Review of Basic Concepts of Probability

Contents

Review of basic concepts of probability

The notation used in probability theory is closely related to Set Theory in Mathematics.

Preliminary definitions

- **Experiment (ε)**: observation of a physical phenomenon. Of each REALIZATION or ESSAY of an experiment, an OUTCOME is obtained.
- **Sample Space (Ω)**: complete set of all the possible outcomes of an experiment.
- **Event (A)**: set of outcomes of an experiment (a subset of the sample space).
- **Certain event**: An event which is sure to occur at every performance of an experiment is called a certain event connected with the experiment ('head or tail' is a certain event when tossing a coin).
- **Impossible event**: An event which cannot occur at any performance of the experiment is called an impossible event (`seven' in case of throwing a die).
- **Inclusion (or implication)** ($A \subset B$). A is included in B (or A implies B) iff the occurrence of event A produces the occurrence of event B.
- **Union of events A and B** ($A \cup B$, or $A+B$): The union of events A and B holds true iff A or B holds true. It is not necessary that all events must hold true.
- **Intersection (or product)** ($A \cap B$, or AB): The intersection of events A and B can only be true iff both events hold true.
- **Complementary event**: An event B is complementary of A, if it is verified if and only if (iff) A is not verified ('head' and 'tail' are complementary to each other in the experiment of tossing a coin).
- **Mutually Exclusive events** (or incompatible): A and B are mutually exclusive iff they are not verified simultaneously ($A \cap B = \emptyset$).
- **Compound event**: A is a compound event iff:

$$\exists B, C \neq \emptyset, \text{ and } B, C \neq A / B, C \subset \text{ or } A = B \cup C$$

Axiomatic definition of probability

Probability is a measure of uncertainty. Once a random experiment is defined, we call probability of the event A the real number P(A)∈ [0,1]. The function P(·):Ω→ [0,1] is called probability measure or probability distribution and must satisfy the following three axioms:

- P(A) ⩾0; the probability is always a non-negative value.
- P(Ω) =1; the probability of the certain event is one (the maximum value).
- ∀ A, B / A ∩ B = ∅ ⇒P(A∪B) = P(A) + P(B); if two events are complementary, the probability of the union event equals the sum of probabilities.

The axiomatic definition of probability allows us to construct a robust theory of probability. To better understand the concept of probability we rely on the `Frequentist definition':

- Frequentist definition of probability: The probability P(A) of an event A is the limit:

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

- Where N is the number of observations and N_A is the number of times that event A occurred.

Conditional probability and Bayes' theorem

Given two events A and B, and considering P(B)≥0;, the conditional probability is defined with the following expression:

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

This definition fulfills the three axioms of the axiomatic definition of probability.

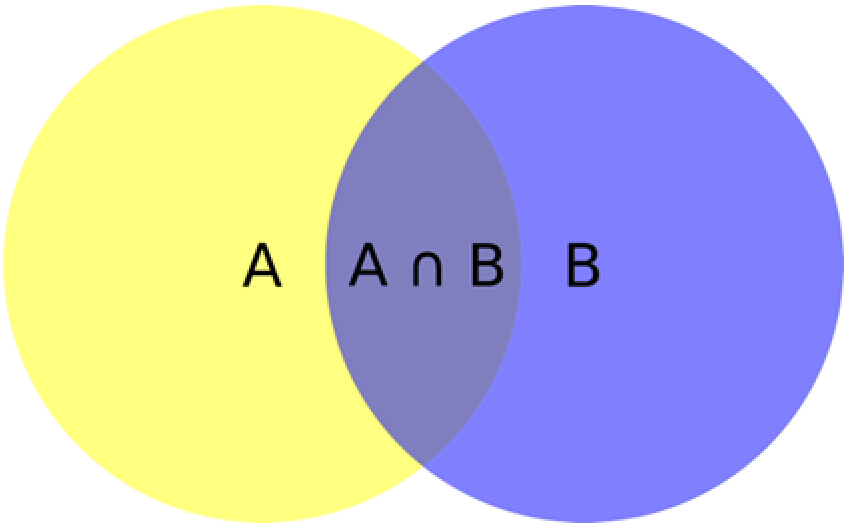

Figure 1 Representation of two events with intersection

P(A/B) can be understood as the probability of the intersection event (A∩B) when event B occurs.

Bayes' Theorem: The conditional probabilities P(A/B) and P(B/A), can be related with the Bayes Theorem.

$$P\left({}^{B}/_{A} \right)P(A) = P\left({}^{A}/_{B} \right)P(B)$$

Partition of the sample space, and total probability

The set of events $\{A_1, A_2, \cdots, A_n\}$ is a Partition of the Sample Space Ω iff:

a. $A_i \cap A_j = \emptyset, \forall i \neq j;$ all the events in the partition are mutually exclusive.
b. $\cup_{i=1}^{n} A_i = \Omega$
c. $P(A_i) > 0, \forall i$

Given a partition $\{A_1, A_2, \cdots, A_n\}$, it is possible to calculate the probability of any event B defined in the sample space Ω, applying the Total Probability theorem, defined with the following expression:

$$P(B) = \sum_{i=1}^{n} P(B/A_i)P(A_i)$$

And applying the Bayes theorem: $P(A/B) = \dfrac{P\left({}^{B}/_{A} \right)P(A)}{\sum_{i=1}^{n} P\left({}^{B}/_{A_i} \right)P(A_i)}$

Example: A transmission system can send one of the following letters in binary code: A=100, B=110, C=111. The prior probability of transmitting each letter is: P(A)=0.4; P(B)=0.4; P(C)=0.2.

a. Randomly select one symbol of the transmitted signal. Calculate the probability of this symbol to be 1.

Answer:

$$P(1) = P(1/A)P(A) + P(1/B)P(B) + P(1/C)P(C) = 0.6$$

b. Suppose the selected symbol is 1. Calculate the most likely transmitted letter.

