On the probability and stochastic models

Danijela Rajter-Ćirić

DEPARTMENT OF MATHEMATICS AND INFORMATICS

FACULTY OF SCIENCE

UNIVERSITY OF NOVI SAD, SERBIA

Sample space and events

Suppose that we will perform an experiment whose outcome is not predictable in advance and let us suppose that the set of all possible outcomes is known. This set is called the *sample space* and here will be denoted by Ω. For instance, if the experiment consists of rolling a die, the sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

where the outcome i stands for the case when the number i appear on the die, $i = 1, 2, 3, 4, 5, 6$.

Sample space and events

An event A is a subset of the sample space Ω. In the example with rolling a die, the event $A = \{2, 4, 6\}$ is the event that an even number appears on the roll. Specially, $\emptyset$ refers to the event consisting of no outcomes.

Since events are sets, all usual set operations are allowed such as, union, intersection, taking a subset,...

We usually denote the intersection of events A and B by AB.

The union of mutually exclusive events $A_1, A_2, \ldots$ (that means that $A_i A_j = \emptyset$ when $i \neq j$) is denoted by $\sum_{i=1}^{\infty} A_i$.

Probability of events

Suppose that, for each event A of the sample space Ω, a number $P(A)$ is defined and satisfies the following conditions:

(i) $P(\Omega) = 1$.

(ii) $0 \leq P(A) \leq 1$.

(iii) For any family of mutually exclusive events $A_1, A_2, \ldots$ the following holds:

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The number $P(A)$ is called the probability of the event A.

Conditional probability

The probability that A occurs given that B has occurred is called the *conditional* probability and it is denoted by $P(A|B)$. It is defined by

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

Note that $P(A|B)$ is well defined only if $P(B) > 0$. If both $P(A) > 0$ and $P(B) > 0$, then from

$$P(A|B) = \frac{P(AB)}{P(B)} \text{ and } P(B|A) = \frac{P(AB)}{P(A)}$$

one obtains

$$P(AB) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A).$$

Independent events

Intuitively, two events A and B are independent if the realization of either of this two has no impact on the probability that the other one occurs:

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B).$$

Based on this, one obtains the formal definition: Two events A and B are *independent* if

$$P(AB) = P(A) \cdot P(B).$$

Independent events

Example Suppose we toss two fair coins. What is the probability that on both coins heads appear?

Solution: To solve this problem, denote by A_1 and A_2 the events that head appears on the first and the second coin, respectively. Then, the desired probability is $P(A_1A_2)$ and, since A_1 and A_2 are independent, can be calculated as

$$P(A_1A_2) = P(A_1)P(A_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Random variables

It is very often that in performing some experiment we are more interested in some functions of outcome then in outcome itself. For instance, in tossing dice we are interested in the sum of two dice and rarely in the actual outcome.

These real valued functions defined on the sample space are *random variables*.

We may assign probabilities to the set of possible values of the random variable since those values are determined by the outcomes of the experiment.

Random variables

For instance, suppose that the experiment consists of tossing two fair coins and denote by H the outcome when head appears on whatever coins and by T the outcome when tail appears. If X denote the number of heads appearing, then X is a random variable taking one of the values 0,1,2 with probabilities

$$P\{X=0\} = P\{(T,T)\} = 1/4,$$
$$P\{X=1\} = P\{(H,T)\} + P\{(T,H)\} = 2/4,$$
$$P\{X=2\} = P\{(H,H)\} = 1/4.$$

Of course, $P\{X=0\} + P\{X=1\} + P\{X=2\} = 1.$

Random variables

The *distribution function* F_X of the random variable X is defined for any real number b, $-\infty < b < \infty$, by

$$F_X(b) = P\{X < b\}.$$

It is nondecreasing function, such that

$$\lim_{b\to-\infty} F_X(b) = F_X(-\infty) = 0, \qquad \lim_{b\to\infty} F_X(b) = F_X(\infty) = 1.$$

All probability questions about X can be answered in terms of distribution function. For instance,

$$P(a \le X < b) = F_X(b) - F_X(a), \quad \text{for all } a, b \in \mathbb{R},\ a < b.$$

