A BRIEF REVIEW OF THE PROBABILITY THEORY

MARIA CAMERON

Contents

1. Definitions

A probability space is a triple consisting of *the set of outcomes*, the set of subsets of the set of outcomes that we want to be able to assign probabilities to called *the σ-algebra*, and *the probability measure*, i.e. a function that assigns probabilities.

- A **sample space** Ω is the set of all possible outcomes.
- An **event** A is a subset of Ω.
- A **σ-algebra** $\mathcal{B}$ is a subset of the set of all subsets of Ω that is closed with respect to set operations. The minimal requirements guaranteeing that the σ-algebra possesses these properties constitute the set of axioms that defines it:
  (1) $\emptyset \in \mathcal{B}$ and $\Omega \in \mathcal{B}$;
  (2) If $B \in \mathcal{B}$ then $B^c \in \mathcal{B}$ (B^c is the complement of B in Ω, i.e., $B^c \equiv \Omega \backslash B$).
  (3) If $\mathcal{A} = \{A_1, \dots, A_n, \dots\}$ is a finite or countable collection in $\mathcal{B}$ then
$$\bigcup_i A_i \in \mathcal{B}.$$
  Corollary: If $\mathcal{A} = \{A_1, \dots, A_n, \dots\}$ is a finite or countable collection in $\mathcal{B}$ then
$$\bigcap_i A_i \in \mathcal{B}.$$
  Indeed,
$$\bigcap_i A_i = \left(\bigcup_i A_i^c\right)^c.$$

Example 1 Suppose you are tossing a die. For a single throw, the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. If you are interested in particular number on the top, the natural choice of the σ-algebra is the set of all subsets of Ω. Then $|\mathcal{B}| = 2^6 = 64$. If you are interested only in whether the outcome is odd or even, then a reasonable choice of σ-algebra is

$$\mathcal{B} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}.$$

If you are interested only whether there is an outcome or not, you can choose the coarsest σ-algebra

$$\mathcal{B} = \{\emptyset, \{1, 2, 3, 4, 5, 6\}\}.$$

Example 2 Suppose you are doing a measurement whose outcome can be any real number. For example, you are living in a one-dimensional world, you are throwing a point object, and measuring its position with respect to a fixed point, i.e. the origin of a coordinate system in your 1D-world. The set of outcomes is $\mathbb{R}$. The most commonly chosen σ-algebra is the so-called *Borel σ-algebra* which is generated by all open sets in $\mathbb{R}$. Thanks to the properties of σ-algebra, the Borel σ-algebra can be generated by all intervals of the form $(-\infty, a]$, where $a \in \mathbb{R}$.

- A **probability measure** P is a function $P : \mathcal{B} \to [0, 1]$ such that
  (1) $P(\Omega) = 1$;
  (2) $0 \le P(A) \le 1$ for all $A \in \mathcal{B}$.
  (3) **Countable additivity:** If $\mathcal{A} = \{A_1, \dots, A_n, \dots\}$ is a finite or countable collection in $\mathcal{B}$ such that $A_i \cap A_j = \emptyset$ for any i, j, then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i).$$

  Corollary: $P(\emptyset) = 0$. Indeed,

$$1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset).$$

  Hence, $P(\emptyset) = 0$.
- A **probability space** is the triple $(\Omega, \mathcal{B}, P)$.
- A **random variable** η is a $\mathcal{B}$-measurable function $\eta : \Omega \to \mathbb{R}$.

  A function is called $\mathcal{B}$-measurable if the preimage of any measurable subset of $\mathbb{R}$ is in $\mathcal{B}$. It is proven in analysis that it is suffices to check that $\{\omega \in \Omega \mid \eta(\omega) \le x\} \in \mathcal{B}$ for any $x \in \mathbb{R}$.
- A **probability distribution function** of a random variable η is defined by

$$F_\eta(x) = P(\{\omega \in \Omega \mid \eta(\omega) \le x\}) = P(\eta \le x).$$

  Theorem 1. *If F is a probability distribution function then*
  (1) *F is nondecreasing, i.e. $x < y$ implies $F(x) \le F(y)$.*
  (2) $\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1.$

(3) *$F(x)$ is continuous from the right for every $x \in \mathbb{R}$, i.e.,*

$$\lim_{y \to x+0} F(y) = F(x).$$

Example 3 Suppose you are tossing a die. Consider the probability space

$$(\Omega = \{1, 2, 3, 4, 5, 6\}, \mathcal{B} = 2^{\Omega}, P(\omega) = \tfrac{1}{6}), \tag{1}$$

where 2^{Ω} is the set of all subsets of Ω, and $\omega \in \Omega = \{1, 2, 3, 4, 5, 6\}$. Consider the random variable $\eta(\omega) = \omega$. The probability distribution function is given by

$$F_\eta(x) = \begin{cases} 0, & x < 1, \\ j/6, & j \le x < j+1, \quad j = 1, 2, 3, 4, 5 \\ 1, & x \ge 6. \end{cases}$$

- Suppose $F'_\eta(x)$ exists. Then $f_\eta(x) \equiv F'_\eta(x)$ is called the **probability density function (pdf)** of the random variable η, and

$$P(x < \eta \le x + dx) = F_\eta(x + dx) - F_\eta(x) = f_\eta(x)dx + o(dx).$$

Example 4 The Gaussian density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

where m and σ are constants. m is the mean, while σ is the standard deviation.

