Protocoles, Réseaux et Processus Stochastiques

Paulo Gonçalves and Pascale Primet Vicat-Blanc

INRIA, ENS-Lyon (LIP)

E-mail : paulo.goncalves@ens-lyon.fr, pascale.primet@ens-lyon.fr

2007-2008

Références

[1] A. Blanc-Lapierre and Fortet. *Théorie des fonctions aléatoires.* Masson edition, 1953.

[2] A. Blanc-Lapierre and B. Picinbono. *Fonctions aléatoires.* Masson edition, 1981.

[3] R. Jain. *The art of computer systems performance analysis (Techniques for Experimental Design, Measurement, Simulation, and Modeling).* John wiley & sons, inc. edition, 1991.

[4] E. Lukacs. *Developments in Characteristic Function Theory.* New York, macmillan co. edition, 1983.

[5] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* Mc Graw Hill, 3 edition, 1991.

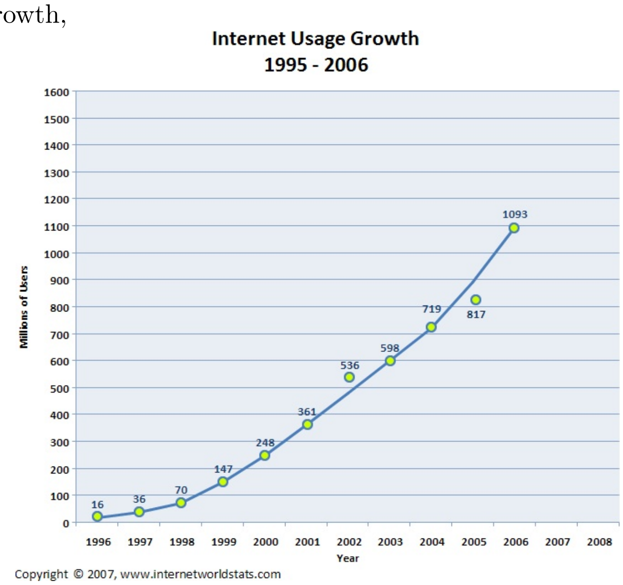
[6] Y. Thomas. *Signaux et systèmes linéaires.* Masson edition, 1992.

[7] A.M. Yaglom. *Correlation Theory of Stationary and Related Random Functions I.* Springer verlag edition, 1987.

Table des matières

1 Motivation

Deterministic approaches fail at handling system of increasing complexity due to :
– usage exponential growth,



– system heterogeneity,
– applications diversity (P2P, VOiP, ftp, http, . . .), . . .

Need for probabilistic / statistical tools for :
– performance analysis (queuing theory, graph theory),
– network modeling (stochastic geometry) [Lecture M1 by E. Fleury “Algorithme pour les Telecom”]
– a better characterization of traffic load (stochastic processes) ,. . .

Applied to networks, objectives of stochastic processes analysis and modeling are :
– to identify relevant indicators of the present system state : e.g., to instantaneously detect attacks, deny of service, overloads,. . .
– to propose accurate predictors of short-term changes : to undertake preventive actions ;
– to design *intelligent* protocols that dynamically adapt to environment (e.g. at router scale) described by real-time measurements : a *data-driven protocol*.

As an illustration, we consider the data flow corresponding to the packet (header) stream on an aggregated link :

timestamp	Prot	ip_src	ip_dst	sport	dport	length	tcp_flags
4746909.842464	T	10.69.7.225	138.96.20.224	8649	47899	1500	A
4746909.842477	T	10.69.7.225	138.96.20.224	8649	47899	1500	A
4746909.849537	T	10.69.7.225	138.96.20.224	8649	47899	1500	A
4746909.849550	T	10.69.7.225	138.96.20.224	8649	47899	1471	FPA
4746909.988728	T	10.69.7.225	138.96.20.224	8649	47899	52	A
4746915.047051	T	10.69.7.227	129.88.70.61	22	36929	52	A
4746917.706674	2	192.168.4.13	224.0.0.22	-	-	32	-
4746918.319298	T	10.69.7.227	129.88.70.61	22	36934	52	A
4746924.992842	T	10.69.7.225	138.96.20.224	8649	47910	60	SA
4746925.000186	T	10.69.7.225	138.96.20.224	8649	47910	1500	A
4746925.000198	T	10.69.7.225	138.96.20.224	8649	47910	919	PA
4746925.000206	T	10.69.7.225	138.96.20.224	8649	47910	97	PA
4746925.007549	T	10.69.7.225	138.96.20.224	8649	47910	1500	A
4746925.007561	T	10.69.7.225	138.96.20.224	8649	47910	1500	A

It is can be converted to different time series, or stochastic processes :

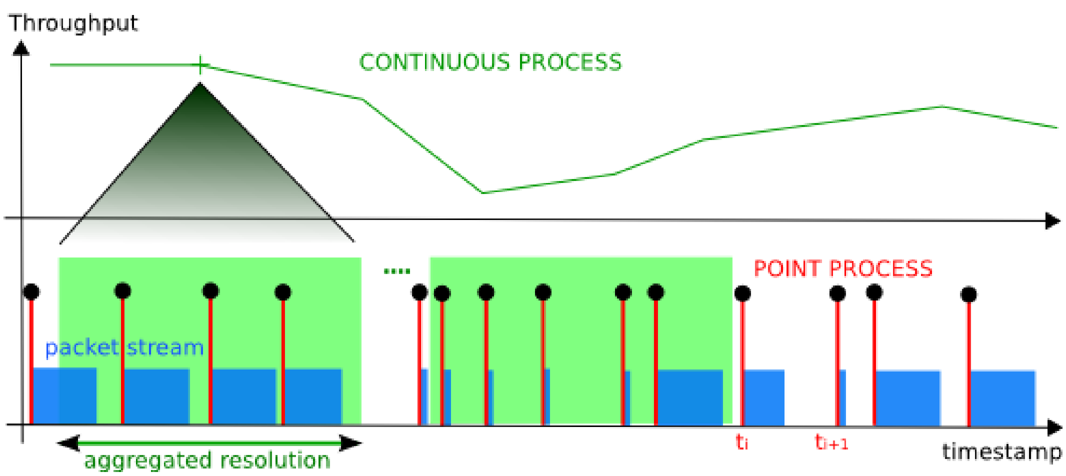


FIG. 1 – Conversion of a packet stream to different time series and stochastic processes.

The purpose of this lecture is then to introduce *statistical signal analysis tools* to characterize these random processes. Hopefully, the estimated quantities will provide with a snapshot, image of the current network state.

2 Probability, random variables and stochastic processes [5]

This section introduces basic concepts and usual notations.

2.1 Probability

2.1.1 Definitions

We shall pose the following notations :

Certain event Ω is a set of elements ;

Experimental outcomes ω_i are the elements of Ω

Event $\mathcal{A}$ is a subset of Ω. The empty set $\{\emptyset\}$ is the impossible event. The event $\{\omega_i\}$ consisting of the single element (experimental outcome) ω_i is an elementary event.