Random variables

Discrete random variables

1. The binomial random variable
2. The Poisson random variable
3. The geometric random variable

Continuous random variables

1. The uniform random variable
2. The exponential random variable
3. The normal random variable

Discrete random variables

A random variable that can take at most a countable number of possible values is said to be discrete. For a discrete random variable X define

$$p(b) = P\{X = b\}$$

and it is called the *probability mass function* of X. If X must take one of the values x_1, x_2,..., then $p(x_i) > 0$, $i = 1, 2, \ldots$ and $p(x) = 0$, for all other values of x. This can be written as

$$X : \begin{pmatrix} x_1 & x_2 & \ldots & x_n & \ldots \\ p(x_1) & p(x_2) & \ldots & p(x_n) & \ldots \end{pmatrix}.$$

The binomial random variable

Suppose that experiment consists of n independent trials during which an event A occurs with the probability p (hence, it does not occur with probability $q = 1 - p$). If X represents the number of occurring of event A in the n trials, then X is a *binomial* random variable with parameters n and p (we write $X : \mathcal{B}(n; p)$), where $n \in \mathbb{N}$ and $0 < p < 1$. Then

$$X : \begin{pmatrix} 0 & 1 & \dots & k & \dots & n \\ p(0) & p(1) & \dots & p(k) & \dots & p(n) \end{pmatrix},$$

where

$$p(i) = \left(\begin{array}{c} n \\ i \end{array} \right) p^i q^{n-i}, \quad i = 0, 1, \ldots, n.$$

Note that,

$$\sum_{i=0}^{n} p(i) = \sum_{i=0}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) p^i q^{n-i} = (p + (1-p))^n = 1.$$

The Poisson Random variable

The random variable X is said to be *Poisson* random variable with parameter $\lambda > 0$ (we write $X : \mathcal{P}(\lambda)$), if

$$X : \begin{pmatrix} 0 & 1 & \dots & n & \dots \\ p(0) & p(1) & \dots & p(n) & \dots \end{pmatrix},$$

where

$$p(i) = \frac{\lambda^i}{i!}\, e^{-\lambda}, \quad i = 0, 1, \dots, n.$$

Then,

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1.$$

One of important properties of Poisson random variable is that it can be used to approximate a binomial random variable when the binomial parameter n is large and p is small. Namely, if $X : \mathcal{B}(n; p)$ and $n \to \infty$ and $p \to 0$, but $np \to const$, then

$$p(k) \to \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots,$$

where $p(k)$ is probability mass function for binomial random variable.

Example Suppose that the number of typographical errors on a single page of some book has a Poisson distribution with parameter $\lambda = 1$. What is the probability that there is at least one error on a certain page?

Solution:

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1} \approx 0.633.$$

Example If the number of accidents occurring on a highway each day is a Poisson random variable with parameter $\lambda = 3$, what is the probability that no accident occur on a certain day?

Solution: $P\{X = 0\} = e^{-3} \approx 0.05.$

The geometric random variable

Suppose that experiment consists of n independent trials during which an event A occurs with the probability p. If X is the number of trials until the first appearance of A, then X is said to be the *geometric* random variable with parameter $0 < p < 1$ (we write $X : \mathcal{G}(p)$), and

$$X : \begin{pmatrix} 1 & 2 & \dots & n & \dots \\ p(1) & p(2) & \dots & p(n) & \dots \end{pmatrix},$$

where

$$p(i) = p\, q^{i-1}, \quad i = 1, 2, \dots, \quad q = 1 - p.$$

Again, we have

$$\sum_{i=1}^{\infty} p(i) = p \sum_{i=1}^{\infty} q^{i-1} = p \sum_{i=0}^{\infty} q^{i} = p \, \frac{1}{1-q} = \frac{p}{p} = 1.$$

Example Luka goes target shooting and he hits the target with the probability 0.7. If he shoots until he misses the target, what is the probability that he shoots exactly three times?

Solution: If X is the number of shooting, then X is a geometric random variable with $p = 0.3$. Then,

$$P\{X = 3\} = (0.7)^2 \cdot 0.3 = 0.147.$$

Continuous random variables

Continuous random variables have uncountable set of possible values. More precisely, X is said to be continuous if there exists a nonnegative function $\varphi_X : \mathbb{R} \mapsto \mathbb{R}^+$, such that

$$P\{X \in B\} = \int_B \varphi_X(t)\, dt, \quad \text{for any set } B \text{ of real numbers.}$$

The function φ_X is called the *probability density function* of random variable X, or just the density function of X.