Example 5 The density of an exponential random variable with parameter $a > 0$ is given by:

$$f(x) = \begin{cases} ae^{-ax}, & x \ge 0, \\ 0, & x < 0. \end{cases}.$$

Example 6 The density of a uniform random variable on an interval $[a, b]$ is

$$f(x) = \frac{1}{b-a} I_{[a,b]}(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \text{otherwise.} \end{cases}.$$

Here $I_{[a,b]}(x)$ is the indicator function of the interval $[a, b]$.

- If the set of outcomes Ω is discrete (finite or countable) and the σ-algebra is the set of all subsets Ω, then the function $P(\omega)$ is often called **the probability mass function**.

2. Expected values and moments

Definition 1. *Let $(\Omega, \mathcal{B}, P)$ be a probability space, and η be a random variable. Then the expected value, or mean, of the random variable η is defined as*

$$E[\eta] = \int_\Omega \eta(\omega) dP. \tag{2}$$

If Ω is a discrete set,

$$E[\eta] = \sum_i \eta(\omega_i) P(\omega_i).$$

Example 7 Suppose you are tossing a die. Consider the probability space (1) and the random variable $\eta(\omega) = \omega$, $\omega = 1, 2, 3, 4, 5, 6$. The expected value of η is

$$E[\eta] = \sum_{j=1}^{6} j \tfrac{1}{6} = 3.5$$

Suppose that the random variable η is fixed. Then we will omit the subscript in the notation of its probability distribution function: $F_\eta(x) \equiv F(x)$.

The integral in Eq. (2) can be rewritten using $F(x)$:

$$E[\eta] = \int_{\mathbb{R}} x P(x < \eta \le x + dx) = \int_{-\infty}^{\infty} x dF(x).$$

If a derivative $f(x)$ of the probability distribution function F exists, then

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

If g is a function defined on the range of the random variable η (on $\eta(\Omega)$), then the expected value of this function is

$$E[g(\eta)] = \int_{-\infty}^{\infty} g(x) dF(x).$$

Moments: Let us take $g(x) = x^n$.

$$E[\eta^n] = \int_{-\infty}^{\infty} x^n dF(x).$$

Central moments: Let us take $g(x) = (x - E[\eta])^n$.

$$E[(\eta - E[\eta])^n] = \int_{-\infty}^{\infty} (x - E[\eta])^n dF(x).$$

Variance = 2nd central moment:

$$\mathrm{Var}(\eta) = E[(\eta - E[\eta])^2) = \int_{-\infty}^{\infty} (x - E[\eta])^2 dF(x).$$

Example 8 Suppose you are tossing a die. Consider the probability space (1) and the random variable $\eta(\omega) = \omega$, $\omega = 1, 2, 3, 4, 5, 6$. The variance of η is

$$\mathrm{Var}(\eta) = \frac{1}{6}\sum_{j=1}^{6}(j-3.5)^2 = \frac{35}{12} = 2.91(6).$$

The standard deviation:

$$\sigma(\eta) = \sqrt{\mathrm{Var}(\eta)}.$$

3. Independence, joint distributions, covariance

- Two events $A, B \in \mathcal{B}$ are **independent** if

$$P(A \cap B) = P(A)P(B).$$

- Two random variables η_1 and η_2 are independent if the events

$$\{\omega \in \Omega \mid \eta_1(\omega) \le x\} \quad \text{and} \quad \{\omega \in \Omega \mid \eta_2(\omega) \le y\} \tag{3}$$

are independent for all $x, y \in \mathbb{R}$.

Example 9 Suppose you are tossing a die twice. Consider the probability space

$$\left(\Omega = \{1,2,3,4,5,6\}^2, \mathcal{B} = 2^{\Omega^2}, P(\{\omega_1, \omega_2\}) = {}^1\!/_{36}\right), \quad 1 \le \omega_1, \omega_2 \le 6. \tag{4}$$

Let η_1 and η_2 be random variables equal to the outcomes of the first and

Table 1. Two throws of a die. Values of the random variables $\xi(\omega_1, \omega_2) = \omega_1 + \omega_2$ (left) and $\beta(\omega_1, \omega_2) = \omega_1 - \omega_2$ (right).