Partition It is a collection of mutually exclusive subsets $\mathcal{A}_i$ of Ω whose union equals Ω.

Example 1 *Let us consider a dice experiment interpreted by two players X, Y.*
Player X says that the outcomes of his experiment are the 6 faces of the dice. The space $\Omega = \{f_1, f_2, f_3, f_4, f_5, f_6\}$ *has* $n = 6$ *elements, and* $2^n = 64$ *subsets. The event (subset)* $\mathcal{A} = \{even\}$ *consists of the three experimental outcomes* $\{f_2, f_4, f_6\}$*.*
Player Y wants to bet on even *or* odd *only. The certain event* $\Omega = \{even, odd\}$ *consists in only two experimental outcomes and among the* $2^n = 4$ *possible subsets, the specific event* $\mathcal{A} = \{even\}$ *is composed of a single outcome.*

Definition 1 (Axiomatic definition) *We assign to each event* $\mathcal{A}$ *a number* $P(\mathcal{A})$ *called the* probability of the event $\mathcal{A}$*. This number must satisfy the three following axiomatic conditions :*

$$\textbf{I.} \quad P(\mathcal{A}) \geq 0 \tag{1}$$
$$\textbf{II.} \quad P(\Omega) = 1 \tag{2}$$
$$\textbf{III.} \quad \text{if } \mathcal{A}\mathcal{B} = \emptyset \text{ (mutually exclusive), then } P(\mathcal{A}+\mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) \tag{3}$$

Remark. In certain case of sets with infinitely many outcomes, it is not possible to assign a probability P satistfying all axioms of definition 1 to all possible subsets (events) of Ω. We are then led to consider only a class $\mathfrak{U}$ of subsets of Ω. A class $\mathfrak{U}$ is a **field** :
– if it is non-empty,
– if $\mathcal{A} \in \mathfrak{U}$ then $\overline{\mathcal{A}} \in \mathfrak{U}$
– if $\mathcal{A} \in \mathfrak{U}$ and $\mathcal{B} \in \mathfrak{U}$ then $\mathcal{A} + \mathcal{B} \in \mathfrak{U}$
Moreover, suppose $\mathcal{A}_1, \ldots, \mathcal{A}_n, \ldots$ is an **infinite** sequence of sets in $\mathfrak{U}$, whose union and intersection also belong to $\mathfrak{U}$, then $\mathfrak{U}$ is called a **Borel field**.

Frequency interpretation. If an experiment is performed n times and the event $\mathcal{A}$ occurs $n_{\mathcal{A}}$ times, then, with a *high degree of certainty*, the relative frequency $n_{\mathcal{A}}/n$ of the occurrence of $\mathcal{A}$ is *close* to $P(\mathcal{A})$.

2.1.2 Properties.

— The probability of the impossible event is 0 : $P(\emptyset) = 0$
— For any event $\mathcal{A}$: $P(\mathcal{A}) = 1 - P(\overline{\mathcal{A}}) \leq 1$ (because $\mathcal{A} + \overline{\mathcal{A}} = \Omega$ and $\mathcal{A}\overline{\mathcal{A}} = \{\emptyset\}$).
— For any $\mathcal{A}$ and $\mathcal{B}$: $P(\mathcal{A}+\mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A}\mathcal{B}) \leq P(\mathcal{A}) + P(\mathcal{B})$

Countable spaces. If the space Ω consists of $N < \infty$ experimental outcomes, the probability of all events can be expressed in terms of the probabilities of the elementary events $\{\omega_i\}$, noted $P\{\omega_i\} = p_i$, with $p_i \geq 0$ and $p_1 + \ldots + p_N = 1$. Hence, if event $\mathcal{A}$ is composed of r elements $\{\omega_1, \ldots, \omega_r\}$, then $P(\mathcal{A}) = P\{\omega_1\} + \ldots + P\{\omega_r\}$ (this relation holds even if Ω consists of an infinite but countable number of elements).

Noncountable spaces. If Ω consists in a noncountable infinity of elements, then its probability cannot be determined in terms of the probabilities of the elementary events.

To fix the ideas, consider Ω the set of real numbers. It can be shown that it is not possible to assign a probability P (satisfying axioms of definition 1) to all subsets of Ω. To construct a probability space on the real line, we shall consider only events corresponding to intervals $x_1 \leq x \leq x_2$, and to their countable unions and intersections. The resulting class is the smallest Borel field containing all half-lines $x \leq x_i$, with x_i any real number. This field contains all open and closed intervals, all points, and most set of points on the real line that usually matter in applications (except e.g. Cantor sets which are not countable unions and intersections of intervals).
It suffices now to assign a probability to the events $x_1 \leq x \leq x_2$ to determine all other probabilities from the axioms. Suppose that $p(x)$ is a function such that :

$$\int_{-\infty}^{\infty} p(x)\, dx = 1 \quad \text{and} \quad p(x) \geq 0.$$

We define the probabilty of the event $\{x \leq x_i\}$ by the integral

$$P\{x \leq x_i\} = \int_{-\infty}^{x_i} p(x)\, dx,$$

which specifies the probability of all events of Ω. In paticular, considering event $\{x_1 < x \leq x_2\}$ consisting of all the points in the interval (x_1, x_2), it is mutually exclusive with event $\{x \leq x_1\}$, and their union equals $\{x \leq x_2\}$. We then have,

$$\begin{aligned} & P\{x \leq x_1\} + P\{x_1 < x \leq x_2\} = P\{x \leq x_2\} \\ \Leftrightarrow \quad & P\{x_1 < x \leq x_2\} = P\{x \leq x_2\} - P\{x \leq x_1\} = \int_{-\infty}^{x_2} p(x)\, dx - \int_{-\infty}^{x_1} p(x)\, dx = \int_{x_1}^{x_2} p(x)\, dx \end{aligned}$$

Notice that if $p(x)$ is bounded, then $P\{x_1 < x \leq x_2\}$ tends to zero as $x_1 \to x_2$. This leads to the conclusion that the probability of all elementary event $\{x_i\}$ of Ω is zero, although the probability of their union equals 1 (this is so, because the total number of elements of Ω is not countable). The experiment outcome $\{x_i\}$ is improbable but not impossible.