Answer:

$$P(A/1) = \frac{P(1/A)P(A)}{P(1/A)P(A)+P(1/B)P(B)+P(1/C)P(C)} = \frac{2}{9}$$

$$P(B/1) = \frac{P(1/B)P(B)}{P(1/A)P(A)+P(1/B)P(B)+P(1/C)P(C)} = \frac{4}{9}$$

$$P(C/1) = \frac{P(1/C)P(C)}{P(1/A)P(A)+P(1/B)P(B)+P(1/C)P(C)} = \frac{1}{3}$$

Therefore, the most likely transmitted letter is B.

3 6 Partition of Sample Space, Bayes Formula with Example

For further explanations about these concepts: [3 6 Partition of Sample Space, Bayes Formula with Example]

Independency

Two events A and B are independent iff:

a. $P(A \cap B) = P(A)P(B)$
b. $P(A/B) = P(A)$
c. $P(B/A) = P(B)$

Random variables

A random variable is defined by assigning a number to each result of a random experiment.

Definition: Given a random experiment which generates a sample space, defined as {Ω, ε, P(·)}, a random variable x is the result of mapping $\Omega \rightarrow \mathbb{R}$, which assigns a real number x to every outcome ω, with the following conditions:

1. The set $X \leq \chi$ is an event for every x.
2. The probabilities $P(x = \infty) = 0$ and $P(x = -\infty) = 0$.

Cumulative distribution function

Random variables are usually characterized with the Cumulative Distribution Function (or just Distribution Function), and the Probability Density Function, which are related.

The Cumulative Distribution Function is defined with the following expression

$$F_x(x_0) \triangleq P(X \leq x_0)$$

Properties of the distribution function:

1. $F_x(-\infty) = 0$, and $F_x(\infty) = 1$
2. It is a monotonically increasing function, i.e. $0 \leq F_x(x_0) \leq 1$
3. $P(X > x_0) = 1 - F_x(x_0)$
4. $P(\alpha < X \leq b) = F_x(b) - F_x(\alpha)$

Example: The distribution function of the random variable built assigning numbers to the results of the experiment `throwing a die' is the following (note that the random variable is discrete):

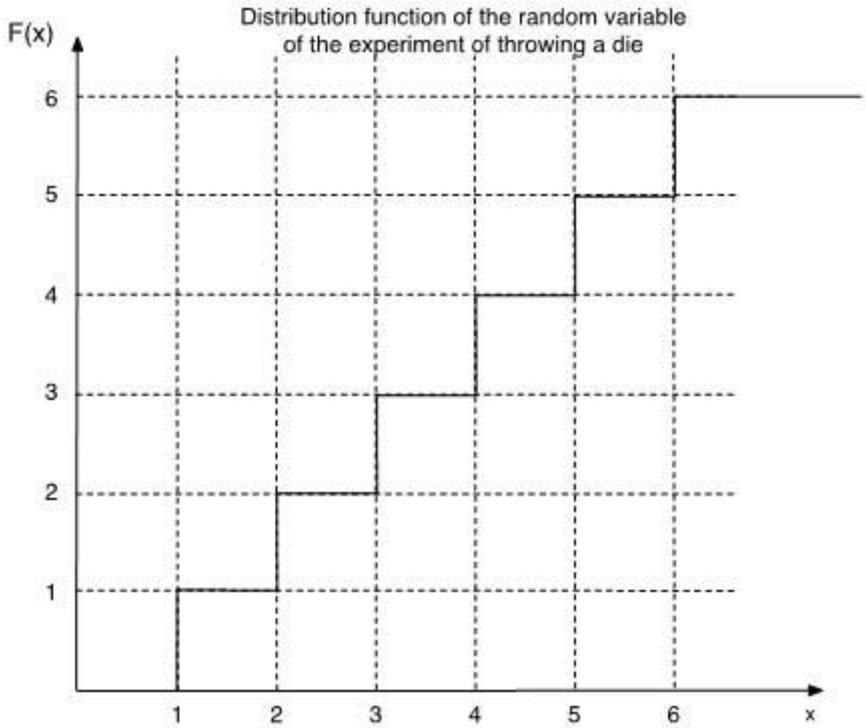

Figure 2 Cumulative Distribution Function of the random variable of the `throwing a die' experiment © University of Alcalá

Probability density function

The Probability Density Function (PDF) gives information about how likely a value of a random variable is. It is not directly the probability but the probability per unit length. For a given random variable X we define the PDF $p_x(x)$ as follows:

$$p_X(x) = \lim_{\Delta \to 0} \frac{P(x < X < x + \Delta)}{\Delta} = \lim_{\Delta \to 0} = \frac{F_X(x+\Delta) - F_X(x)}{\Delta}$$

$$p_X(x) = \frac{dF_X(x)}{dx}$$

For step discontinuities in the CDF, the Dirac delta function is used to describe the PDF.

Properties:

1. $P(x \leq x_0) = F_X(x_0) = \int_{-\infty}^{x_0} p_X(x)dx$
2. $p_X(x) \geq 0, \forall x$
3. $\int_{-\infty}^{+\infty} p_X(x)dx = 1$
4. $P(a < x < b) = \int_a^b p_x(x)dx$

The PDF of discrete random variables accumulate the probability in a finite and numerable number of values, which are the possible values. Its PDF can be expressed using the Dirac delta function:

$$p_X(x) = \sum_{i-1}^{N} P(X = x_i)\delta(x - x_i)$$

Expectation, and moments of a random variable

Suppose a random variable X, from which N samples are known ($X_1, X_2, X_3, \ldots, X_n$). Expectation is defined with the following expression:

$$E[x] = \lim_{N \to \infty} \frac{X_1 + X_2 + X_3 + X_N}{N} = \int_{-\infty}^{\infty} x p_X(x)dx$$

Moments

For any positive integer r, the r-th moment of the random variable characterized by the PDF $p_x(X)$ are given by the following expression:

$$\mu_r = E[x^r] = \int_{-\infty}^{\infty} x^r p_X(x)dx$$

For discrete random variables, the moments are calculated as:

$$\mu_r = E[x^r] = \sum_{i=1}^{n} x_t^r P(X = x_i)$$

The first moment defines the **mean** of the random variable (μ), while the second, the mean squared value.