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	-1	0	1	2	3	4
3	-2	-1	0	1	2	3
4	-3	-2	-1	0	1	2
5	-4	-3	-2	-1	0	1
6	-5	-4	-3	-2	-1	0

the second throws respectively. These random variables are independent. Now consider the random variables $\eta(\omega_1, \omega_2) = \omega_1$ and $\xi(\omega_1, \omega_2) = \omega_1 + \omega_2$ (see Table 1, left). We can show that η and ξ are dependent by taking e.g., $x = 1$ and $y = 2$ in Eq. (3):

$$P(\eta \le 1 \;\&\; \xi \le 2) = \tfrac{1}{36} \ne P(\eta \le 1)P(\xi \le 2) = \tfrac{1}{6} \cdot \tfrac{1}{36} = \tfrac{1}{216}.$$

Finally, we consider the random variables $\xi(\omega_1, \omega_2) = \omega_1 + \omega_2$ and $\beta(\omega_1, \omega_2) = \omega_1 - \omega_2$ (see Table 1, right). We can show that they are dependent by taking e.g., $x = 2$ and $y = -1$ in Eq. (3):

$$P(\xi \le 2 \;\&\; \beta \le -1) = 0 \neq P(\xi \le 2)P(\beta \le -1) = \tfrac{1}{36} \cdot \tfrac{15}{36} = \tfrac{5}{432}.$$

- The joint distribution function of two random variables η_1 and η_2 is given by

$$F_{\eta_1\eta_2}(x, y) = P\left(\{\omega \in \Omega \mid \eta_1(\omega) \le x,\ \eta_2(\omega) \le y\}\right) = P\left(\eta_1(\omega) \le x,\ \eta_2(\omega) \le y\right).$$

- If the second mixed derivative of $F_{\eta_1\eta_2}$ exists, it is called the **joint probability density** of η_1 and η_2 and denoted by

$$f_{\eta_1\eta_2}(x, y) := \frac{\partial F_{\eta_1\eta_2}(x, y)}{\partial x \partial y}.$$

In this case,

$$F_{\eta_1,\eta_2}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{\eta_1\eta_2}(x, y) dx dy.$$

Exercise Show that two random variables are independent if and only if

$$F_{\eta_1\eta_2}(x, y) = F_{\eta_1}(x)F_{\eta_2}(y).$$

Furthermore, if the joint pdf $f_{\eta_1\eta_2}(x, y)$ exists, then η_1 and η_2 are independent iff

$$f_{\eta_1\eta_2}(x, y) = f_{\eta_1}(x)f_{\eta_2}(y).$$

- Given the joint pdf $f_{\eta_1\eta_2}$, one can obtain $f_{\eta_1}(x)$ by

$$f_{\eta_1}(x) = \int_{-\infty}^{\infty} f_{\eta_1\eta_2}(x, y) dy.$$

In this equation, f_{η_1} is called a **marginal** of $f_{\eta_1\eta_2}$, and the variable η_2 is **integrated out**.

- **Properties of expected value and variance** It follows from the definition, that the expected value is a linear functional:

$$E[a\eta_1 + b\eta_2] = aE[\eta_1] + bE[\eta_2]. \tag{5}$$

-

$$\mathrm{Var}(a\eta) = a^2\mathrm{Var}(\eta). \tag{6}$$

- If η_1 and η_2 are independent, then

$$\mathrm{Var}(\eta_1 + \eta_2) = \mathrm{Var}(\eta_1) + \mathrm{Var}(\eta_2). \tag{7}$$

If η_1 and η_2 are dependent, (7) is not true: take $\eta_1 = \eta_2$. In general,

$$\mathrm{Var}(\eta_1 + \eta_2) = \mathrm{Var}(\eta_1) + \mathrm{Var}(\eta_2) + 2\mathrm{Cov}(\eta_1, \eta_2), \tag{8}$$

where $\mathrm{Cov}(\eta_1, \eta_2)$ is the covariance of η_1 and η_2 – see below. You will see below that (7) does not imply that η_1 and η_2 are independent, only that they are uncorrelated.

Example 10 Suppose you are tossing a die twice. Consider the probability space and random variables introduced in Example 9. Then

$$E[\xi] = E[\eta_1 + \eta_2] = E[\eta_1] + E[\eta_2] = 7.$$

$$E[\beta] = E[\eta_1 - \eta_2] = E[\eta_1] + E[-\eta_2] = 0.$$

$$\mathrm{Var}[\xi] = \mathrm{Var}[\eta_1 + \eta_2] = \mathrm{Var}[\eta_1] + \mathrm{Var}[\eta_2] = \tfrac{35}{6} = 5.8(3).$$

$$\mathrm{Var}[\beta] = \mathrm{Var}[\eta_1 - \eta_2] = \mathrm{Var}[\eta_1] + \mathrm{Var}[-\eta_2] = \mathrm{Var}[\eta_1] + \mathrm{Var}[\eta_2] = \tfrac{35}{6} = 5.8(3).$$

