A Fundamentals of Stochastics

Theorem A.1 (Axioms of Probability)**.** *The following axioms of probability can be defined:*

(1) The probability $\mathrm{P}(A)$ *of an outcome* A *in an experiment is a uniquely determined non-negative real number, which can be at most equal to 1, thus it holds that*

$$0 \leq \mathrm{P}(A) \leq 1 \ . \tag{A.1}$$

(2) For a certain event A *of an experiment it holds that*

$$\mathrm{P}(A) = 1 \ . \tag{A.2}$$

For equivalent events B *and* C *in an experiment it holds that*

$$\mathrm{P}(B) = \mathrm{P}(C) \ . \tag{A.3}$$

(3) If two events B *and* C *in an experiment are mutually exclusive, then it holds that*

$$\mathrm{P}(B + C) = \mathrm{P}(B) + \mathrm{P}(C) \ . \tag{A.4}$$

Definition A.1 (Random Variable, Stochastic Variable)**.** A function X is called a random variable or stochastic variable if it is assigned to a random experiment and possesses the following properties:

(1) The values of X are real numbers and

(2) for every number a and every interval I on the number line, the probability of the event "X has the value a" or "X lies in the interval I" is in accordance with the axioms of probability.

In this context, a random experiment is an experiment in which the result of a single execution can be expressed by a single number.

If X is a *discrete random variable*, then the corresponding probabilities $p_1 = \mathrm{P}(X = x_1)$, $p_2 = \mathrm{P}(X = x_2), \ldots$ can be assigned to all values of X, hereinafter denoted by $x_1, x_2, \ldots$. The function

$$f(x) = \begin{cases} p_j & \text{for } x = x_j \\ 0 & \text{for all other } x \end{cases} \tag{A.5}$$

is then called the *probability function*. Since the random variable X always takes a value

x_j, it must also hold that the sum of all probabilities

$$\sum_j f(x_j) = 1 \tag{A.6}$$

is 1. The probability that the random variable X lies in the interval $a < X \leq b$ is then easily calculated in the form

$$\mathrm{P}(a < X \leq b) = \sum_{a<x_j\leq b} f(x_j) \ . \tag{A.7}$$

If the probability $\mathrm{P}(X \leq x)$ is plotted as a function of x, i.e., the probability of an experiment whose results are $X \leq x$, then one obtains the *probability distribution function*

$$F(x) = \mathrm{P}(X \leq x) = \sum_{x_j\leq x} f(x_j) \ . \tag{A.8}$$

For a *continuous random variable* X, the probability distribution function can be represented in integral form

$$F(x) = \int_{-\infty}^{x} f(v)\,\mathrm{d}v \quad \text{with} \quad F(\infty) = 1 \tag{A.9}$$

where the integrand $f(v)$ is a non-negative and, except for finitely many points, continuous function, which is also called the *probability density function*.

Exercise A.1. Show that for the continuous random variable X the relation

$$\mathrm{P}(a < X \leq b) = \int_a^b f(v)\,\mathrm{d}v = F(b) - F(a)$$

holds.

Figure A.1 shows the typical course of the probability distribution function of a discrete and a continuous random variable.

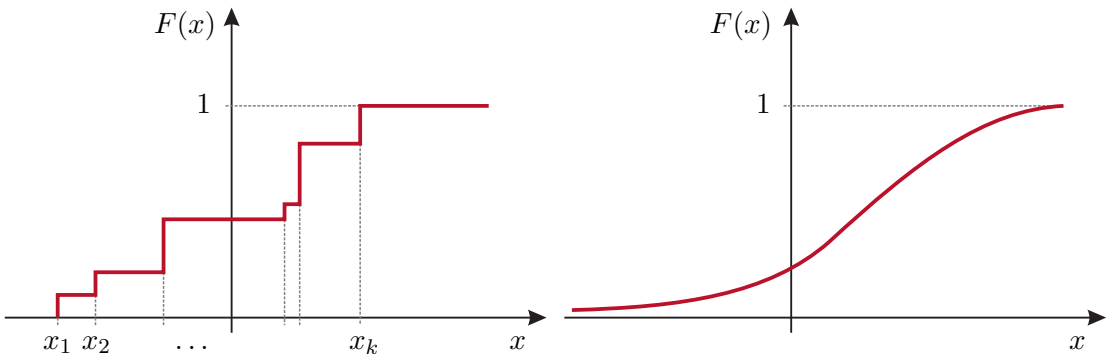


Figure A.1: On the probability distribution function.