Moments can be defined about a given number, for example, the mean. Moments about the mean are known as central moments, and calculated as follows:

$$\mu_{0r} = E\left[(x-\mu)^r\right] = \int_{-\infty}^{\infty}(x-\mu)^r p_X(x)dx$$

For discrete random variables, the central moments are calculated as follows:

$$\mu_{0r} = \sum_{i=1}^{n}(x_i-\mu)^r P(X = x_i)$$

The **Variance** σ_{x^2} is a measure of dispersion around the mean value. It is the central moment for r=2.

$$\sigma_x^2 = E\left[(x - \mu)^2\right] = \int_{-\infty}^{\infty}(x-\mu)^2 p_X(x)dx$$

In the case of discrete random variables, it is calculated as follows:

$$\sigma_x^2 = \sum_{n=-1}^{N}(x_i - \mu)^2 P(X = x_i)$$

The Gaussian probability density function

A continuous scalar random variable x is said to be *normally distributed* with parameters μ and σ^2 and expressed as $x\text{-}N(\mu, \sigma^2)$, if its probability density function is given by:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

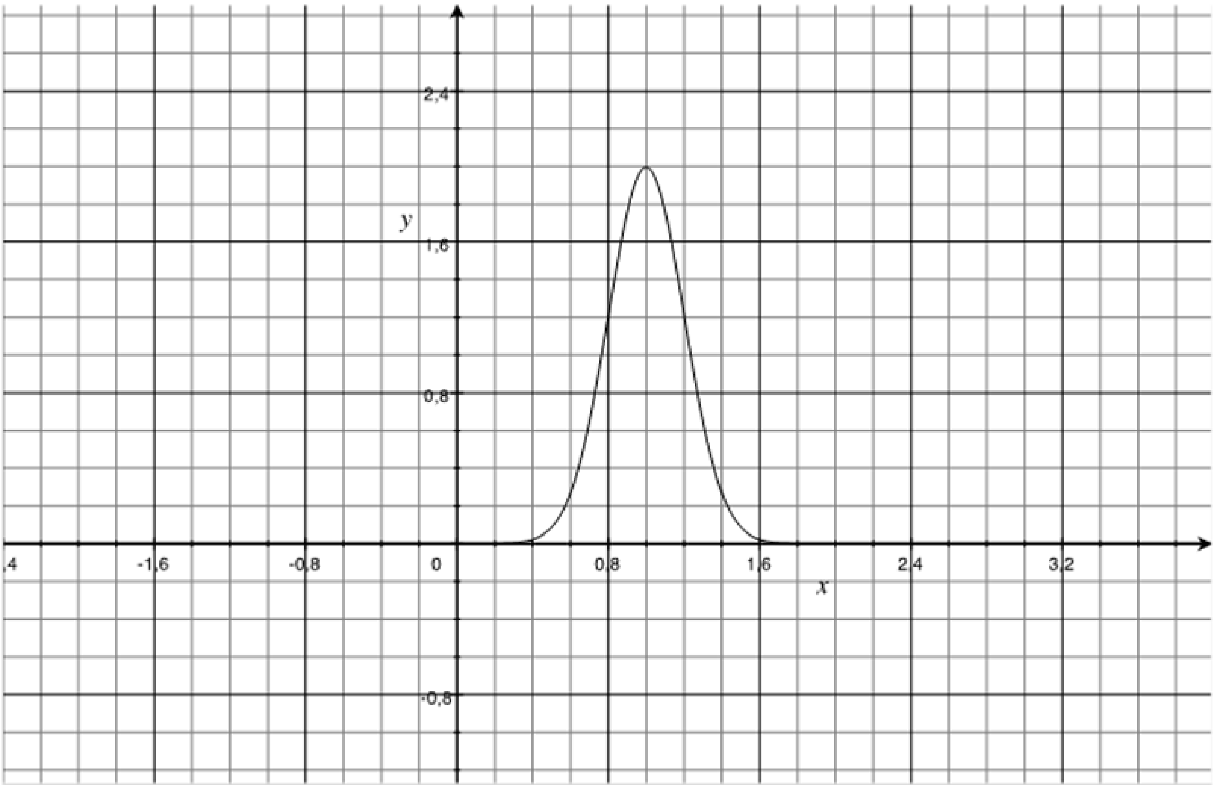

Figure 3 Gaussian PDF with μ=1 and σ^2 = 0.2

Note that the probability density function depends only on two parameters, the mean and the variance. If we define a new random variable with this transformation $z=(x-\mu)/\sigma$, a new normal or Gaussian random variable with zero mean and unity variance is obtained. It is called standard normal.

- **Calculation of probabilities with the Normal or Gaussian PDF**

Probabilities must be calculated by integrating the PDF. The integral of a Normal function only can be solved numerically, as a primitive function cannot be obtained. This calculation is quite common in communications.

The `complementary error function', or Function Q, is defined as follows:

$$Q(A) = \frac{1}{\sqrt{2\pi}} \int_A^{\infty} e^{-\frac{x^2}{2}} dx$$

The values of Q(x) have been calculated numerically and can be found in tables in many books. The calculation of probabilities of normal random variables is usually carried out using this function. The Q(x) function is monotonically decreasing. Some features are:

$$Q(-\infty) = 1;\ Q(0) = \frac{1}{2};\ Q(\infty) = 0;\ Q(-x) = 1 - Q(x)$$

Example: Given a standard normal random variable, calculate the probability of the random variable to be smaller than or equal to one.

$$P(x \leq 1) = \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 - \int_{1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 - Q(1)$$

Functions of random variables

We are interested in the study of functions where the independent variable is random. There are random phenomena obtained by processing the result of a random experiment, defining functions of random variables, where the applied processing defines the function.