By choosing $B = (-\infty, \infty)$ one conclude that φ_X must satisfy

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} \varphi_X(t)\, dt.$$

By choosing $B = [a, b]$ one obtains

$$P\{a < X < b\} = P\{X \in [a, b]\} = \int_{a}^{b} \varphi_X(t)\, dt.$$

If $a = b$ it follows that

$$P\{X = a\} = \int_{a}^{a} \varphi_X(t)\, dt = 0.$$

By choosing $B = (-\infty, b)$, one obtains

$$F_X(b) = \int_{-\infty}^{b} \varphi_X(t)\, dt, \quad \text{for any } b \in \mathbb{R},$$

where F_X is the distribution function of random variable X.

Differentiating both sides gives

$$\frac{d}{db} F_X(b) = \varphi_X(b).$$

The uniform random variable

A continuous random variable X is said to be *uniformly distributed* over the interval (a, b), $a, b \in \mathbb{R}$ and $a < b$, (we write $X : \mathcal{U}(a, b)$), if its density function is

$$\varphi_X(x) = \begin{cases} \dfrac{1}{b-a}, & x \in (a, b) \\ 0, & x \notin (a, b) \end{cases} .$$

For example, if X is the number randomly selected from the interval (a, b), then $X : \mathcal{U}(a, b)$.

The distribution function of $X : \mathcal{U}(a,b)$ is

$$F_X(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ 1, & x > b \end{cases} .$$

Example We select a number from the interval (0,10) by chance. What is the probability that it is less then 3?

Solution: If X is the number we select, then it is uniformly distributed over the interval (0,10). Therefore,

$$P\{X < 3\} = \int_0^3 \frac{dx}{10} = \frac{3}{10}.$$

The normal random variable

A continuous random variable is said to be normally distributed with parameters $m \in \mathbb{R}$ and $\sigma > 0$, (we write $X : \mathcal{N}(m; \sigma^2)$, if its density function is

$$\varphi_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-m)^2}{2\sigma^2}},\ x \in \mathbb{R}.$$

The distribution function of $X : \mathcal{N}(m; \sigma^2)$ is

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt,\ x \in \mathbb{R}.$$

The important property is that, if X is normally distributed with parameters m and σ^2, then $Y = aX + b$ is also normally distributed but with parameters $am + b$ and $a^2\sigma^2 > 0$.

In case when $m = 0$ and $\sigma^2 = 1$ one obtains so-called standard normal distribution, $\mathcal{N}(0; 1)$ which has many applications (for instance, in statistics).

Jointly distributed random variables

We are often interested in probability statements concerning two or more random variables. To deal with such probabilities, we define, for any two random variables X and Y, the *joint cumulative probability distribution function* of X and Y by

$$F_{XY}(a,b) = P\{X < a,\ Y < b\}, \quad -\infty < a, b < \infty.$$

The distribution of X is obtained from the joint distribution of X and Y as follows:

$$F_X(a) = P\{X < a\} = P\{x < a,\ Y < \infty\} = F_{XY}(a, \infty).$$

Similarly, $F_Y(b) = F_{XY}(\infty, b)$.

Independent random variables

The random variables X and Y are said to be *independent* if, for all a, b,

$$P\{X < a,\ Y < b\} = P\{X < a\} \cdot P\{Y < b\}.$$

In other words, X and Y are independent if, for all a and b, the events $\{X < a\}$ and $\{X < b\}$ are independent. In the terms of distribution functions, X and Y are independent if

$$F_{XY}(a, b) = F_X(a) \cdot F_Y(b), \qquad \text{for all } a, b.$$

Expectation and variance

In probability theory, the expectation (expected value, mean) of a random variable is the weighted average of all possible values that this random variable can take on.

The expectation of the random variable X, $E(X)$, is

* $E(X) = \sum_i x_i p(x_i)$, in case X is discrete with the probability mass function $P\{X = x_i\} = p(x_i)$;
* $E(X) = \int_{-\infty}^{\infty} x\ \varphi_X(x)\ dx$, in case X is continuous with the density function $\varphi_X(x)$,

assuming that sum and integral above are absolutely convergent.

Let Y be a function of random variable X, that is $Y = g(X)$. Then

* $E(Y) = E(g(X)) = \sum_i g(x_i)p(x_i)$, in case X is discrete with the probability mass function $p(x_i)$.
* $E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x)\ \varphi_X(x)\ dx$, in case X is continuous with the probability density function $\varphi_X(x)$,

assuming that the sum and the integral above are absolutely convergent.