Example 11 Consider the Bernoulli random variable

$$\eta = \begin{cases} 1, & P(1) = p, \\ 0, & P(0) = 1 - p. \end{cases} \tag{9}$$

Its expected value and variance are

$$E[\eta] = 1 \cdot p + 0 \cdot (1 - p) = p,$$

$$\mathrm{Var}(\eta) = (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) = p(1 - p).$$

Now consider the sum of n independent copies of η:

$$\xi := \sum_{i=1}^{n} \eta_i.$$

Using Eq. (5) we calculate $E[\xi]$:

$$E[\xi] = \sum_{\imath=1}^{n} E[\eta_i] = np.$$

Since η_i, $1 \le i \le n$, are independent, we can calculate $\mathrm{Var}(\xi)$ using Eq. (7):

$$\mathrm{Var}(\xi) = \sum_{i=1}^{n} \mathrm{Var}(\eta_i) = np(1 - p).$$

Finally, consider the average of n independent copies of η:

$$\zeta := \frac{1}{n} \sum_{i=1}^{n} \eta_i \equiv \frac{\xi}{n}.$$

Using Eqs. (5) and (6), we find

$$E[\zeta] = p,$$

$$\mathrm{Var}(\zeta) = \mathrm{Var}\left(\frac{\xi}{n}\right) = \frac{1}{n^2}\mathrm{Var}(\xi) = \frac{p(1 - p)}{n}.$$

- The **covariance** of two random variables η_1 and η_2 is defined by

$$\mathrm{Cov}(\eta_1, \eta_2) = E[(\eta_1 - E[\eta_1])(\eta_2 - E[\eta_2])].$$

Remark If η_1 and η_2 are independent, then $\mathrm{Cov}(\eta_1, \eta_2) = 0$. If $\mathrm{Cov}(\eta_1, \eta_2) = 0$ then η_1 and η_2 are uncorrelated. Note that uncorrelated random variables are not necessarily independent.

Example 12 Suppose you are tossing a die twice. Consider the probability space and random variables introduced in Example 9. As we have established in Example 9, ξ and β are dependent. However, they are uncorrelated. Indeed,

$$\mathrm{Cov}(\xi, \beta) = \sum_{1\le\omega_1\le6,\ 1\le\omega_2\le6} (\omega_1 + \omega_2 - 7)(\omega_1 - \omega_2)P(\{\omega_1, \omega_2\})$$

$$= \frac{1}{36}\left(\sum_{\omega_1<\omega_2} (\omega_1 + \omega_2 - 7)(\omega_1 - \omega_2) + \sum_{\omega_1>\omega_2} (\omega_1 + \omega_2 - 7)(\omega_1 - \omega_2)\right) = 0.$$

Example 13 A vector-valued random variable $\eta = [\eta_1, \dots, \eta_n]$ is jointly Gaussian if

$$P(x_1 < \eta_1 \le x_1 + dx_1, \dots, x_n < \eta_n \le x_n + dx_n) = \frac{1}{Z}e^{-\frac{1}{2}(x-m)^\top A^{-1}(x-m)}dx + o(dx),$$

where $x = [x_1, \dots, x_n]^\top$, $m = [m_1, \dots, m_n]^\top$, $dx = dx_1 \dots dx_n$, and A is a symmetric positive definite matrix. The normalization constant Z is given by

$$Z = (2\pi)^{n/2}|A|^{1/2}, \quad \text{where} \quad |A| = \det A.$$

In the case of jointly Gaussian random variables, the covariance matrix C whose entries are

$$C_{ij} = E[(\eta_i - E[\eta_i])(\eta_j - E[\eta_j])]$$

is equal to A. Two jointly Gaussian random variables are independent if and only if they are uncorrelated.

4. Chebyshev's inequality

Chebyshev's inequality holds for any random variable. It is a very useful theoretical tool for proving various estimates. In practice, it often gives too rough estimates which is a consequence of its universality. Chebyshev's inequality is not improvable, as we can construct a random variable for which it turns into an equality.

Theorem 2. *Let η be a random variable. Suppose $g(x)$ is a nonnegative, nondecreasing function (i.e., $g(x) \ge 0$, $g(a) \le g(b)$ whenever $a < b$). Then for any $a \in \mathbb{R}$*

$$P(\eta \ge a) \le \frac{E[g(\eta)]}{g(a)}. \tag{10}$$

Proof.

$$E[g(\eta)] = \int_{-\infty}^{\infty} g(x)dF(x)$$
$$\ge \int_a^{\infty} g(x)dF(x) \ge g(a)\int_a^{\infty} dF(x) = g(a)P(\eta \ge a).$$