The result obtained with a function of random variable is random as well. The objective is to characterize statistically the new random variable, from the known characterization of the original random variable: we know the PDF or the CDF of random variable X, and we would like to know the PDF of CDF of the new random variable Y=g(X).

Cumulative distribution function of Y=g(X)

According to the definition of CDF, $F_y(y) = P(Y \leq y) = P(g(X) \leq y)$

- Example: $y = ax + b$, with $a,\ b\ \in \mathbb{R}$

The procedure to find the CDF consists in finding the range of x which fulfills $ax + b \leq y$;

$$F_y(y) = P\left\{x \leq \frac{y-b}{a}\right\} = F_x\left(\frac{y-b}{a}\right),\ a > 0$$

$$F_y(y) = P\left\{x \geq \frac{y-b}{a}\right\} = 1 - F_x\left(\frac{y-b}{a}\right),\ a < 0$$

Probability density function of Y=g(X)

To find the PDF of the new random variable, the `Fundamental Theorem' is applied, as follows:

- Find the roots of equation y=g(x), obtaining the set x_n
- Find the desired PDF applying the following expression:

$$p_y(y) = \sum_n \frac{p_x(x_n)}{|g^1(x_n)|}$$

Expectation of a function of random variable

The mean value of Y=g(X) is obtained with the following expression:

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x)p_x(x)dx$$

The main application of the functions of random variables is the generation of random variables with a desired PDF from random variables with well known PDFs, such as Gaussians or Uniform distributed variables.

Joint characterization of two random variables

Two random variables can be jointly characterized using the 'Joint Probability Density Function' or the 'Joint Cumulative Distribution Function', defined as follows:

$$F_{xy}(x,\ y) = P(X \leq x, Y \leq y)$$

$$p_{xy}(x,\ y) = \frac{\partial F_{xy}(x,y)}{\partial x \partial y}$$

If the two variables are independent, the joint PDF is the product of marginal probability density functions:

$$p_{xy}(x,\ y) = p_x(x)p_y(y)$$

This reasoning can be extended to sets of `n' random variables, concluding that the joint PDF of a set of random variables is the product of the marginal probability density functions.

$$p_z(z) = p(x_1,\ x_2,\ \ldots,\ x_n)\ =\ \prod_{i=1}^{n} p(x_i)$$

Example: Find the PDF of the random variable resulting from the sum of two independent random variables, if the marginal PDF of the variables are known.

$$Z = X + Y$$

This is applied to the characterization of noise which results from adding two different kind of interferences. The objective is to find the PDF of random variable z, as a function of the PDFs of the random variables x and y.

The procedure to be applied consists in finding the cumulative distribution function of z:

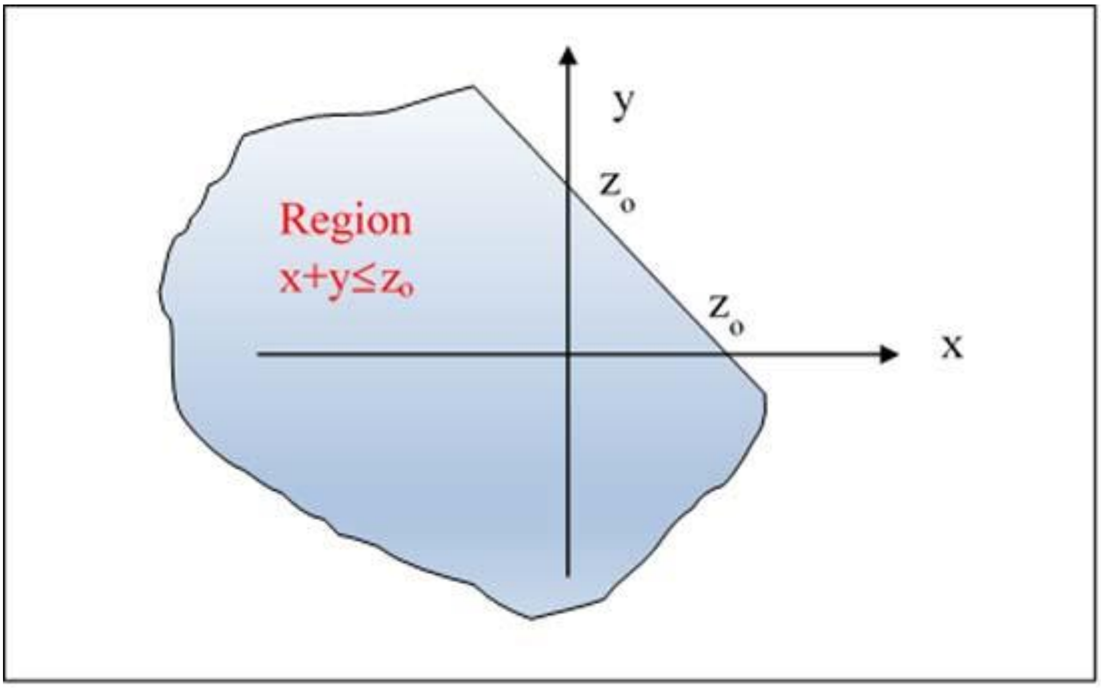

Figure 4 Graphical representation of integration area to find the CDF of the sum of two random variables

$$F_z(z_0)\ =\ P(z \leq z_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{z_0 - y} p_{xy}(x, y)dxdy$$

If both random variables are independent: $p_{xy}(x,y) = p_x(x)p_y(y)$

$$F_z(z_0) = \int_{-\infty}^{\infty} p_y(y) \int_{-\infty}^{z_0-y} p_x(x)dxdy$$

$$p_z(z_0) = \frac{dF_z(z_0)}{dz_0} \int_{-\infty}^{\infty} p_y(y)p_x(z_0 - y)dy$$

The PDF of the sum of two independent random variables is equal to the CONVOLUTION of the marginal PDFs. An important consequence is the fact that the sum of normal or Gaussian random variables is always Gaussian, since the convolution of Gaussian functions is also Gaussian.