Conditional probability.
The conditional probability of an event $\mathcal{A}$ assuming the event $\mathcal{M}$, denoted $P(\mathcal{A} \mid \mathcal{M})$, is defined by the ratio

$$P(\mathcal{A} \mid \mathcal{M}) = \frac{P(\mathcal{A}\mathcal{M})}{P(\mathcal{M})}.$$

Consequently :

$$\text{if} \quad \mathcal{M} \subset \mathcal{A} \quad \text{then} \quad P(\mathcal{A} \mid \mathcal{M}) = 1$$

$$\text{if} \quad \mathcal{A} \subset \mathcal{M} \quad \text{then} \quad P(\mathcal{A} \mid \mathcal{M}) = \frac{P(\mathcal{A})}{P(\mathcal{M})} \geq P(\mathcal{A})$$

Theorem 1 (Bayes' theorem) *Let $\mathcal{U} = [\mathcal{A}_1, \ldots, \mathcal{A}_n]$ be a partition of Ω and $\mathcal{B}$, an arbitrary event, then*

$$P(\mathcal{A}_i \mid \mathcal{B}) = \frac{P(\mathcal{B} \mid \mathcal{A}_i) P(\mathcal{A}_i)}{P(\mathcal{B} \mid \mathcal{A}_1) P(\mathcal{A}_1) + \ldots + P(\mathcal{B} \mid \mathcal{A}_n) P(\mathcal{A}_n)} \tag{4}$$

Independence. Two events $\mathcal{A}$ and $\mathcal{B}$ are called independent if

$$P(\mathcal{A}\mathcal{B}) = P(\mathcal{A})\, P(\mathcal{B}).$$

As a corollary, if $\mathcal{A}$ and $\mathcal{B}$ are two independent events, so are events $\overline{\mathcal{A}}$ and $\mathcal{B}$ and events $\overline{\mathcal{A}}$ and $\overline{\mathcal{B}}$.

2.2 Random variable

2.2.1 Definitions

A random variable (RV) is a number $\mathbf{x}(\omega)$ assigned to every outcome of an experiment (e.g. the gain in a dice game). It is thus a function $\mathbf{x}$ whose domain is the set Ω of experimental outcomes, and the range a set of numbers ($\mathbb{R}, \mathbb{C}, \ldots$)
It is then natural to relate the probability of a RV, to the probability of the corresponding generated events. So, the notation $\{\mathbf{x} \leq x\}$ represents a subset of Ω consisting of all outcomes ω such that $\mathbf{x}(\omega) \leq x$. Thus, $\{\mathbf{x} \leq x\}$ is not a set of numbers, but a set of experimental outcomes.

Definition 2 (Random Variable) *A RV* $\mathbf{x}$ *is a process of assigning a number* $\mathbf{x}(\omega)$ *to every outcome* ω *:*

$$\mathbf{x} : \begin{cases} \Omega & \longrightarrow & \mathbb{R} \quad (\mathbb{C}, \ldots) \\ \omega & \longrightarrow & \mathbf{x}(\omega) = x \end{cases}$$

Function $\mathbf{x}$ *must satisfy the following two conditions :*

I. *The set* $\{\mathbf{x} \leq x\}$ *is an event for every* x*.*
II. *The probability of the event* $\{\mathbf{x} = \infty\}$ *and* $\{\mathbf{x} = -\infty\}$ *equals 0 :*

$$P\{\mathbf{x} = \infty\} = 0 \qquad P\{\mathbf{x} = -\infty\} = 0$$

As the elements of the set Ω contained in the event $\{\mathbf{x} \leq x\}$ vary with x, the probability $P\{\mathbf{x} \leq x\}$ is a number that also depends on x.

Definition 3 (Distribution function) *The (cumulative) distribution function of a RV* $\mathbf{x}$ *is the function*

$$F_x(x) = P\{\mathbf{x} \leq x\}, \quad \textit{for all } x \textit{ from } -\infty \textit{ to } \infty$$

Some properties.
1. $F(-\infty) = 0\,;\, F(\infty) = 1$
2. if $x_1 < x_2$ then $F(x_1) \leq F(x_2)$: nondecreasing function
3. $P\{\mathbf{x} > x\} = 1 - F(x)$
4. $P\{x_1 < \mathbf{x} \leq x_2\} = F(x_2) - F(x_1)$
5. $P\{\mathbf{x} = x\} = F(x) - F(x^-)$. In particular if F is continuous (i.e. RV $\mathbf{x}$ is of continuous-type), then $P\{\mathbf{x} = x\} = 0$.

Definition 4 (Density function) *The derivative*

$$p(x) = \frac{dF(x)}{dx}$$

is called the density function of RV $\mathbf{x}$*.*

Properties of p straightforwardly derive from those of F.
Remark. Often, we consider RV $\mathbf{x}$ having specific distribution or density functions without any reference to a particular probability space Ω. For this, given a particular function F (or p), we need to systematically consider as our space Ω the set of all real numbers, and as its events all intervals on the real line and their unions and intersections. To each event (experiment outcome) $\{x \leq x_1\}$ is assigned the probability $P\{x \leq x_1\} = F(x_1)$, and the random variable is $\mathbf{x}(x) = x$. Thus x is the outcome of the experiment and the corresponding value of the RV $\mathbf{x}$.

2.2.2 Examples of common densities

Normal. A RV is called *normal* if or *gaussian* if its density is the normal curve,

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

The corresponding distribution $F(x) = \int_{-\infty}^{x} p(x)\, dx$ is the Gauss error function (erf) noted $\mathbb{G}_{m,\sigma}(x)$.

Remark : Importance of normal RV stems from the *Central Limit Theorem*. Given n independent RVs $\mathbf{x}_i$, we form the sum

$$\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_n$$

Under certain general conditions [1], the distribution of $\mathbf{x}$ of $\mathbf{x}$ approaches (equals in law) a normal distribution with mean $m = m_1 + \cdots + m_n$ and variance $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$, i.e. :

$$F_{\mathbf{x}}(x) \underset{n\to\infty}{\longrightarrow} \mathbb{G}_{m,\sigma}(x)$$

Uniform. A RV is called uniform between x_1 and x_2 if its density is constant and equal to $|x_2 - x_1|^{-1}$ on the interval (x_1, x_2) and 0 elsewhere. The corresponding distribution is a linearly increasing function such that, $F(x_1) = 0$ and $F(x_2) = 1$.

Binomial (Bernoulli trial). In the experiment that consists in analyzing a length n binary word, an outcome is a sequence $w_1 \cdots \omega_n$ of k on'es and $n-k$ zero's where $k = 0, \ldots, n$. We define the RV $\mathbf{x}$ equal to the number of one's in the sequence

$$\mathbf{x}(\omega_1 \cdots \omega_n) = k$$

If $P\{1\} = p$, the probability that $\mathbf{x} = k$ follows a **binomial law** :

$$P\{\mathbf{x} = k\} = C_n^k p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

and $\mathbf{x}$ is of discrete-type (lattice-type), its density is a sum of pulses :

$$p(x) = \sum_{k=0}^{k=n} C_n^k p^k (1-p)^{n-k} \delta(x-k)$$

and the corresponding distribution is a staircase function in the interval $(0, n)$:

$$F(x) = \sum_{k=0}^{k=m} C_n^k p^k (1-p)^{n-k}, \quad m \le x < m+1.$$

Remark. When n and np are both *large*, the binomial distribution can be approximated by a Gauss error function (DeMoivre-Laplace theorem) : $F(x) \simeq \mathbb{G}_{0,1}\left((x-np)/\sqrt{np(1-p)}\right)$