Definition A.2 (Expected Value)**.** The *expected value* $\mathrm{E}(X)$ of a random variable X (also called mean value or 1st moment) is defined in the case of a discrete distribution as

$$\mathrm{E}(X) = \sum_j x_j f(x_j) \tag{A.10}$$

and in the case of a continuous distribution as

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x\ . \tag{A.11}$$

For a function $g(X)$ of the random variable X, it generally holds that the expected value of the function $g(X)$ can be calculated in the form

$$\mathrm{E}(g(X)) = \sum_j g(x_j) f(x_j) \quad \text{or} \quad \mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x \tag{A.12}$$

Exercise A.2. Show the validity of the following important equation

$$\mathrm{E}(\alpha g(X) + \beta h(X)) = \alpha\,\mathrm{E}(g(X)) + \beta\,\mathrm{E}(h(X))$$

with the constants α and β. Show that this also implies

$$\mathrm{E}(\mathrm{E}(X)) = \mathrm{E}(X) \quad \text{and} \quad \mathrm{E}(X\,\mathrm{E}(X)) = \mathrm{E}(X)^2$$

Definition A.3 (Variance)**.** The *variance* σ_X^2 of a random variable X measures the quadratic deviation from the expected value and is defined as

$$\sigma_X^2 = \mathrm{E}\Big([X - \mathrm{E}(X)]^2\Big) = \mathrm{E}\Big(X^2 - 2X\,\mathrm{E}(X) + \mathrm{E}(X)^2\Big) = \mathrm{E}\Big(X^2\Big) - \mathrm{E}(X)^2 \tag{A.13}$$

The variance is also referred to as the second central moment, and its positive square root σ_X as the standard deviation.

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

The previous considerations can now be easily extended to *multiple random variables*. For two discrete random variables X and Y, the probability distribution function takes the form

$$F(x,y) = \mathrm{P}(X \leq x, Y \leq y) = \sum_{x_k \leq x} \sum_{y_j \leq y} f(x_k, y_j) \tag{A.14}$$

with

$$f(x,y) = \begin{cases} p_{kj} & \text{for } x = x_k,\ y = y_j \\ 0 & \text{for all other } (x,y) \end{cases} \quad \text{with} \quad \sum_k \sum_j f(x_k, y_j) = 1 \tag{A.15}$$

and in the continuous case it holds that

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(v,w)\,\mathrm{d}v\,\mathrm{d}w \quad \text{with} \quad F(\infty,\infty) = 1\ . \tag{A.16}$$

The so-called *marginal distributions* are given by the following expressions

$$F_1(x) = \mathrm{P}(X \leq x, Y \text{ arbitrary}) = \sum_{x_k \leq x} \underbrace{\sum_j f(x_k, y_j)}_{f_1(x_k)} \tag{A.17}$$

or

$$F_1(x) = \mathrm{P}(X \leq x, Y \text{ arbitrary}) = \int_{-\infty}^{x} \underbrace{\int_{-\infty}^{\infty} f(v,w)\,\mathrm{d}v}_{f_1(w)}\,\mathrm{d}w \tag{A.18}$$

and

$$F_2(y) = \mathrm{P}(X \text{ arbitrary}, Y \leq y) = \sum_{y_j \leq y} \underbrace{\sum_k f(x_k, y_j)}_{f_2(y_j)} \tag{A.19}$$

or

$$F_2(y) = \mathrm{P}(X \text{ arbitrary}, Y \leq y) = \int_{-\infty}^{y} \underbrace{\int_{-\infty}^{\infty} f(v,w)\,\mathrm{d}w}_{f_2(v)}\,\mathrm{d}v \tag{A.20}$$

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

Definition A.4 (Independence of Random Variables)**.** Two random variables X and Y are independent if and only if for every pair of events of the form

$$a_1 < X \leq b_1 \quad \text{and} \quad a_2 < Y \leq b_2 \tag{A.21}$$

the relation

$$\mathrm{P}(a_1 < X \leq b_1, a_2 < Y \leq b_2) = \mathrm{P}(a_1 < X \leq b_1)\,\mathrm{P}(a_2 < Y \leq b_2) \tag{A.22}$$

and thus also

$$f(x,y) = f_1(x) f_2(y) \quad \text{and} \quad F(x,y) = F_1(x) F_2(y) \tag{A.23}$$

with $f_1(x)$, $f_2(y)$, $F_1(x)$ and $F_2(y)$ according to (A.17)–(A.20) holds.