An important question arises: Why don't we use the Fourier transform, so that the Fourier transform of the desired PDF is obtained as the product of the marginal PDFs Fourier transforms?

- **Characteristic Function**: It is defined as the Fourier transform of the probability density function.

$$c_x(f) = E\left\{e^{j2\pi \mathrm{fx}}\right\} = \int_{-\infty}^{\infty} p_x(x)e^{j2\pi \mathrm{fx}}dx$$

Central limit theorem

The Central limit theorem states that the sum of a sufficiently large number of independent random variables tends toward a Normal random variable, independently of the distributions of the random variables which are added. A consequence of this theorem is this: the sample mean (normalized sum of realizations of a random variable) is another random variable, which probability density function tends to be Gaussian if the number or samples is large enough.

A condition to this theorem is that the variance of the added independent random variables must tend to infinity when this number also tends to infinity.

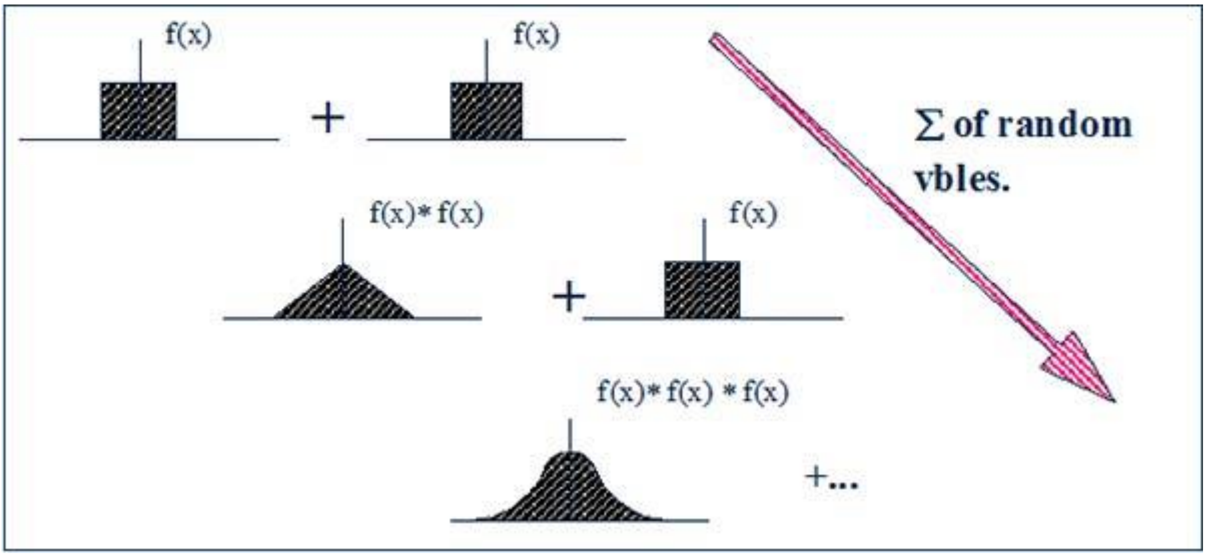

Figure 5 Probability density function of the sum of independent uniform random variables

Random signals (Stochastic processes)

A random signal (also known as Stochastic Process) is a function of one or more deterministic variables (usually time, in communications), and one or more random variables:

X (t, ω) where t is a deterministic variable, and ω is a random variable.

Interpretation:

- For fix ω: it is a time function known as "sample" or "realization".
- For fix t: it is a function of random variable.
- For fix t and ω: it is just a number.

The random signal is characterized with the Cumulative Distribution Function and the first order Probability Density Function:

$$F_x(x,\ t) \ = \ P\left[X(t) \leq x\right]$$

$$p_x(x,\ t) \ = \ \frac{\partial F_x(x,t)}{\partial x}$$

The moments of a random signal are functions of the deterministic variable. For example, if the deterministic variable is time, the mean is a time function.