□

Given a random variable η we define a random variable

$$\xi := |\eta - E[\eta]|.$$

Define

$$g(x) = \begin{cases} x^2, & x \geq 0, \\ 0, & x < 0. \end{cases}.$$

Plugging this into Eq. (10) we obtain

$$P(|\eta - E[\eta]| \geq a) \leq \frac{\mathrm{Var}(\eta)}{a^2}.$$

Example 14 Suppose you are tossing a die twice. Consider the probability space and random variables introduced in Example 9. We will compare the exact probabilities with their Chebyshev estimates.

$$P(|\xi - 7| \geq 1) = P(\xi \neq 7) = 1 - \tfrac{6}{36} = \tfrac{5}{6} = 0.8(3), \quad \frac{\mathrm{Var}(\xi)}{1} = \tfrac{35}{6} = 5.8(3);$$

$$P(|\xi - 7| \geq 2) = P(\xi \leq 5 \text{ or } \xi \geq 9) = \tfrac{20}{36} = \tfrac{5}{9} = 0.(5), \quad \frac{\mathrm{Var}(\xi)}{4} = \tfrac{35}{24} = 1.458(3);$$

$$P(|\xi - 7| \geq 3) = P(\xi \leq 4 \text{ or } \xi \geq 10) = \tfrac{12}{36} = \tfrac{1}{3} = 0.(3), \quad \frac{\mathrm{Var}(\xi)}{9} = \tfrac{35}{54} = 0.6(481);$$

$$P(|\xi - 7| \geq 4) = P(\xi \in \{2, 3, 11, 12\}) = \tfrac{6}{36} = \tfrac{1}{6} = 0.1(6), \quad \frac{\mathrm{Var}(\xi)}{16} = \tfrac{35}{96} = 0.36458(3);$$

$$P(|\xi - 7| \geq 5) = P(\xi \in \{2, 12\}) = \tfrac{2}{36} = \tfrac{1}{18} = 0.0(5), \quad \frac{\mathrm{Var}(\xi)}{25} = \tfrac{35}{150} = 0.2(3);$$

Choosing $a = k\sigma$ we get

$$P(|\eta - E[\eta]| \geq k\sigma) \leq \frac{1}{k^2}.$$

This means that for *any* random variable η defined on *any* probability space we have that the probability that η deviates from its expected value by at least k standard deviations does not exceed $1/k^2$.

The bounds given Chebyshev's inequality cannot be improved in principle, because they are exact for the random variable

$$\eta = \begin{cases} 1, & P = \frac{1}{2k^2}, \\ 0, & P = 1 - \frac{1}{k^2}, \\ -1, & P = \frac{1}{2k^2}. \end{cases}.$$

It is easy to check that $E[\eta] = 0$, $\mathrm{Var}(\eta) = \frac{1}{k^2}$. Hence

$$P(|\eta| \geq 1) = \tfrac{1}{k^2} = \frac{\mathrm{Var}(\eta)}{1^2},$$

i.e. Chebyshev's inequality turns into equality.

5. Types of convergence of random variables

Suppose we have a sequence of random variables $\{\eta_1, \eta_2, \ldots\}$. In probability theory, there exist several different notions of convergence of a sequence of random variables $\{\eta_1, \eta_2, \ldots\}$ to some limit random variable η.

- $\{\eta_1, \eta_2, \ldots\}$ **converges in distribution** or **converges weakly**, or **converges in law** to η if

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{for every} \quad x \quad \text{where} \quad F(x) \quad \text{is continuous,} \tag{11}$$

  where F_n and F are the probability distribution functions of η_n and η respectively.

  Remark Convergence of pdfs $f_n(x)$ implies convergence of $F_n(x)$. The converse is not true in general. For example, consider $F_n(x) = x - \frac{1}{2\pi n}\sin(2\pi n x)$, $x \in (0, 1)$. The corresponding pdf is $f_n(x) = 1 - \cos(2\pi n x)$, $x \in (0, 1)$. $\{F_n(x)\}$ converges to $F(x) = x$, i.e., to the uniform distribution, while $\{f_n(x)\}$ does not converge at all.

Remark In the discrete case, the convergence of probability mass functions $f(k) := P(\eta = k)$ implies the convergence of the probability distribution functions.

Example 15 Consider the sum of n independent copies of the Bernoulli random variable as in Example 11:

$$\xi = \sum_{i=1}^{n} \eta_i, \quad \text{where} \quad \eta_i = \begin{cases} 1, & P(1) = p, \\ 0, & P(0) = 1 - p. \end{cases} \tag{12}$$

Its probability distribution is the binomial distribution given by

$$f(k; n, p) \equiv P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}, \tag{13}$$

where $\binom{n}{k}$ is the number of k-combinations of the set of n elements:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Now we let $n \to \infty$ and $p \to 0$ in such a manner that the product np (i.e., the expected value of ξ) remains constant. We introduce the parameter

$$\lambda := np.$$

Consider the sequence of random variables ξ_n where ξ_n is the sum of n independent copies of Bernoulli random variable with $p = \lambda/n$, i.e,

$$\xi_n = \sum_{i=1}^{n} \eta_i^{(n)}, \quad \text{where} \quad \eta_i^{(n)} = \begin{cases} 1, & P(1) = \lambda/n, \\ 0, & P(0) = 1 - \lambda/n. \end{cases} \tag{14}$$