Theorem A.2 (Expected Value and Covariance of Two Random Variables)**.** *For the expected value of a function $g(X,Y)$ of two random variables X and Y, it holds that*

$$\mathrm{E}(g(X,Y)) = \sum_k \sum_j g(x_k, y_j) f(x_k, y_j) \tag{A.24}$$

or

$$\mathrm{E}(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\,\mathrm{d}x\,\mathrm{d}y \ . \tag{A.25}$$

The expected value $\mathrm{E}(X+Y)$ of a sum of random variables X and Y is calculated as

$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y) \ . \tag{A.26}$$

If two random variables X and Y are independent, then the following relation holds (shown here for the discrete case using (A.19))

$$\mathrm{E}(XY) = \sum_k \sum_j x_k y_j f(x_k, y_j) = \sum_k \sum_j x_k y_j f_1(x_k) f_2(y_j) = \mathrm{E}(X)\,\mathrm{E}(Y) \ . \tag{A.27}$$

If a random variable Z results from the sum of two random variables X and Y, then according to (A.13) the variance of Z is

$$\begin{aligned} \sigma_Z^2 &= \mathrm{E}\left(Z^2\right) - \mathrm{E}(Z)^2 = \mathrm{E}\left(X^2\right) + 2\,\mathrm{E}(XY) + \mathrm{E}\left(Y^2\right) \\ &\quad - \mathrm{E}(X)^2 - 2\,\mathrm{E}(X)\,\mathrm{E}(Y) - \mathrm{E}(Y)^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY} \end{aligned} \tag{A.28}$$

with the so-called covariance

$$\sigma_{XY} = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y) \ . \tag{A.29}$$

It is easily verified that (A.29) *can also be written in the form*

$$\sigma_{XY} = \mathrm{E}([X - \mathrm{E}(X)][Y - \mathrm{E}(Y)]) \tag{A.30}$$

If the random variables X and Y are independent, then according to (A.27) $\sigma_{XY} = 0$.

Definition A.5 (Correlation Coefficient and Covariance Matrix)**.** The quotient

$$r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{A.31}$$

is called the *correlation coefficient* between the random variables X and Y. If $r = 0$, then X and Y are *uncorrelated*. It can also be seen that two independent random variables X and Y must also be uncorrelated because of $\sigma_{XY} = 0$. The correlation coefficient $-1 \leq r \leq 1$ provides a measure of the *linear dependence* of X and Y. For a vector-valued random variable $\mathbf{X}^{\mathrm{T}} = \begin{bmatrix} X_1 & X_2 & \dots & X_n \end{bmatrix}$, it now holds that

$$\mathrm{E}\left(\mathbf{X}^{\mathrm{T}}\right) = \begin{bmatrix} \mathrm{E}(X_1) & \mathrm{E}(X_2) & \dots & \mathrm{E}(X_n) \end{bmatrix} \tag{A.32}$$

and the *covariance matrix* of the vector-valued random variable $\mathbf{X}$ is understood to be the matrix

$$\mathrm{cov}(\mathbf{X}) = \mathrm{E}\left([\mathbf{X} - \mathrm{E}(\mathbf{X})][\mathbf{X} - \mathrm{E}(\mathbf{X})]^{\mathrm{T}}\right) = \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1 X_2} & \cdots & \sigma_{X_1 X_n} \\ \sigma_{X_2 X_1} & \sigma_{X_2}^2 & \cdots & \sigma_{X_2 X_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{X_n X_1} & \sigma_{X_n X_2} & \cdots & \sigma_{X_n}^2 \end{bmatrix} . \tag{A.33}$$

It can be seen that the covariance matrix is symmetric.

