triangleq p(X \leq q)$ – *cumulative distribution function*.

Basic rules of probability (1/3)

Probability of a union of two events

Given two events, A and B, probability of A or B defines as:

$$p(A \vee B) = \begin{cases} p(A) + p(B) & A \text{ and } B \text{ is mutually exclusive;} \\ p(A) + p(B) - p(A \wedge B) & \text{otherwise.} \end{cases}$$

Joint probabilities (product rule)

Probability of the joint event A and B defines as:

$$p(A, B) = p(A \wedge B) = p(A|B) \cdot p(B).$$

Conditional probability

Probability of event A, given that even B is true, defines as:

$$p(A|B) = \frac{p(A, B)}{p(B)}, p(B) > 0.$$

Basic rules of probability (2/3)

Marginal distribution

Given a joint distribution on two events $p(A, B)$, marginal distribution defines as:

$$p(A) = \sum_b p(A, B) = \sum_b p(A|B = b) \cdot p(B = b),$$
$$p(B) = \sum_a p(B, A) = \sum_a p(B|A = a) \cdot p(A = a).$$

Bayes' rule

$$p(X = x|Y = y) = \frac{p(X = x, Y = y)}{P(Y = y)},$$
$$p(X = x|Y = y) = \frac{p(X = x) \cdot p(Y = y|X = x)}{\sum_{\acute{x}} p(X = \acute{x}) \cdot p(Y = y|X = \acute{x})}.$$

Basic rules of probability (3/3)

Independence and conditional independence

Event A and B are unconditionally (marginally) independent if

$$A \perp B \Leftrightarrow p(A,B) = p(A) \cdot p(B).$$

Event A and B are conditionally independent given C if

$$A \perp B|C \Leftrightarrow p(A,B|C) = p(A|C) \cdot p(B|C).$$

Theorem: $A \perp B|C$ if and only if (iff) there exist function $g(\cdot)$and $h(\cdot)$ such that

$$p(A,B|C) = g(A,C) \cdot h(B,C).$$

Parameters of distributions (1/5)

Mean value

For discrete random variable:

$$\mathbb{E}[X] \triangleq \sum_{x \in \mathcal{X}} x \cdot p(x)\,,$$

For continuous random variable:

$$\mathbb{E}[X] \triangleq \int_{-\infty}^{+\infty} x \cdot p(x) dx\,.$$

Properties:

1. Expected value of constant – $\mathbb{E}[c] = c, c = const$;
2. Linearity – $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$;
3. Iterated expectation – $\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X|Y]\big]$;
4. Functional non-invariance – $\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(X) dP \neq g(\mathbb{E}[X])$.

Parameters of distributions (2/5)

Variance

$$var[X] \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2],$$
$$var[X] \triangleq \mathbb{E}[X^2] - \mathbb{E}^2[X],$$
$$std[X] \triangleq \sqrt{var[X]}.$$

Properties:

1. Non-negativity – $var[X] \geq 0$;
2. Variance of constant value –$var[c] = 0, c = const$;
3. Invariance to changes of location parameters:

$$var[X + c] = var[X], c = const;$$

4. Variance of scaled values – $var[cX] = c^2 var[X]$;
5. Variance of a sum of several variables:

$$var[aX \pm bY] = a^2 var[X] + b^2 var[Y] \pm 2(ab)Cov[X, Y],$$
$$Cov[X, Y] = \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])].$$

Parameters of distributions (3/5)

Quantiles

Let us denote inverse F^{-1} for monotonically increasing function (cumulative distribution function) F. Then $F^{-1}(\alpha), \alpha \in [0;1]$ is the value of x_α such that $p(X \leq x_\alpha) = \alpha$. This is called the **$\boldsymbol{\alpha}$ −quantile of F**.

Median – quantile $F^{-1}(0.5)$;

Lower and upper quantile – $F^{-1}(0.25)$ and $F^{-1}(0.75)$;

Interquartile range – $IQR = F^{-1}(0.75) - F^{-1}(0.25)$.

Parameters of distributions (4/5)

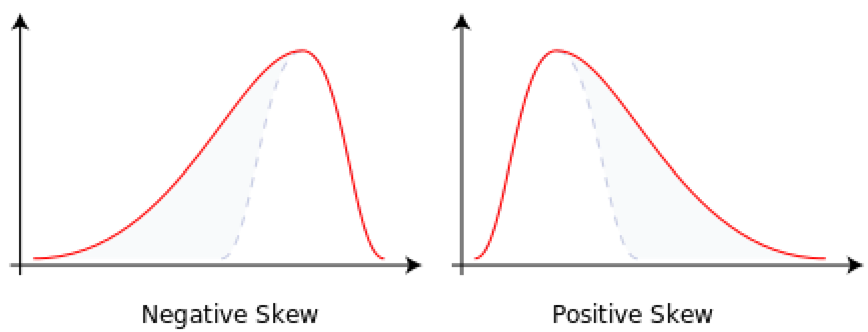
Skewness

$$\gamma_1 \triangleq \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mathbb{E}[(X-\mu)^3]}{(\mathbb{E}[(X-\mu)^2])^{3/2}},$$

where μ – mean value; σ – standard deviation.

Negative skew – the ***left tail*** of distribution is longer; the mass of the distribution is concentrated on the right;

Positive skew – the ***right tail*** of distribution is longer; the mass of the distribution is concentrated on the left.



Parameters of distributions (5/x)

Kurtosis

$$\gamma_1 \triangleq \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mathbb{E}[(X-\mu)^3]}{(\mathbb{E}[(X-\mu)^2])^2} - 3,$$

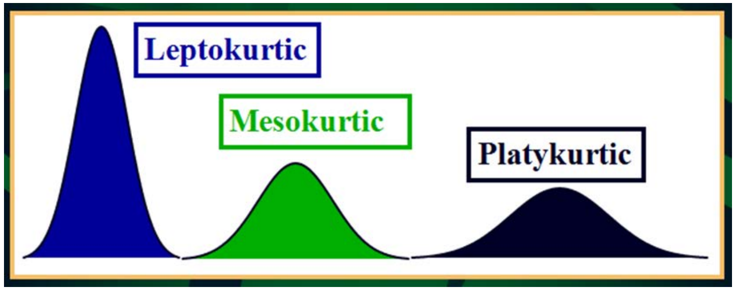
where μ − mean value; σ − standard deviation.

For Gaussian (normal) distribution kurtosis is equal to 3.

Mesokurtic – distribution with zero kurtosis;

Leptokurtic – distribution with positive kurtosis;

Platykurtic – distribution with negative kurtosis.



Conclusion

- History of probability theory was considered;
- Definition and basic rules for probabilities were presented;
- Key parameters of distribution and methods for theirs estimation were presented.