[1] Sufficient conditions :

$$\sigma_1^2 + \cdots + \sigma_n^2 \underset{n\to\infty}{\longrightarrow} \infty$$

$$\int_{-\infty}^{\infty} x^\alpha f_i(x)\, dx < K < \infty, \quad \text{for all } i \text{ and for some } \alpha > 2$$

Poisson. A RV $\mathbf{x}$ is Poisson distributed with parameter a if it takes the values $0, 1 \ldots, n, \ldots$ with

$$P\{\mathbf{x} = k\} = e^{-a}\frac{a^k}{k!}, \quad k = 0, 1, \ldots$$

Thus, $\mathbf{x}$ is of lattice-type with density

$$p(x) = e^{-a}\sum_{k=0}^{\infty}\frac{a^k}{k!}\delta(x-k)$$

Interpretation. We place at random n points in the interval $(-T/2, T/2)$ An outcome ω is a set of points $\mathbf{t}_i$ on the t axis, and we define the RV $\mathbf{n}$ such that its value $\mathbf{n}(\omega)$ equals the number of points $\mathbf{t}_i$ lying in the interval (t_1, t_2) of length $t_2 - t_1 = t_a$. We have

$$\begin{aligned} P\{k \text{ in } t_a\} &= P\{\mathbf{n} = k\} = C_n^k \left(\tfrac{t_a}{T}\right)^k \left(1 - \tfrac{t_a}{T}\right)^{n-k} \\ &\simeq e^{-nt_a/T}\frac{(nt_a/T)^k}{k!} \quad \text{when } n \gg a \text{ and } T \gg t_a \text{ (DeMoivre-Laplace thm)} \\ &= P_{poisson}(k) \text{ with } a = nt_a/T \end{aligned}$$

and moreover, if $n \to \infty$ and $T \to \infty$ but the density $\lambda = n/T$ kept fixed then

$$P\{k \text{ in } t_a\} = P\{\mathbf{n} = k\} = e^{-\lambda t_a}\frac{(\lambda t_a)^k}{k!}.$$

Exponential. (Exercise). In the Poisson points experiments, let us consider a fix event t_0 and we are interested in the first next point t_1 occurring after t_0. Let us pose $\mathbf{x} = t_1 - t_0$, the RV, and let us find $p(x)$.

$$\begin{aligned} F(x) = P\{\mathbf{x} \le x\} &= P\{\text{there is at least one point in } (t_0, t_0 + x)\} \\ &= 1 - P\{\text{there is no point in } (t_0, t_0 + x)\} \\ &= 1 - P\{0 \text{ in } (t_0, t_0 + x)\} \\ &= 1 - e^{-\lambda x} \end{aligned}$$

and

$$p(x) = \frac{dF(x)}{dx} = \lambda e^{-\lambda x}\mathbb{U}_{[0,\infty[}$$

2.2.3 Function of a random variable

Let us consider $\mathbf{x}$ a RV and form the new RV $\mathbf{y} = g(\mathbf{x})$, where g is a function defined on the support of $\mathbf{x}$. Then, we have

$$F_{\mathbf{y}}(y) = P\{\mathbf{y} \le y\} = P\{g(\mathbf{x}) \le y\}$$

We also know that

$$P\{y < g(\mathbf{x}) \le y + dy\} = F_{\mathbf{y}}(y + dy) - F_{\mathbf{y}}(y) = p_{\mathbf{y}}(y)\,dy$$

From the example depicted in figure 2, we get,

$$p_{\mathbf{y}}(y)dy = \sum_{i=1,\ldots,n} P_{\mathbf{x}}\{x_i \le \mathbf{x} < x_i + dx_i\} = \sum_{i=1,\ldots,n} p_{\mathbf{x}}(x_i)\,dx_i,$$

where the $\{x_i,\ i = 1 \ldots n\}$ are the roots of $g(x) = y$. If g is a continuous function, $dx_i = dy/|g'(x_i)|$ and

$$p_{\mathbf{y}}(y) = \sum_{i=1}^{n}\frac{p_{\mathbf{x}}(x_i)}{|g'(x_i)|}.$$

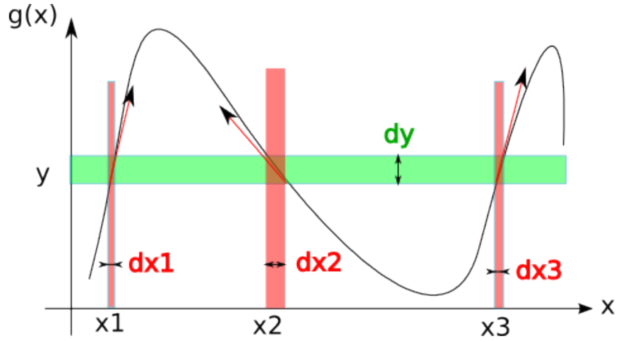


FIG. 2 – Function of a RV.

Example. $\mathbf{y} \sim N(m, \sigma)$ is a normal RV, determine the density of $\mathbf{x} = \exp\{\mathbf{y}\}$.

$$p_{\mathbf{x}}(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log(x)-m)^2}{2\sigma^2}} \quad : \text{ Log-Normal}$$

Inverse problem. Let us consider a uniform RV $\mathbf{x}$ in $(0, 1)$. Determine the function g, such that the distribution of the RV $\mathbf{y} = g(\mathbf{x})$ is a specific function $F_{\mathbf{y}}(y)$.
Let us pose $\mathbf{x} = F_{\mathbf{y}}(\mathbf{y}) \in (0, 1)$, then :

$$\begin{aligned} F_{\mathbf{x}}(x) = P\{\mathbf{x} \leq x\} &= P\{F_{\mathbf{y}}(\mathbf{y}) \leq F_{\mathbf{y}}(y)\} \\ &= P\{\mathbf{y} \leq y\} \quad \text{(because } F_{\mathbf{y}} \text{ is monotonous)} \\ &= F_{\mathbf{y}}(y) = x \quad \text{(by definition)} \\ \Leftrightarrow \quad & F_{\mathbf{x}}(x) = x, \quad \text{for } x \in (0, 1). \end{aligned}$$

Hence, the solution $\mathbf{y} = g(\mathbf{x}) = F_{\mathbf{y}}^{-1}(\mathbf{x})$.

2.2.4 Moments.

Mean. The expected value of a RV $\mathbf{x}$ is, by definition, the integral :

$$\mathbb{E}\{\mathbf{x}\} = \int_{-\infty}^{\infty} x\, p(x)\, dx \;=\; m_{\mathbf{x}}$$

which can also be interpreted as a *Lebesgue integral* :

$$\mathbb{E}\{\mathbf{x}\} = \int_{\Omega} \mathbf{x}\, dF(x)$$

Mean of $g(\mathbf{x})$: Given a RV $\mathbf{x}$ and a function $g(x)$, the mean of RV $\mathbf{y} = g(\mathbf{x})$ is equal to :

$$\mathbb{E}\{\mathbf{y}\} = \int_{-\infty}^{\infty} y\, p_{\mathbf{y}}(y)\, dy \;=\; \int_{-\infty}^{\infty} g(x)\, p_{\mathbf{x}}(x)\, dx$$

Variance. The variance of a RV $\mathbf{x}$ is by definition the intregral :

$$\sigma^2 = \mathbb{E}\{(\mathbf{x} - \mathbb{E}\{\mathbf{x}\})^2\} \;=\; \int_{-\infty}^{\infty} (x - m)^2\, p(x)\, dx \;=\; \mathbb{E}\{\mathbf{x}^2\} - \mathbb{E}^2\{\mathbf{x}\}$$

The positive constant σ is called the *standard deviation* of $\mathbf{x}$.