Exercise A.3. Show that the relation

$$\mathrm{E}\left(\|\mathbf{X} - \mathrm{E}(\mathbf{X})\|_2^2\right) = \mathrm{E}\left([\mathbf{X} - \mathrm{E}(\mathbf{X})]^{\mathrm{T}}[\mathbf{X} - \mathrm{E}(\mathbf{X})]\right) = \mathrm{trace}(\mathrm{cov}(\mathbf{X}))$$

holds with $\mathrm{trace}(\mathbf{S}) = \sum_i s_{ii}$.

From what has been said so far, it follows that the covariance matrix of a vector-valued random variable $\mathbf{X}$, whose components X_i, X_j for $i \neq j = 1, \dots, n$ are uncorrelated, is a diagonal matrix.

For stochastic time signals originating from statistically identical signal sources, there exists not only a single realization $x_1(t)$ but a whole family (ensemble) of random time functions $\{x_j(t)\}$. This ensemble of random time functions or random sequences in the time-discrete case is called a *continuous-time* or *discrete-time stochastic process* $\mathrm{x}(t)$. A single realization $x_j(t)$ for a fixed j is also called a *sample function*. For every fixed point in time t, $\mathrm{x}(t)$ is a random variable with a probability distribution function (see (A.8))

$$F(x, t) = \mathrm{P}(\mathrm{x}(t) \leq x) \ . \tag{A.34}$$

The probability density function is then calculated according to (A.9) as

$$f(x,t) = \frac{\partial F(x,t)}{\partial x} \ . \tag{A.35}$$

Definition A.6 (Mean, Auto- and Cross-Correlation Function of a Stochastic Process)**.** The mean $\eta_{\mathrm{x}}(t)$ of a stochastic process $\mathrm{x}(t)$ is (see also (A.11))

$$\eta_{\mathrm{x}}(t) = \mathrm{E}(\mathrm{x}(t)) = \int_{-\infty}^{\infty} x f(x,t)\,\mathrm{d}x \ . \tag{A.36}$$

The *autocorrelation function* $\Phi_{\mathrm{xx}}(t_1,t_2)$ of a stochastic process $\mathrm{x}(t)$ is understood to be the expected value of the product $\mathrm{x}(t_1)\mathrm{x}(t_2)$

$$\Phi_{\mathrm{xx}}(t_1,t_2) = \mathrm{E}(\mathrm{x}(t_1)\mathrm{x}(t_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1,t_1) x_2 f(x_2,t_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \ . \tag{A.37}$$

The *cross-correlation function* $\Phi_{\mathrm{xy}}(t_1,t_2)$ of two stochastic processes $\mathrm{x}(t)$ and $\mathrm{y}(t)$ is given by the relation

$$\Phi_{\mathrm{xy}}(t_1,t_2) = \mathrm{E}(\mathrm{x}(t_1)\mathrm{y}(t_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,t_1) y f(y,t_2)\,\mathrm{d}x\,\mathrm{d}y. \tag{A.38}$$

Definition A.7 (Auto- and Cross-Covariance Function)**.** The *autocovariance function* $C_{\mathrm{xx}}(t_1,t_2)$ of a stochastic process $\mathrm{x}(t)$ is

$$C_{\mathrm{xx}}(t_1,t_2) = \mathrm{E}([\mathrm{x}(t_1)-\eta_{\mathrm{x}}(t_1)][\mathrm{x}(t_2)-\eta_{\mathrm{x}}(t_2)]) = \Phi_{\mathrm{xx}}(t_1,t_2) - \eta_{\mathrm{x}}(t_1)\eta_{\mathrm{x}}(t_2) \ . \tag{A.39}$$

The *cross-covariance function* $C_{\mathrm{xy}}(t_1,t_2)$ of two stochastic processes $\mathrm{x}(t)$ and $\mathrm{y}(t)$ is analogously given by

$$C_{\mathrm{xy}}(t_1,t_2) = \mathrm{E}([\mathrm{x}(t_1)-\eta_{\mathrm{x}}(t_1)][\mathrm{y}(t_2)-\eta_{\mathrm{y}}(t_2)]) = \Phi_{\mathrm{xy}}(t_1,t_2) - \eta_{\mathrm{x}}(t_1)\eta_{\mathrm{y}}(t_2) \ . \tag{A.40}$$

Exercise A.4. Show the validity of the right-hand identity of (A.39).