$$\eta(t) = E[X(t)] = \int_{-\infty}^{\infty} x \cdot p_x(x, t)dx$$

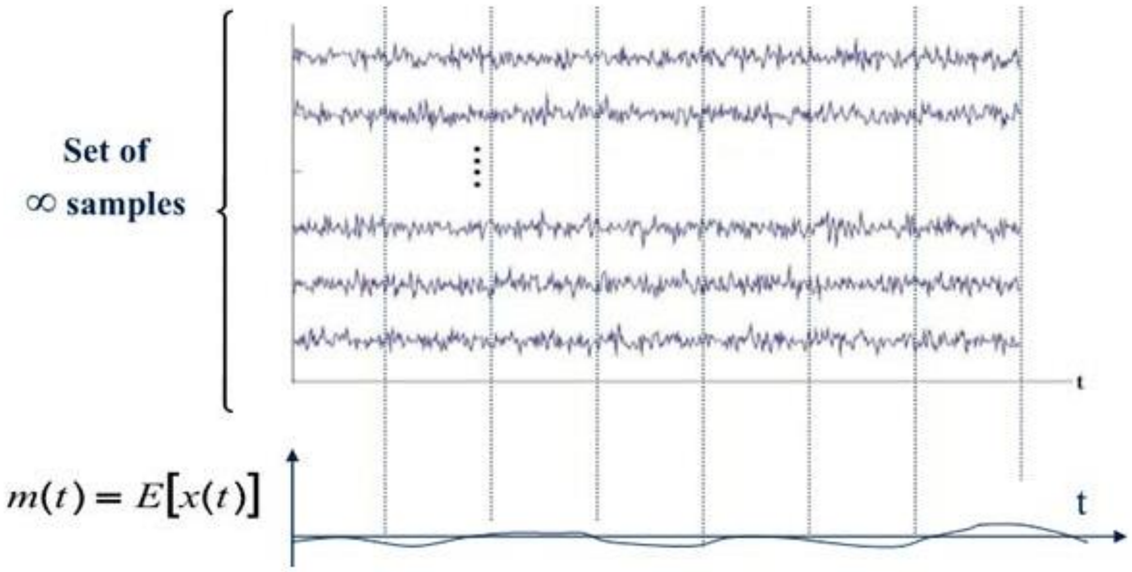

Figure 6 Illustration of random signal meaning

For the complete statistical characterization of random signals, the relationships among random variables obtained particularizing the random signals at different time instant is needed. As the number of random variables can be infinite, the complete characterization can be impossible. The relationship between pairs of random variables is usually studied, to obtain more information about the random signal properties, with the **second order cumulative distribution function** and the **second order probability density function**:

$$F_X(x_1, x_2, t_1, t_2) = P[X(t_1) \leq x_1; X(t_2) \leq x_2]$$

$$p_x(x_1, x_2, t_1, t_2) = \frac{\partial F_x(x_1, x_2, t_1, t_2)}{\partial x_1 \partial x_2}$$

These are the joint cumulative distribution function and the joint probability density function of two random variables, result of the particularization of the random signal in two time instants. Generally, these functions depends on the time instants.

Autocorrelation function

Correlation measures how predictable a random variable is from another one. The autocorrelation function is a function which depends on the time instants values defining the random variables and measures the predictability degree of two random variables defined in different time instants in a random signal. It is defined with the following expression:

$$R_{xx}(t_1, t_2) = E\left[X(t_1)X(t_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_x(x_1, x_2, t_1, t_2)dx_1 dx_2$$

When time instants are equal, we define the Mean Squared Value:

$$R_{xx}(t,\ t) = E\left[X(t)^2\right] = \int_{-\infty}^{\infty} x^2 p_x(x, t)dx$$

Autocovariance

$$C_{xx}(t_1,t_2) = E\left\{\left[X(t_1) - \eta_x(t_1)\right]\left[X(t_2) - \eta_x(t_2)\right]\right\} = R_{xx}(t_1, t_2) - \eta_x(t_1)\eta_x(t_2)$$

It measures how predictable a zero-mean random variable is from another zero-mean random variable. To define the zero-mean random variable, the mean is subtracted from the random variable.

The Variance is obtained when both time instants are equal:

$$C_{xx}(t,t) = E\left\{\left[X(t) - \eta_x(t)\right]\left[X(t) - \eta_x(t)\right]\right\} = R_{xx}(t, t) - \eta_x(t)\eta_x(t)$$

Note that in a random signal, the mean and the variance are time functions. When the variance is zero, the spread around the mean is null, and a DETERMINISTIC signal is obtained.

Cross-correlation of two random signals

Similarly, the predictability between random variables obtained with two random signals could be studied, with the Cross-correlation:

$$R_{xy}(t_1, t_2) = E\left[X(t_1)Y(t_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 y_2 p_{xy}(x_1, y_2, t_1, t_2)dx_1 dy_2$$

Orthogonality: Two random variables are orthogonal if the cross-correlation is zero. Additionally, two random signals are orthogonal if all pairs of random variables from the random signals are orthogonal.

Cross-covariance of two random signals

It is defined with the following expression:

$$C_{xy}(t_1, t_2) = R_{xy}(t_1, t_2) - \eta_x(t_1)\eta_y(t_2)$$

Uncorrelation: Two random variables are said to be `uncorrelated' if the cross-covariance is zero. Two random signals are uncorrelated if all pairs of random variables defined in the random signals are uncorrelated.

Stationarity

- A random signal X(t) is STATIONARY in STRICT SENSE, if its statistical properties do not depend on time, that is, X(t) and X(t+τ) have the same statistical properties (same CDFs, PDFs, second order CDFs and PDFs, etc.).
- A random signal X(t) is STATIONARY in WIDE SENSE, IFF:
  - The mean is constant (does not depend on time): $E[X(t)] = \eta$.

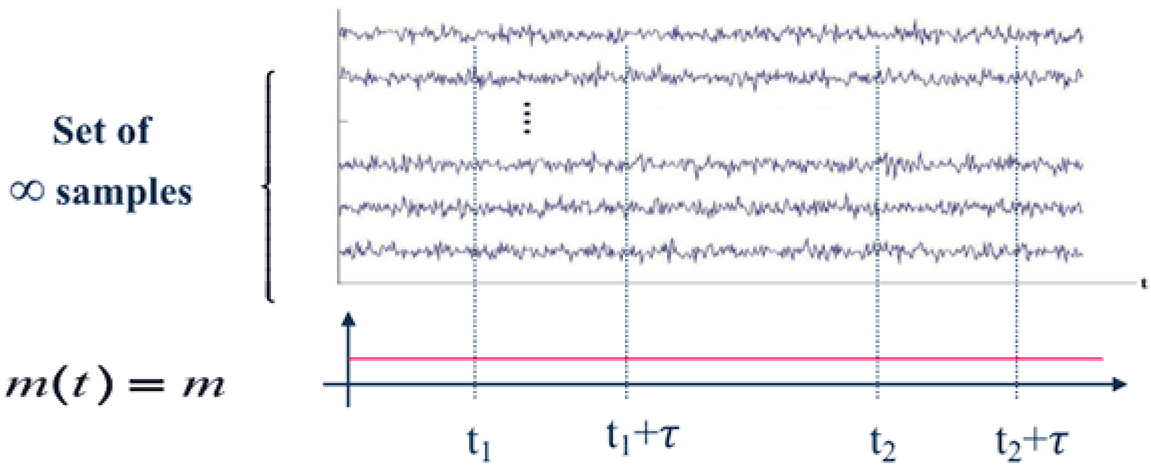

Figure 7 Meaning of "stationarity" in the mean

- The autocorrelation only depends on the time instants difference:

$$R_{xx}(t_1, t_2) = R_{xx}(t_2 - t_2) = R_{xx}(\tau)$$

- The first order PDF is independent of time:

$$p(x,t) = \frac{\partial F_x(x,t)}{\partial x} = p_x(x)$$

Ergodicity

Characterization of random signals is very important in telecommunication. The available information is usually scarce, limited to one random sample or realization of the random process or random signal.