Moments.

$$m_n \;=\; \mathbb{E}\{\mathbf{x}^n\} \;=\; \int_{-\infty}^{\infty} x^n\, p(x)\, dx$$

Remark : Two RV $\mathbf{x}$ and $\mathbf{y}$ are equal in law (or in distribution) if all their moments of finite order are equal.

Central moments.

$$\mu_n \;=\; \mathbb{E}\{(\mathbf{x} - \mathbb{E}\{\mathbf{x}\})^n\} \;=\; \int_{-\infty}^{\infty} (x-m)^n\, p(x)\, dx$$

Binomial formula : $\mu_n \;=\; \sum_{k=0}^{n} C_n^k\, m_k\, (-m)^{n-k}$ and $m_n \;=\; \sum_{k=0}^{n} C_n^k\, \mu_k\, (m)^{n-k}$

Absolute moments.

$$\mathbb{E}\{|\mathbf{x}|^n\} \;=\; \int_{-\infty}^{\infty} |x|^n\; p(x)\, dx \quad \text{and} \quad \mathbb{E}\{|\mathbf{x} - m|^n\} \;=\; \int_{-\infty}^{\infty} |x-m|^n\; p(x)\, dx$$

Example : $\mathbf{x} \sim N(m, \sigma)$ a normal RV, then

$$\mathbb{E}\{\mathbf{x}^n\} \;=\; \begin{cases} 0 & n = 2k+1 \\ 1 \cdot 3 \cdots (n-1)\sigma^n & n = 2k \end{cases}$$

$$\mathbb{E}\{|\mathbf{x}|^n\} \;=\; \begin{cases} 2^k\, k!\, \sigma^{2k+1}\sqrt{2/\pi} & n = 2k+1 \\ 1 \cdot 3 \cdots (n-1)\sigma^n & n = 2k \end{cases}$$

Inequalities.

Theorem 2 (Tchebycheff Inequality) *For any $\varepsilon > 0$,*

$$P\{|\mathbf{x} - m| \geq \epsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

This theorems bounds the probability that a RV $\mathbf{x}$ deviates from its mean.

Theorem 3 (Markov Inequality) *If $p(x) = 0$ for all $x < 0$, then, for any $\alpha > 0$,*

$$P\{\mathbf{x} \geq \alpha\} \leq \frac{m}{\alpha}$$

This theorem bounds the probability for large deviations of RV $\mathbf{x}$.

Heavy-Tailed distribution. Markov inequality indicates that $P\{\mathbf{x} \geq x\}$ can not decrease slower than $1/x$ as $x \to \infty$, and in particular, if it decreases exponentially fast, then all moments $\mathbb{E}\{\mathbf{x}^n\}$ of $\mathbf{x}$ are finite.

Definition 5 (Heavy-tailed distribution) *A RV $\mathbf{x}$ is called **heavy-tailed** with tail parameter α if*

$$P\{|\mathbf{x}| > x\} \;=\; x^{-\alpha}\, L(x)$$

with $L(x)$ a slowly varying function[2].

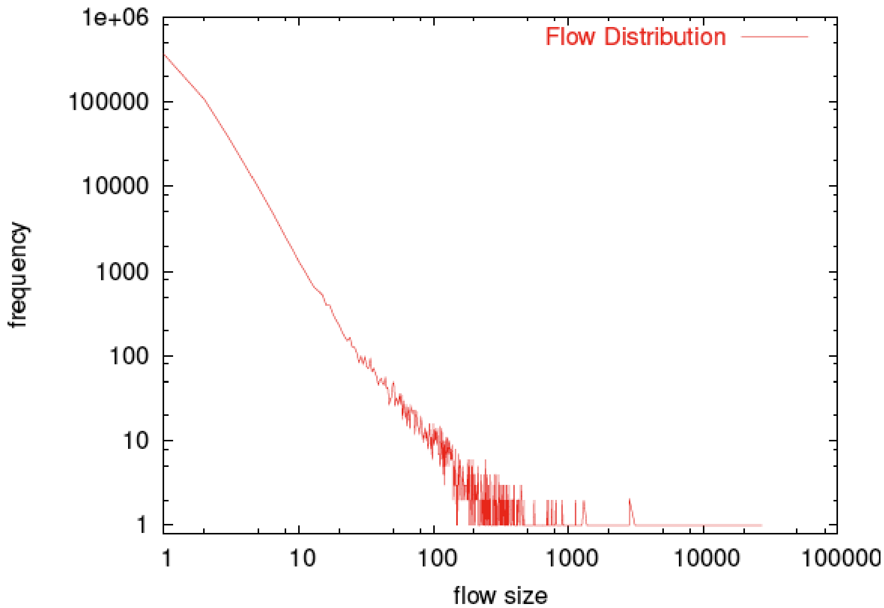


FIG. 3 – Estimated density function of the flow size on a TCP/IP traffic link.

For these RV, it is clear that moments are finite exactly up to order α.
Examples of heavy-tailed distributions
- Pareto distribution ;
- Log-normal distribution ;
- Weibull distribution ;
- Burr distribution ;
- Log-gamma distribution. Those that are two-tailed include :
- Cauchy distribution, itself a special case of
- t-distribution ;
- Stable Distribution family, excepting the special case of the normal distribution within that family.

2.2.5 Characteristic function

The characteristic function (or *moment generating function*) of a RV is by definition the integral :

$$\Phi(\nu) = \int_{-\infty}^{\infty} p(x)\, e^{i\nu x}\, dx \;\; = \;\; \int_{\Omega} e^{i\nu x} dF(x) = \mathbb{E}\left\{e^{i\nu \mathbf{x}}\right\}$$

Differentiation Φ n times, we obtain

$$\Phi^{(n)}(\nu) = \mathbb{E}\left\{(i\nu)^n\, e^{i\nu \mathbf{x}}\right\}, \quad \text{and in particular,} \quad \Phi^{(n)}(0) = i^n\, m_n$$

Then, expanding Φ into a series near the origin, we obtain

$$\Phi(\nu) = \sum_{n=0}^{\infty} \frac{m_n}{n!} (i\nu)^n$$

[2] $L(tx)/L(x) \to 1$, as $x \to \infty$.

thus the name "moment generating function" for Φ. Since $p(x)$ can be entirely determined in terms of Φ, the density of a RV is uniquely determined if all moments are known.