The variance of the random variable X is defined as

$$D(X) = E\left((X - E(X))^2\right).$$

One easily calculates $D(X) = E(X^2) - E^2(X)$.

Since the variance is always nonnegative, one can define $\sigma(X) = \sqrt{D(X)}$ and it is called the *standard deviation* of the random variable X.

For random variable X having the expectation and the variance we can define

$$X^* = \frac{X - E(X)}{\sqrt{D(X)}}.$$

One can easily show

$$E(X^*) = 0 \text{ and } D(X^*) = 1.$$

Stochastic processes

A stochastic process $\{X(t),\ t \in T\}$ is a collection of random variables.

(It means that, for each $t \in T$, $X(t)$ is a random variable.)

The index t is often interpreted as time. Therefore, we refer to $X(t)$ as the *state* of the process at time t.

For example, $X(t)$ might be the number of customers in the supermarket at time t.

The set T is called the *index* set of the process.

If T is countable set then the stochastic process is said to be a *discrete-time* process.

On the other hand, if T is an interval of the real line, the stochastic process is said to be a *continuous-time* process.

Usually, by $\{X_n,\ n = 0, 1, \ldots\}$ we denote a discrete-time process indexed by nonnegative integers, while $\{X(t),\ t \geq 0\}$ usually denotes a continuous-time stochastic process indexed by nonnegative real numbers.

The state space of a stochastic process is the set of all possible values that the random variables $X(t)$ can assume. Thus, a stochastic process is a family of random variables that describes the evolution through the time of some (physical) process.

Markov chains

Consider a stochastic process $\{X_n,\ n = 0, 1, 2, \ldots\}$ that takes a finite or countable number of possible values. The set of possible values of the process will be denoted by the set of nonnegative integers $\{0, 1, 2, \ldots\}$.

If $X_n = i$, then the process is said to be in state i at time n.

Supposition: Whenever the process is in state i, there is a fixed probability P_{ij} that next it will be in state j:
$P_{ij} = P\{X_{n+1} = j \mid X_n = i\}$. Then we suppose

$$P\{X_{n+1} = j \mid X_n = i,\ X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P_{ij}$$

for all states $i_0,\ i_1, \ldots, i_{n-1},\ i,\ j$ and all $n = 0, 1, 2 \ldots$.

The property in the previous slide may be interpreted as stating that conditional distribution of any future state X_{n+1} given the past states $X_0,\ X_1, \ldots, X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present state!

A process having this property is called Markov chain.

The matrix $P = [P_{ij}]_{ij}$ is called transition probability matrix.

We define n-step transition probability P_{ij}^n to be the probability that a process in state i will be in state j after n additional transforms:

$$P_{ij}^n = P\{X_{k+n} = j \mid X_k = i\},\ \ i, j \geq 0,\ n = 0, 1, 2, \ldots.$$

The Chapmen-Kolmogorov equations provide a method for computing n-step transition probabilities:

$$P_{ij}^{n+m} = \sum P_{ik}^{n} P_{kj}^{m}, \quad \text{for all } n, m = 0, 1, 2, \ldots, \text{ and all } i, j.$$

Explanation: $P_{ik}^{n} P_{kj}^{m}$ represents the probability that starting in i the process will go to the state j in $n + m$ transitions through a path which takes it into state k at the nth transition. Summing over all intermediate states k gives the probability that the process will be in state j after $n + m$ transitions.

If we denote by P_n the n-step transition probability matrix P_{ij}^{n}, $P_n = [P_{ij}^{n}]_{ij}$, then

$$P_{n+m} = P^n P^m, \quad \text{and consequently } P_n = P^n.$$

Examples [Sheldon M. Ross, Probability Models]

Example 1: (Forecasting the weather) Suppose that the chance of rain tomorrow depends on previous weather conditions only through whether or not it is raining today and not on past weather conditions. Suppose also that if it rains today, then it will rain tomorrow with probability α and if does not rain today, then it will rain tomorrow with probability β.

We will define the corresponding Markov chain and its transition probability matrix.

For $\alpha = 0.7$ and $\beta = 0.4$ we will calculate the probability that it will rain four days from today given that it is raining today.