Plugging in $p = \lambda/n$ in the results of Example 11 we find the expected value and the variance:

$$E[\xi_n] = n\frac{\lambda}{n} = \lambda.$$

$$\text{Var}(\xi_n) = n\frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right) = \lambda\left(1 - \frac{\lambda}{n}\right).$$

We will show that the sequence ξ_n converges to the Poisson random variable with parameter λ in distribution. Consider the limit

$$\lim_{n\to\infty} f\left(k; n, \frac{\lambda}{n}\right) = \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} =$$

$$\frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{n^k} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k}$$

The first limit in the equation above is 1 as $n(n-1)\dots(n-k+1) = n^k + O(n^{k-1})$. The second limit can be calculated using the well-known fact that

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e.$$

Hence

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

The third limit is 1. Therefore,

$$\lim_{n\to\infty}\frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda},$$

which is the Poisson distribution with parameter λ.

- $\{\eta_1, \eta_2, \dots\}$ **converges in probability** to η if for any $\epsilon > 0$

(15) $$\lim_{n\to\infty} P(|\eta_n - \eta| \geq \epsilon) = 0$$

Remark Convergence in probability implies convergence in distribution.

Proof. We will prove this fact for the case of scalar random variables. We have $\lim_{n\to\infty} P(|\eta_n - \eta| \geq \epsilon) = 0$, we need to prove $\lim_{n\to\infty} P(\eta_n \leq x) = P(\eta \leq x)$ for every x where F_η is continuous. First we show an auxiliary fact that for any two random variables ξ and ζ, $x \in \mathbb{R}$ and $\epsilon > 0$

(16) $$P(\xi \leq a) \leq P(\zeta \leq a+\epsilon) + P(|\xi - \zeta| > \epsilon).$$

Indeed,

$$\begin{aligned}P(\xi \leq a) &= P(\xi \leq a \;\&\; \zeta \leq a+\epsilon) + P(\xi \leq a \;\&\; \zeta > a+\epsilon)\\ &\leq P(\zeta \leq a+\epsilon) + P(\xi - \zeta \leq a - \zeta \;\&\; a - \zeta < -\epsilon)\\ &\leq P(\zeta \leq a+\epsilon) + P(\zeta - \xi < -\epsilon)\\ &\leq P(\zeta \leq a+\epsilon) + P(\zeta - \xi < -\epsilon) + P(\zeta - \xi > \epsilon)\\ &= P(\zeta \leq a+\epsilon) + P(|\zeta - \xi| > \epsilon).\end{aligned}$$

Applying Eq. (16) to $\xi = \eta_n$ and $\zeta = \eta$ with $a = x$ and $a = x - \epsilon$, we get

$$\begin{aligned}P(\eta_n \leq x) &\leq P(\eta \leq x+\epsilon) + P(|\eta_n - \eta| > \epsilon)\\ P(\eta \leq x-\epsilon) &\leq P(\eta_n \leq x) + P(|\eta_n - \eta| > \epsilon).\end{aligned}$$

$$P(\eta \leq x-\epsilon) - P(|\eta_n - \eta| > \epsilon) \leq P(\eta_n \leq x) \leq P(\eta \leq x+\epsilon) + P(|\eta_n - \eta| > \epsilon).$$

Taking the limit $n \to \infty$ and taking into account that $\lim_{i\to\infty} P(|\eta_n - \eta| \geq \epsilon) = 0$, we get

$$F_\eta(x-\epsilon) \leq \lim_{n\to\infty} F_{\eta_n}(x) \leq F_\eta(x+\epsilon).$$

If x is a point of continuity of F_η,

$$\lim_{\epsilon\to 0} F_\eta(x-\epsilon) = \lim_{\epsilon\to 0} F_\eta(x+\epsilon) = F_\eta(x).$$

Therefore, taking the limit $\epsilon \to 0$ we obtain the weak convergence:

$$\lim_{n\to\infty} F_{\eta_n}(x) = F_\eta(x)$$

for any x where $F_\eta(x)$ is continuous. □

Remark The converse is, generally, not true. However, convergence in distribution to a *constant* random variable implies convergence in probability.

- $\{\eta_1, \eta_2, \ldots\}$ **converges almost surely** or **almost everywhere** or **with probability 1** or **strongly** to η if

(17) $$P\left(\lim_{n\to\infty} \eta_n = \eta\right) = 1.$$

  Remark Convergence almost surely implies convergence in probability (by Fatou's lemma) and in distribution.