Remark : Sometimes, the *second characteristic function* $\Psi(\nu) = \log \Phi(\nu)$ is prefered, and cumulants of order n defined

$$\lambda_n = (-i)^n \Psi^{(n)}(0)$$

which lead to[3]

$$\Psi(\nu) = \lambda_1(i\nu) + \frac{1}{2}\lambda_2(i\nu)^2 + \cdots + \frac{1}{n!}\lambda_n(i\nu)^n + \cdots$$

The second characteristic function is particularly interesting in the normal case, since then

$$\Psi(\nu) = \log\left(e^{im\nu - \frac{1}{2}\sigma^2\nu^2}\right) = im\nu - \frac{1}{2}\sigma^2\nu^2$$

where we identify[4]

$$\lambda_1 = m \;\;;\;\; \lambda_2 = \sigma^2 \;\text{ and }\; \lambda_n = 0,\; \forall n \geq 3$$

Remark : Characteristic functions are positive-definite functions (as Fourier transform of a positive function).

2.2.6 Two random variables

We briefly generalize notions defined for a random variable to the case of two (or several) RVs.

Joint distribution. Let $\mathbf{x}$ and $\mathbf{y}$ be two RV, we define the joint distribution

$$F_{\mathbf{xy}}(x,y) = P\{\mathbf{x} \leq x \text{ and } \mathbf{y} \leq y\}$$

Consequences :
- $F(-\infty, y) = F(x, -\infty) = 0$
- $F(\infty, \infty) = 1$
- $P\{x_1 < \mathbf{x} \leq x_2 \text{ and } \mathbf{y} \leq y\} = F(x_2, y) - F(x_1, y)$
- $P\{\mathbf{x} \leq x \text{ and } y_1 < \mathbf{y} \leq y_2\} = F(x, y_2) - F(x, y_1)$
- $P\{x_1 < \mathbf{x} \leq x_2 \text{ and } y_1 < \mathbf{y} \leq y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$

Joint density.

$$p_{\mathbf{xy}}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{\mathbf{xy}}(x,y)$$

Marginals. The marginal distributions and the marginal densities of two RV $\mathbf{x}$ and $\mathbf{y}$ are expressed in terms of their joint distribution (respectively their joint density) according to :

$$F_{\mathbf{x}}(x) = F_{\mathbf{xy}}(x, \infty) \quad ; \quad F_{\mathbf{y}}(y) = F_{\mathbf{xy}}(\infty, y)$$

and

$$p_{\mathbf{x}}(x) = \frac{\partial F_{\mathbf{xy}}(x,\infty)}{\partial x} = \int_{-\infty}^{\infty} p_{\mathbf{xy}}(x,y)\,dy \quad ; \quad p_{\mathbf{y}}(y) = \frac{\partial F_{\mathbf{xy}}(\infty,y)}{\partial y} = \int_{-\infty}^{\infty} p_{\mathbf{xy}}(x,y)\,dx$$

[3] Clearly, $\Psi(0) = \log \Phi(0) = \log(1) = 0$.
[4] Edgeworth development of the Gaussian.

Independence.

$$p(x,y) = p(x)p(y) \quad \text{since} \quad P\{\mathbf{x} \le x \text{ and } \mathbf{y} \le y\} = P\{\mathbf{x} \le x\}P\{\mathbf{y} \le y\}$$

Example : Joint normality. RV $\mathbf{x}$ and $\mathbf{y}$ are joint normal if their joint density takes on the form :

$$p_{\mathbf{xy}}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-m_x)^2}{\sigma_x^2} - 2r\frac{(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right]}$$

In this expression m_x and m_y are the expected values of $\mathbf{x}$ and $\mathbf{y}$, and σ_x^2 and σ_y^2, their variance. $|r| < 1$ is the correlation factor. When $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated $r = 0$, and moreover independent since :

$$p_{\mathbf{xy}}(x,y) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} \cdot \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}} = p_{\mathbf{x}}(x) \cdot p_{\mathbf{y}}(y)$$

where $p_{\mathbf{x}}$ and $p_{\mathbf{y}}$ are the marginal distributions.

Function of two RV. Let $\mathbf{x}$ and $\mathbf{y}$ be two RV, and let us form the RV $\mathbf{z} = g(\mathbf{x},\mathbf{y})$. The domain D_z is the region of the xy plane such that $g(x,y) \le z$. Then

$$F_{\mathbf{z}}(z) = P\{\mathbf{z} \le z\} = P\{(\mathbf{x},\mathbf{y}) \in D_z\} = \iint_{D_z} p(x,y)\,dx\,dy$$

Also,

$$\mathbb{E}\{g(\mathbf{x},\mathbf{y})\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)p(x,y)\,dx\,dy = \int_{\Omega_x}\int_{\Omega_y} g(\mathbf{x},\mathbf{y})dF(\mathbf{x},\mathbf{y})$$

Example : Let us consider $\mathbf{x}$ and $\mathbf{y}$ two independent RV, and form the new RV $\mathbf{z} = \mathbf{x} + \mathbf{y}$. The domain D_z such that $x + y \le z$ is defined by $(x \le z - y, y)$, thus

$$F_{\mathbf{z}}(z) = \int_{y=-\infty}^{\infty} \underbrace{\int_{x=-\infty}^{z-y} p_{\mathbf{xy}}(x,y)\,dx}_{\text{primitive of } p_{\mathbf{xy}}(x,y) \text{ in } x} \quad dy$$

and

$$p_{\mathbf{z}}(z)\,dz = dF_{\mathbf{z}}(z) = \int_{y=-\infty}^{\infty} p_{\mathbf{xy}}(z-y,y)\,dy\,dz = \int_{-\infty}^{\infty} p_{\mathbf{x}}(z-y)\,p_{\mathbf{y}}(y)\,dy\,dz$$
$$\Leftrightarrow \quad p_{\mathbf{z}}(z) = p_{\mathbf{x}} \underset{z}{*} p_{\mathbf{y}}$$

Joint moments.

$$m_{kr} = \mathbb{E}\{\mathbf{x}^k\mathbf{y}^r\} = \iint x^k\,y^r\,p_{\mathbf{xy}}(x,y)\,dx\,dy$$

is a joint moment of the RV $\mathbf{x}$ and $\mathbf{y}$ of order $n = k + r$, and similarly for characteristic function, cumulants. . .