A Stochastic Process or Random Signal is said to be an **Ergodic Process** if its statistical properties can be deduced from a single, sufficiently long, random sample of the process (realization). In these signals it is possible to substitute the ensemble averages with averages in time. The averages calculated in the time domain with one random sample of the process are constant (mean, variance, etc), therefore only Stationary Processes can be Ergodic:

ERGODICITY ⇒ STATIONARITY

- **Averages in time:**
- The "**sample mean**" of continuous-time x(t) and discrete-time x[n] signals, respectively, are given by:

$$<n_x>_T = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} X(t)dt \qquad <n_x>_N = \frac{1}{2N+1}\sum_{n=-N}^{N} X[n]$$

- The "**sample variance**" of continuous-time x(t) and discrete-time x[n] signals, respectively, are given by:

$$<\sigma_X^2>_T = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} (X(t) - (\mu_X)_T)^2 dt \qquad <\sigma_X^2>_N = \frac{1}{2N+1}\sum_{n=-N}^{N} (X[n] - <\mu_X>_N)^2$$

The sample mean and sample variance of a random signal are random variables, because their values depend on the particular sample function over which they are calculated.

- The **time-averaged sample Autocorrelation Function** (ACF) of continuous-time x(t) and discrete-time x[n] signals, respectively, are given by:

$$<R_{XX}>_T = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} X(t)X^*(t+\tau)dt$$

$$<R_{XX}>_N = \frac{1}{2N+1}\sum_{n=-N}^{N} X[n]X^*[n+m]$$

In an ergodic process, these averages in time converge to the expectations (probabilistic mean, probabilistic variance and autocorrelation):

$$\lim_{T\to\infty} <\eta_x>_T = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} X(t)dt \;=\; E[X(t)] \;= \eta_x$$

$$\lim_{T\to\infty} <\sigma_X^2>_T = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} (X(t)-(\mu_X)_T)^2 dt \;=\; E\left[\left(X(t)-\eta_x\right)^2\right] \;=\; \sigma_X^2$$

$$\lim_{T\to\infty} <R_{XX}>_T = \lim_{T\to\infty} \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} X(t)X^*(t+\tau)dt \;=\; E[(X(t)X^*(t+\tau)] \;=\; R_{XX}(\tau)$$

Equivalently for discrete-time ergodic random signals:

$$\lim_{N\to\infty} <\eta_X>_N \;=\; \lim_{N\to\infty} \frac{1}{2N+1}\sum_{n=-N}^{N} X[n] \;=\; E[X[n]] \;=\; \eta_X$$

$$\lim_{N\to\infty} <\sigma_X^2>_N \;=\; \lim_{N\to\infty} \frac{1}{2N+1}\sum_{n=-N}^{N} (X[n] \;-<\mu_x>_N)^2 = \; E\left[\left(X[n]-<\mu_X>_N\right)^2\right] \;=\; \sigma_X^2$$

$$\lim_{N\to\infty} <R_{XX}>_N \;=\; \lim_{N\to\infty} \frac{1}{2N+1}\sum_{n=-N}^{N} X[n]X^*[n+m] \;=\; E[X[n]X^*[n+m]] \;=\; R_{XX}[m]$$

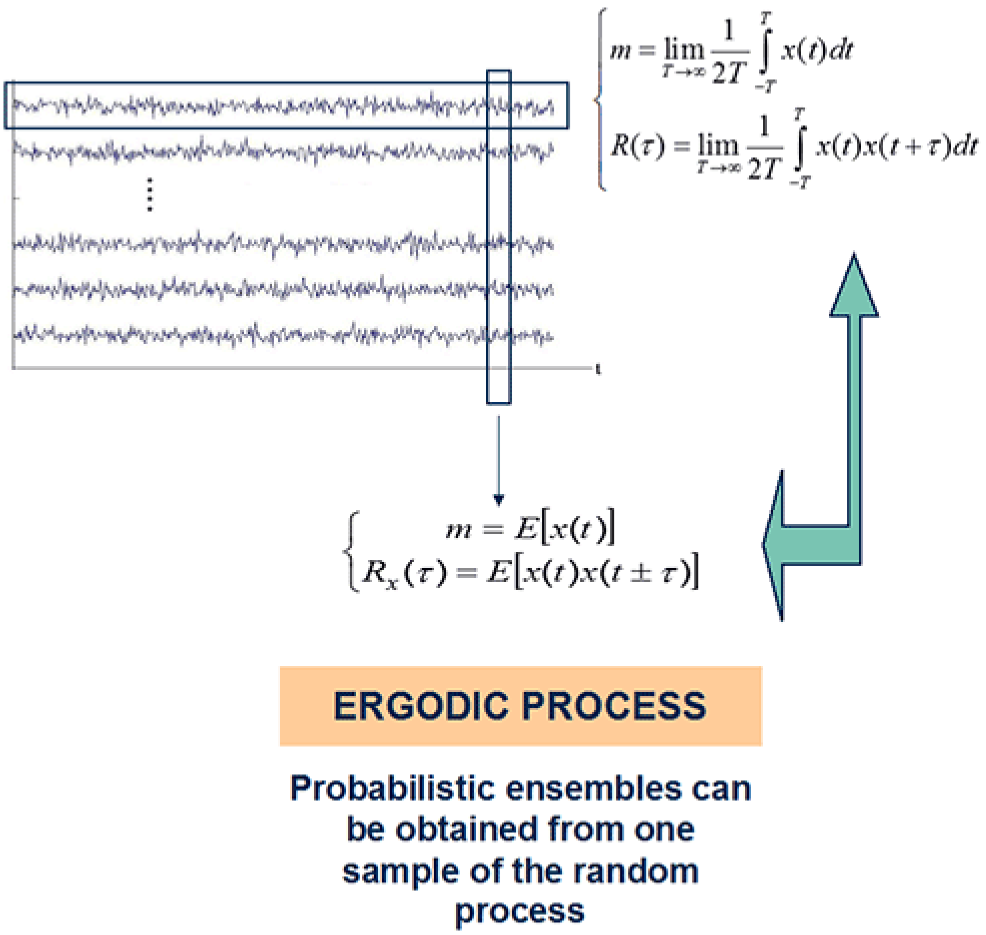

Figure 8 Illustration of “ergodicity”