- To summarize,

(18) $$\boxed{\eta_i \to \eta \text{ almost surely}} \Rightarrow \boxed{\eta_i \to \eta \text{ in probability}} \Rightarrow \boxed{\eta_i \to \eta \text{ in distribution}}$$

6. Laws of Large Numbers and the Central Limit Theorem

- Let $\{\eta_1, \eta_2, \ldots\}$ be a sequence of random variables with finite expected values $\{m_1 = E[\eta_1], m_2 = E[\eta_2], \ldots\}$. Define

$$\xi_n = \frac{1}{n}\sum_{i=1}^{n} \eta_i, \quad \bar{\xi}_n = \frac{1}{n}\sum_{i=1}^{n} m_i.$$

Definition 2. (1) *The sequence of random variables η_n satisfies the Law of Large Numbers if $\xi_n - \bar{\xi}_n$ converges to zero in probability, i.e., for any $\epsilon > 0$*

$$\lim_{n\to\infty} P(|\xi_n - \bar{\xi}_n| > \epsilon) = 0.$$

(2) *The sequence of random variables η_n satisfies the Strong Law of Large Numbers if $\xi_n - \bar{\xi}_n$ converges to zero almost surely, i.e., for almost all $\omega \in \Omega$*

$$\lim_{n\to\infty} \xi_n - \bar{\xi}_n = 0.$$

- If the random variables η_n are independent and if $\mathrm{Var}(\eta_i) \le V < \infty$, then the Law of Large Numbers holds by the Chebyshev Inequality (10):

$$P(|\xi_n - \bar{\xi}_n| > \epsilon) = P\left(\left|\sum_{i=1}^{n} \eta_i - \sum_{i=1}^{n} m_i\right| > n\epsilon\right)$$
$$\le \frac{\mathrm{Var}(\eta_1 + \ldots + \eta_n)}{\epsilon^2 n^2} \le \frac{V}{\epsilon^2 n} \to 0 \quad \text{as} \quad n \to \infty.$$

- **Theorem 3.** *(Khinchin) A sequence of independent identically distributed random variables $\{\eta_i\}$ with $\mathbb{E}[\eta_i] = m$ and $\mathbb{E}[|\eta_i|] < \infty$ satisfies the Law of Large Numbers.*

- **Theorem 4.** *(Kolmogorov) A sequence of independent identically distributed random variables with finite expected value and variance satisfies the Strong Law of Large Numbers.*

- **Theorem 5. (The central limit theorem)** *Let $\{\eta_1, \eta_2, \ldots\}$ be a sequence of independent identically distributed (i.i.d.) random variables with $m = E[\eta_i]$ and $0 < \sigma^2 = \mathrm{Var}(\eta_i) < \infty$, then*

$$\frac{(\sum_{i=1}^n \eta_i) - nm}{\sigma\sqrt{n}} \longrightarrow N(0,1) \quad \text{in distribution,} \tag{19}$$

i.e., converges weakly to the standard normal distribution $N(0,1)$ (i.e., the Gaussian distribution with mean 0 and variance 1) as $n \to \infty$.

A proof via Fourier transform can be found in [1]. Another proof making use of characteristic functions can be found in [2].

Remark Eq. (19) can be recasted as

$$\frac{1}{n}\sum_{i=1}^n \eta_i \longrightarrow N\left(m, \frac{\sigma^2}{n}\right) \quad \text{in distribution,} \tag{20}$$

i.e., the average of the first n i.i.d. random variables η_i converges in distribution to the Gaussian random variable with mean $m = E[\eta_i]$ and variance σ^2/n.

7. Conditional probability and conditional expectation

- The conditional probability of an event B given that the event A has happened is given by

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

Note that if A and B are independent, then $P(A \cap B) = P(A)P(B)$ and hence

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

Example 16 Suppose you are tossing a die twice. Consider the probability space (4). Let A be the event that the outcome of the first throw is even, and B be the event that the sum of the outcomes is ≥ 10. Then (see Table 1)

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{4/36}{1/2} = \frac{2}{9}.$$

Note that $P(B) = 1/6 < P(B|A)$. Hence the events A and B are dependent.

If the event A is fixed, then $P(B|A)$ defines a probability measure on $(\Omega, \mathcal{B})$.