Definition A.8 (Stationary Stochastic Process)**.** A stochastic process $\mathrm{x}(t)$ is called *strictly stationary* if its statistical properties are invariant under time shifts, i.e., $\mathrm{x}(t)$ and $\mathrm{x}(t+c)$ have identical statistical properties for all c. The stochastic process $\mathrm{x}(t)$ is *stationary in a wider sense* if the mean $\eta_{\mathrm{x}}(t) = \eta_{\mathrm{x}}$ is constant and the autocorrelation function depends only on the time difference $\tau = t_2 - t_1$, i.e., $\Phi_{\mathrm{xx}}(\tau) = \mathrm{E}(\mathrm{x}(t)\mathrm{x}(t+\tau))$.

The functions (A.36)–(A.40) shown so far take into account all realizations, i.e., the entire ensemble, of a stochastic process. According to the so-called *ergodic hypothesis*, the functions (A.36)–(A.40) can also be written using only a single sample function if *infinitely long time intervals* are considered. Ergodic processes are also stationary, but the inverse does not generally hold. This results in the following expressions, with the continuous-time case on the left and the discrete-time case on the right:

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

(1) Mean value

$$\eta_x = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\,\mathrm{d}t \quad \text{or} \quad \eta_x = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} x_k \tag{A.41}$$

(2) Autocorrelation function

$$\Phi_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)\,\mathrm{d}t \quad \text{or} \quad \Phi_{xx}(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} x_k x_{k+\tau} \tag{A.42}$$

(3) Cross-correlation function

$$\Phi_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t+\tau)\,\mathrm{d}t \quad \text{or} \quad \Phi_{xy}(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} x_k y_{k+\tau} \tag{A.43}$$

Exercise A.5. Show that $\Phi_{xx}(\tau)$ satisfies the following properties:

1. $\Phi_{xx}(\tau) = \Phi_{xx}(-\tau)$
2. $\Phi_{xx}(0) = \mathrm{E}\left(x(t)^2\right)$
3. $\Phi_{xx}(\infty) = \mathrm{E}(x(t))^2$
4. $\Phi_{xx}(\tau) \leq \Phi_{xx}(0)$

Exercise A.6. Show that $\Phi_{xy}(\tau)$ satisfies the following properties:

1. $\Phi_{xy}(\tau) = \Phi_{yx}(-\tau)$
2. $\Phi_{xy}(0) = \mathrm{E}(x(t)y(t))$
3. $\Phi_{xy}(\infty) = \mathrm{E}(x(t))\,\mathrm{E}(y(t))$
4. $\Phi_{xy}(\tau) \leq \frac{1}{2}[\Phi_{xx}(0) + \Phi_{yy}(0)]$

Exercise A.7. What are the expressions for the autovariance and autocovariance functions in the ergodic case?

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

Definition A.9 (White Noise)**.** A stochastic process $\mathrm{v}(t)$ is called *(strict) white noise* if for all time points $t_1 \neq t_2$, $\mathrm{v}(t_1)$ and $\mathrm{v}(t_2)$ are statistically independent, or

$$C_{\mathrm{vv}}(\tau) = C_0\delta(\tau) \quad \text{with} \quad \delta(\tau) = \begin{cases} 1 & \text{for } \tau = 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and } C_0 > 0\ . \tag{A.44}$$

It is further assumed that the mean $\eta_{\mathrm{v}} = 0$.

Remark: It can be shown that in the continuous-time case the average power of the white noise is infinite and therefore this process cannot be ideally realized. In contrast, in the discrete-time case, where the signal sequence values are at a finite distance from each other, the power remains finite, so that ideal white noise is also realizable.

©W. Kemmetmüller, A. Kugi, Automation and Control Institute, TU Wien

A.1 References

[A.1] R. Isermann, *Identifikation dynamischer Systeme 1 und 2*, 2nd ed. Berlin, Deutschland: Springer, 1992.

[A.2] E. Kreyszig, *Statistische Methoden und ihre Anwendungen*, 7th ed. Gttingen, Deutschland: Vandenhoeck & Ruprecht, 1998.

[A.3] D. Luenberger, *Optimization by Vector Space Methods*. New York, USA: John Wiley & Sons, 1969.

[A.4] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, USA: McGraw-Hill, 2002.