First order – Expectation operator $\mathbb{E}$ is linear, then

$$\begin{aligned}\mathbb{E}\{\mathbf{x}+\mathbf{y}\} &= \mathbb{E}_{\mathbf{xy}}\{\mathbf{x}\} + \mathbb{E}_{\mathbf{xy}}\{\mathbf{y}\} \\ &= \mathbb{E}_{\mathbf{x}}\{\mathbf{x}\} + \mathbb{E}_{\mathbf{y}}\{\mathbf{y}\}\end{aligned}$$

Second order – In general $\mathbb{E}\{\mathbf{xy}\} \neq \mathbb{E}\{\mathbf{x}\}\mathbb{E}\{\mathbf{y}\}$, but

if $\mathbf{x}$ and $\mathbf{y}$ are independent $\Leftrightarrow$ $p_{\mathbf{xy}}(x,y) = p_{\mathbf{x}}(x)p_{\mathbf{y}}(y)$ $\Rightarrow$ $\mathbb{E}\{\mathbf{xy}\} = \mathbb{E}\{\mathbf{x}\}\mathbb{E}\{\mathbf{y}\}$

$\mathbb{E}\{\mathbf{xy}\} = \mathbb{E}\{\mathbf{x}\}\mathbb{E}\{\mathbf{y}\}$ $\Rightarrow$ $\mathbf{x}$ and $\mathbf{y}$ are uncorrelated (but not necessarily independent)

$\mathbb{E}\{\mathbf{xy}\} = 0$ $\Rightarrow$ $\mathbf{x}$ and $\mathbf{y}$ are orthogonal

Covariance.

$$\begin{aligned} c_{\mathbf{xy}} &= \mathbb{E}(\mathbf{x}-\mathbb{E}\{\mathbf{x}\})(\mathbf{y}-\mathbb{E}\{\mathbf{y}\}) \\ &= \mathbb{E}\mathbf{xy} - \mathbb{E}\{\mathbf{x}\mathbb{E}\mathbf{y}\} - \mathbb{E}\{\mathbf{y}\mathbb{E}\mathbf{x}\} + \mathbb{E}\{\mathbb{E}\mathbf{x}\mathbb{E}\mathbf{y}\} \\ &= \mathbb{E}\{\mathbf{xy}\} - \mathbb{E}\{\mathbf{x}\}\mathbb{E}\{\mathbf{y}\} \end{aligned}$$

Correlation coefficient.

$$r = \frac{c_{\mathbf{xy}}}{\sigma_x \sigma_y} \quad \text{and} \quad |r| < 1$$

Proof : form the quantity $\mathbb{E}\left\{[a(\mathbf{x}-m_x)+(\mathbf{y}-m_y)]^2\right\} = a^2\sigma_x^2 + 2ac_{\mathbf{xy}} + \sigma_y^2 \geq 0,\ \forall a$, which admits no real roots. Hence, the (reduced) discriminant[5] equals $\Delta = c_{\mathbf{xy}}^2 - \sigma_x^2\sigma_y^2 \leq 0 \ \Rightarrow c_{\mathbf{xy}} \leq \sigma_x\sigma_y$

Interpretation. $\mathbb{E}\{\mathbf{xy}\}$ is equivalent to the inner product of $\mathbf{x}$ and $\mathbf{y}$. Thus, $\mathbf{x}$ and $\mathbf{y}$ are orthogonal *iff* $\mathbb{E}\{\mathbf{xy}\} = 0$.

Sequence of random variables Let $\underline{\mathbf{x}} = [\mathbf{x}_1; \mathbf{x}_2; \cdots; \mathbf{x}_n]$ a column vector of RV $\mathbf{x}_i$.

Mean. $\mathbb{E}\{\underline{\mathbf{x}}\} = [\mathbb{E}\mathbf{x}_1; \mathbb{E}\mathbf{x}_2; \cdots; \mathbb{E}\mathbf{x}_n]$

Variances.

$$\begin{aligned} V_{\mathbf{x}} &= \mathbb{E}\{\underline{\mathbf{x}} \odot \underline{\mathbf{x}}\} - \mathbb{E}\{\underline{\mathbf{x}}\} \odot \mathbb{E}\{\underline{\mathbf{x}}\} = \left[\mathbb{E}\mathbf{x}_1^2 - m_1^2; \mathbb{E}\mathbf{x}_2^2 - m_2^2\ ; \cdots;\ \mathbb{E}\mathbf{x}_n^2 - m_n^2\right] \\ &= \left[\sigma_{x_1}^2; \sigma_{x_2}^2; \cdots; \sigma_{x_n}^2\right] \end{aligned}$$

Covariance matrix.

$$\underline{\underline{C_{\mathbf{xx}}}} = \begin{bmatrix} C_{1,1} & \cdots & C_{1,n} \\ \vdots & \dots & \vdots \\ C_{n,1} & \cdots & C_{n,n} \end{bmatrix} = \mathbb{E}\{\underline{\mathbf{x}}\,\underline{\mathbf{x}}^t\} - (\mathbb{E}\{\underline{\mathbf{x}}\})\ (\mathbb{E}\{\underline{\mathbf{x}}\})^t$$

and where $C_{i,j} = \mathbb{E}\mathbf{x}_i\mathbf{x}_j - (\mathbb{E}\mathbf{x}_i)(\mathbb{E}\mathbf{x}_j)$.
The correlation matrix $\underline{\underline{R_{\mathbf{xx}}}}$ is similarly defined with elements $R_{i,j} = \mathbb{E}\mathbf{x}_i\mathbf{x}_j$.

[5] $(\Delta = b^2 - ac)$

2.3 Processes

2.3.1 Deterministic signals as a guideline

We first need to define the functional space of signals with finite energy :

$$f(t) \in L^2(\mathbb{R}) \ \text{ iff } \ E_f = \int_{-\infty}^{\infty} |f(t)|^2 \, dt < \infty$$

$L^2(\mathbb{R})$ is an Hilbert space with inner product $\langle f, g \rangle = \int f(t) \, g^*(t) \, dt$

Periodic signals. To handle periodic signals, which are not of finite energy, we introduce the concept of power :

$$\begin{array}{ll} \text{instantaneous power} & p_f(t) = |f(t)|^2 \\ \text{average power} & P_f(t,T) = \dfrac{1}{T} \displaystyle\int_{t-\frac{T}{2}}^{t+\frac{T}{2}} p_f(u) \, du \end{array}$$

A signal f is of finite power iff $P_f = \lim_{T\to\infty} P_f(t,T) < \infty$. Note that $E_f < \infty$ yields $P_f = 0$.