Power spectral density of stationary processes

The Power Spectral Density is defined as the Fourier transform of the autocorrelation function. It is only defined for Stationary Processes. It provides information about the distribution of power in the frequency domain:

$$S_{XX}(\Omega) = \int_{-\infty}^{\infty} R_{XX}(\tau)e^{-j\Omega\tau}d\tau \leftrightarrow R_{XX}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XX}(\Omega)e^{j\Omega\tau}d\Omega$$

$$S_{XX}(\omega) = \sum_{n=-1}^{N} R_{XX}(m)e^{-j\omega m} \leftrightarrow R_{XX}[m] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{XX}(\omega)e^{j\omega m}d\omega$$

Properties:

1. The Power Spectral Density is a non-negative function.
2. The random process power is obtained by integrating the power spectral density:

$$P_x = R_{XX}(0) \ = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{XX}(\Omega)d\Omega \text{ (for continuous time processes)}$$

$$P_x = R_{XX}(0) \ = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{XX}(\omega)d\omega \text{ (for discrete time processes)}$$

3. The Power Spectral Density is an even function If the process is real:

   $S_{XX}(\omega) \ = S_{XX}(-\omega)$

White noise

White noise is a very special kind of stationary random process. It is a model very useful in communications, which describes, for example, thermal noise generated by electronic devices, or noise in many band-limited communication channels.

Definition: X(t) is white noise in strict sense IFF X(t_1) and X(t_2) are independent random variables for any t_1 and t_2.

Definition: X(t) is White noise in wide sense IFF the autocovariance function is zero:

$$C_{XX}(t_1, t_2) \ = R_{XX}(t_1, t_2) \ - \eta_x(t_1)\eta_x(t_2) \ = \ 0, \ \forall t_1, t_2$$

If a process is white noise in strict sense, it is also white noise in wide sense. Therefore, for white noise, the autocorrelation function is equal to the product of means, for any $t_1 \neq t_2$:

$$R_{XX}(t_1, t_2) \ = \eta_X(t_1, t_2) \ - \eta_x(t_1)\eta_x(t_2) \ = \ \eta_X^2, \ \forall t_1 \neq \ t_2$$

The autocorrelation function of white noise can be expressed as follows:

$$R_{XX}(\tau) \ = \sigma_X^2 \delta(\tau) \ + \ \eta_X^2 \text{ (continuous time white noise)}$$

$$R_{XX}[m] \ = \sigma_X^2 \delta[m] \ + \ \eta_X^2 \text{ (discrete time white noise)}$$

The power spectral density is constant plus a delta function in the origin, due to the mean value of the process in the time domain.

Linear and time invariant systems, with random signals at the input

Figure 9 Linear and Time Invariant system with random input

There are very useful relations between the moments of the input and output of the LTI (linear and time invariant) system. The cross-correlation function of the input and output processes can be related to the autocorrelation function of the input, and the impulse response of the LTI system.

Here, only the relations for discrete time systems and signals are presented. There are similar relations for continuous time processes.

$$\eta_Y = n_X H(0) = \eta_X \sum_{n=-\infty}^{\infty} h[n]$$

$$R_{XY}[m] = R_{XX}[m] * h[-m] \Leftrightarrow S_{XY}(\omega) = S_{XX}(\omega) H*(\omega)$$

$$R_{YX}[m] = R_{XX}[m] * h[m] \Leftrightarrow S_{YX}(\omega) = S_{XX}(\omega) H(\omega)$$

$$R_{YY}[m] = R_{XY}[m] * h[m] \Leftrightarrow S_{YY}(\omega) = S_{XY}(\omega) H(\omega)$$

$$R_{YY}[m] = R_{YX}[m] * h[-m] \Leftrightarrow S_{YY}(\omega) = S_{YX}(\omega) H*(\omega)$$

$$R_{YY}[m] = R_{XX}[m] * h[m] * h[-m] \Leftrightarrow S_{YY}(\omega) = S_{XX}(\omega) |H(\omega)|^2$$

Summary

You should now be able to:

- Describe the concept of random experiment, probability and conditional probability.
- Calculate the probability of events.
- Describe and characterize random variables, and calculate probabilities using the probability density function.
- Characterize sets of independent random variables with the joint probability density function.
- Describe what is a “random process” or “random signal”.

- Describe the meaning of stationarity and ergodicity.
- Given a LTI system with random input, calculate moments of the output, the cross-correlation function of the input and the output, and the autocorrelation function of the output, using the impulse response of the system.

Further reading

1. Gianluca Bontempi & Souhaib Ben Taieb, "*Statistical Foundations of Machine Learning*", Open-Access Textbooks (available on-line at: https://www.otexts.org/sfml)
2. Hossein Pishro-Nik, "Introduction to Probability, Statistics, and Random Processes" (available on-line at: https://www.probabilitycourse.com)
3. Probability lessons by ActuarialPath (Probability: Lesson 1- Basics of Set Theory)
4. Athanasious Papoulis, "*Probability, Random Variables and Sthochastic Process, fourth edition*", McGraw Hill Education, 2015.