Solution: If we say that the process is in state 0 when it rains and in state 1 when it does nor rain, then one has a two-state Markov chain whose transition probability matrix is given by

$$P = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix}.$$

Specially, for $\alpha = 0.7$ and $\beta = 0.4$ the one-step transition probability matrix is

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}.$$

The two-step transition probability matrix is

$$P_2 = P^2 = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} . \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix} .$$

The four-step transition probability matrix is

$$P_4 = P^4 = (P^2)^2 = \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix} .$$

So, the probability that it will rain four days from today given that it is raining today is $P_{00}^4 = 0.5749$.

Example 2: (A communication system) Consider a communication system which transmits the digits 0 and 1. Each digit transmitted must pass through several stages, at each of which there is a probability p that the digit entered will be unchanged when it leaves. Letting X_n denote the digit entering the nth stage, then $\{X_n,\ n = 1, 2, \ldots\}$ is a two-state Markov chain having a transition probability matrix

$$P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

Example 3: (A random walk model) A Markov chain whose state space is given by the integers $i = 0, \pm 1, \pm 2, \ldots$ is said to be a random walk if, for some number $0 < p < 1$,

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \quad i = 0, \pm 1, \pm 2, \ldots.$$

It is called random walk since we may think of it as being a model for an individual walking on a straight line where one at each point of time either takes one step to the right with the probability p or one step to the left with the probability $1 - p$.

Example 4: (A gambling model) Consider a gambler who, at each play of the game, either wins 1 dollar with probability p or loses 1 dollar with probability $1-p$. If we suppose that our gambler quits playing either when he goes broke or he attains a fortune of N dollars, then the gambler's fortune is a Markov chain having transition probabilities

$$P_{i,i+1} = p = 1 - P_{i,i-1}, \quad i = 1, 2, \ldots, N-1$$
$$P_{00} = P_{NN} = 1.$$

States 0 and N are called absorbing states since once entered they are never left. The process is actually a finite state random walk with absorbing barriers (states 0 and N).

Counting process

A stochastic process $\{N(t),\ t \geq 0\}$ is said to be a counting process if $N(t)$ is the total number of events that occur by time t.

From its definition we see that counting process $N(t)$ must satisfy:

(i) $N(t) \geq 0$.

(ii) $N(t)$ is integer valued.

(iii) If $s < t$, then $N(s) \leq N(t)$.

(iv) For $s < t$, $N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$.

Poisson process

The counting process $\{N(t),\ t \geq 0\}$ is said to be a Poisson process having rate λ, $\lambda > 0$, if

(i) N(0)=0.

(ii) The process has independent increments (the number of events that occur in disjoint time intervals are independent).

(iii) The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all $s, t \geq 0$,

$$P\{N(t+s) - N(s) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \ldots$$

The alternative definition of a Poisson process, equivalent to the previous definition, is:

The counting process $\{N(t),\ t \geq 0\}$ is said to be a Poisson process having rate λ, $\lambda > 0$, if

(i) N(0)=0.

(ii) The process has independent and stationary increments.

(iii) $P\{N(h) = 1\} = \lambda h + o(h)$.

(iv) $P\{N(h) \geq 2\} = o(h)$.

Consider a Poisson process and denote the time of the first event by T_1. Further, for $n > 1$, let T_n denote the elapsed time between the $(n-1)$st and nth event.

The sequence $\{T_n,\ n = 1, 2, \ldots\}$ is called the sequence of interarrival times.

$T_n,\ n = 1, 2, \ldots$ are independent identically distributed exponential random variables having mean $1/\lambda$.

The arrival (waiting) time of the nth event is $S_n = T_1 + \ldots + T_n$ and it has the probability density function

$$\varphi_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \quad t \geq 0.$$

Example: Suppose that people immigrate into a territory at a Poisson rate $\lambda = 1$ per day.

a) What is the expected time until the tenth immigrant arrives?

b) What is the probability that the elapsed time between the tenth and the eleventh arrival exceeds two days?

Solution:

a) $E(S_{10}) = 10/\lambda = 10$ days.

b) $P\{T_{11} > 2\} = e^{-2\lambda} = e^{-2} \approx 0.133.$

Some of the applications of Poisson processes:

- In queueing theory
- In medicine (for example, in tracking the number of HIV infections)
- In actuarial mathematics (insurance theory, ruin theory)
- In reliability theory (for example, estimating software reliability)
- In biology (Birth and death processes...)

Generalizations of the Poisson process

Some generalizations of the Poisson process are

- Nonhomogeneous Poisson process
- Compound Poisson process
- Conditional or mixed Poisson process

Thank you!