- If η is a random variable on Ω, then conditional expectation of η given the event A is

$$E[\eta|A] = \int_\Omega \eta(\omega)P(d\omega|A) = \int_\Omega \eta(\omega)\frac{P(d\omega \cap A)}{P(A)} = \frac{\int_A \eta(\omega)P(dw)}{P(A)}.$$

Example 17 . Suppose you are tossing a die twice. Consider the probability space (4). Let A be the event that the outcome of the first throw is even, and η be the random variable whose value is the sum of outcomes, i.e., $\eta(\{\omega_1, \omega_2\}) = \omega_1 + \omega_2$. Then

$$E[\eta|A] = \sum_{\omega_1=1}^{6} \sum_{\omega_2=1}^{6} (\omega_1 + \omega_2) P(\{\omega_1, \omega_2\} \mid \omega_1 \in \{2,4,6\}).$$

Let us calculate $P(\{\omega_1, \omega_2\} \mid \omega_1 \in \{2,4,6\})$.

$$P(\{\omega_1, \omega_2\} \mid \omega_1 \in \{2,4,6\}) = \frac{P(\{\omega_1, \omega_2\} \cap (\omega_1 \in \{2,4,6\}))}{P(\omega_1 \in \{2,4,6\})}$$

$$= \begin{cases} 0, & \omega_1 \in \{1,3,5\}, \\ \frac{P(\{\omega_1,\omega_2\})}{P(\omega_1 \in \{2,4,6\})} = \frac{1/36}{1/2} = \frac{1}{18}, & \omega_1 \in \{2,4,6\}. \end{cases}$$

Now we continue our calculation:

$$E[\omega_1 + \omega_2 \mid \omega_1 \in \{2,4,6\}] = \sum_{\omega_1 \in \{2,4,6\}} \sum_{\omega_2=1}^{6} (\omega_1 + \omega_2) \frac{1}{18} = \frac{135}{18} = 7.5.$$

Note that $E[\eta] = 7 \neq E[\eta|A] = 7.5$.

- Now we show how one can construct new random variables using conditional probability. For simplicity, we start with partitioning the set of outcomes Ω into a finite or countable number of disjoint measurable subsets:

$$\Omega = \bigcup_i A_i, \quad \text{where} \quad A_i \in \mathcal{B}, \quad A_i \cap A_j = \emptyset.$$

Definition 3. *Let η be a random variable on the probability space $(\Omega, \mathcal{B}, P)$. Let $\mathcal{A} = \{A_i\}$ be a partition of Ω as above. Define a new random variable $E[\eta|\mathcal{A}]$ as follows:*

$$E[\eta|\mathcal{A}] = \sum_i E[\eta|A_i]\chi(A_i), \tag{21}$$

where $\chi(A_i)$ is the indicator function of A_i:

$$\chi(A_i;\omega) = \begin{cases} 1, & \omega \in A_i, \\ 0, & \omega \notin A_i. \end{cases}$$

Remark Note that $E[\eta|\mathcal{A}]$ is a random variable as it is a function of the outcome ω. Indeed,

$$E[\eta|\mathcal{A}](\omega) = E[\eta|A_i] \text{ where } A_i \ni \omega.$$

Example 18 Suppose you are tossing a die twice. Let us partition the set of outcomes as follows:

$$\Omega = \bigcup_{i=1}^{6} \{(\omega_1, \omega_2) \mid \omega_1 = i\}.$$

The corresponding partition $\mathcal{A}$ is

$$\mathcal{A} = \{\{(\omega_1, \omega_2) \mid \omega_1 = i\}\}_{i=1}^{6}.$$

Take the random variable $\xi = \omega_1 + \omega_2$ (see Table 1, left), the sum of numbers on the top. Construct a new random variable

$$E[\xi|\mathcal{A}] = \sum_{i=1}^{6} E[\xi|\omega_1 = i]\chi(\omega_1 = i) = \sum_{i=1}^{6} (i + 3.5)\chi(\omega_1 = i)$$
$$= 4.5\chi(\omega_1 = 1) + 5.5\chi(\omega_1 = 2) + 6.5\chi(\omega_1 = 3)$$
$$+ 7.5\chi(\omega_1 = 4) + 8.5\chi(\omega_1 = 5) + 9.5\chi(\omega_1 = 6).$$

- Now we define the conditional expectation of one random variable η given the other random variable θ. First we assume that θ assumes a finite or countable number of values $\{\theta_1, \theta_2, \ldots\}$. Define the partition $\mathcal{A}$ where

$$A_i = \{\omega \in \Omega \mid \theta = \theta_i\}.$$

Definition 4. *We define a new random variable $E[\eta|\theta]$ as a the following function of the random variable θ:*

$$E[\eta|\theta] := E[\eta|\mathcal{A}], \quad \text{i.e.,} \quad E[\eta|\theta] = E[\eta|A_i] \ \text{ if } \ \theta = \theta_i.$$

Example 19 Suppose you are tossing a die twice. Let (ω_1, ω_2) be the numbers on the top. Define random variables $\xi = \omega_1 + \omega_2$ and $\theta = \omega_1$. Then it follows from our calculation from the previous example that

$$E[\xi|\theta] = 3.5 + \theta.$$

References

[1] A. Chorin and O. Hald, Stochastic Tools in Mathematics and Science, 3rd edition, Springer 2013
[2] L. Koralov and Ya. Sinai, Theory of probability and stochastic processes, 2nd edition, Springer, 2007