Fourier transform. If $f(t) \in L^2(\mathbb{R})$, the Fourier transform of f reads :

$$\mathcal{F}[f] = \tilde{f}(\nu) = \int_{-\infty}^{\infty} f(t) \, e^{-i2\pi\nu t} \, dt, \quad \text{with } \tilde{f} \in L^2(\mathbb{R})$$

and its inverse

$$f(t) = \mathcal{F}^{-1}[\tilde{f}] = \int_{-\infty}^{\infty} \tilde{f}(\nu) \, e^{i2\pi\nu t} \, dt.$$

We can interpret the FT as the inner product of the analyzed signal with $e_\nu = e^{i2\pi\nu t}$ the eigenfunction of all time invariant linear operators, and

$$\left\{ \begin{array}{l} f(t) = \displaystyle\int \langle f, e_\nu \rangle e_\nu(t) \, d\nu \\ \tilde{f}(\nu) = \displaystyle\int \langle f, \delta_u \rangle \, \delta_u(t) \, du \ \text{ where } \delta_u(t) = \delta(t-u) = \mathcal{F}[e^{i2\pi\nu u}] \end{array} \right.$$

Moreover, it is an isometry relation (**Parseval relation**) :

$$E_f = \langle f, f \rangle = \int |f(t)|^2 \, dt = \int |\tilde{f}(\nu)|^2 \, d\nu \quad ; \quad E_{fg} = \langle f, g \rangle = \int f(t) \, g^*(t) \, dt = \int \tilde{f}(\nu) \tilde{g}^*(\nu) \, du.$$

Energy (Power) Density Spectrum.

$$S_f(\nu) = |\tilde{f}(\nu)|^2$$

It is the frequency counterpart of instantaneous power $p_f(t)$, and $\int S_f(\nu) \, d\nu = E_f$

Remark. For signals of finite power, we introduce the intermediary quantity $S_{f,T}(\nu) = 1/T \, |\mathcal{F}[f(t) \, \Pi_T(t)]|^2$ and $S_f(\nu) = \lim_{T\to\infty} S_{f,T}(\nu)$, so that $\int S_f(\nu) \, d\nu = P_f$.

Correlation.

Theorem 4 (Wiener-Khintchine) *For finite energy signals $f \in L^2(\mathbb{R})$, we define the function*

$$C_f(\tau) := \int_{-\infty}^{\infty} S_f(\nu)\, e^{i2\pi\nu\tau}\, d\nu = \mathcal{F}^{-1}[S_f] = \int_{-\infty}^{\infty} f^*(t)\, f(t-\tau)\, d\tau.$$

Proof.

$$\begin{aligned}
\int S_f(\nu)\, e^{i2\pi\nu\tau}\, d\nu &= \int \tilde{f}(\nu)\tilde{f}^*(\nu)\, e^{i2\pi\nu\tau}\, d\nu \\
&= \int \tilde{f}(\nu) \left(\int f(t) e^{-i2\pi\nu t}\, dt \right)^* e^{i2\pi\nu\tau}\, d\nu \\
&= \int f^*(t) \int \tilde{f}(\nu)\, e^{i2\pi\nu(t+\tau)}\, d\nu\, dt \\
&= \int f^*(t)\, f(t+\tau)\, dt
\end{aligned}$$

We check easily that :
- $C_f(0) = E_f$
- $|C_f(\tau)| \leq C_f(0)$ (Schwarz' inequality)
- Concentration of C_f implies a spread out S_f on the frequency axis, and reciprocally (Gabor-Heisenberg principle)

Remark. For finite power signals,

$$\begin{aligned}
C_x(\tau) &= \lim_{T\to\infty} \frac{1}{T} \int (f(t)\Pi_T(t))^* \left(f(t+\tau)\Pi_T(t+\tau)\right) dt \\
&= \lim_{T\to\infty} \frac{1}{T} \int S_{f,T}(\nu)\, e^{i2\pi\nu\tau}\, d\nu \;=\; \int S_f(\nu)\, e^{i2\pi\nu\tau}\, d\nu \quad \text{(Wiener-Khintchine)}
\end{aligned}$$

Discussion. In a way, the correlation function $C_x(\tau)$ measures the averaged (along the time axis) product of $f(t)$ with $f(t+\tau)$. If t is considered an event of some probability space Ω and $\mathbf{f}(t)$ a realization of some associated RV (with same distribution for all $t \in \Omega$) , then $C_f(\tau)$ *"resembles"* the ensemble average $\mathbb{E}\{\mathbf{f}(t)\mathbf{f}(t+\tau)\}$ defining the correlation of two RVs.
For this correspondence to make sense, RV $\mathbf{f}$ needs to be **stationary**, then

$$\mathbb{E} \text{ for a RV} \quad \sim \quad \int \ldots dt \text{ for a continuous realization } f(t) \text{ of } \mathbf{f}$$

More generally though, we now investigate how it is possible to transpose to stochastic processes, notions such those of *spectrum density and correlation function...*

2.3.2 Stochastic processes

Definition 6 (Stochastic process)

A stochastic process is a rule that consist to assign to each event ω of a probability space Ω a function of time

$$\begin{aligned}
\Omega &\mapsto \mathbb{S} \text{ (some signal space)} \\
\omega &\mapsto \mathbf{x}(t;\omega) \quad \text{or } \mathbf{x}[n;\omega] \text{ in discrete time}
\end{aligned}$$

At a given time t, $\mathbf{x}(t,\omega)$ is considered as a RV and the following properties derive :

– First order properties

Distribution $F_{\mathbf{x}}(x,t) = P\{\mathbf{x}(t,\omega) \leq x\}$ and similarly for $p_{\mathbf{x}}(x,t)$

Mean $m(t) = \mathbb{E}\mathbf{x}(t,\omega)$

– Second order properties

Joint distribution $F(x_1,x_2,t_1,t_2) = P\{\mathbf{x}(t_1,\omega) \leq x_1 \text{ and } \mathbf{x}(t_2,\omega) \leq x_2)$

$$p(x_1,x_2,t_1,t_2) = \frac{\partial^2 F(x_1,x_2,t_1,t_2)}{\partial x_1 \partial x_2}$$

Auto-correlation $R_{\mathbf{x}} = \mathbb{E}\mathbf{x}(t_1,\omega)\mathbf{x}^*(t_2,\omega)$

Variance $R_{\mathbf{x}}(t,t) = \mathbb{E}|\mathbf{x}(t,\omega)|^2 - |m|^2(t)$

Auto-covariance $C_{\mathbf{x}}(t_1,t_2) = \mathbb{E}\mathbf{x}(t_1,\omega)\mathbf{x}^*(t_2,\omega) - m(t_1)m^*(t_2)$

– High order properties

Joint distribution $F_{\mathbf{x}}(x_1,\cdots,x_n\,;\,t_1,\cdots,t_n) = P\{\mathbf{x}(t_1) \leq x_1\ ,\ \cdots\ ,\ \mathbf{x}(t_n) \leq x_n\}$

– Equality in distribution : $\mathbf{x} \overset{d}{=} \mathbf{y} \;\Leftrightarrow\; F_{\mathbf{x}}(x,t) = F_{\mathbf{y}}(y,t)\ \forall\, t$

– Equality in mean square (MS) : $\mathbb{E}_\Omega|\mathbf{x}(t) - \mathbf{y}(t)|^2 = 0 \;\Leftrightarrow\; \mathbf{x} = \mathbf{y}$ almost surely (a.s.) or with probability 1 (which does not necessarily imply that for a given event ω, $X(t,\omega) = Y(t,\omega)$ a.s.)

Example. Given a set of Poisson points t_i with density λ, we form the process $\mathbf{x}(t) =$ {number of points t_i} within the period $(0,t)$. Then, it can be shown (cf [5], p.108) that

$$\mathbb{E}\mathbf{x}(t) = \lambda t \quad \text{and} \quad \mathbb{E}\mathbf{x}^2(t) = (\lambda t)^2 + (\lambda t)$$

Then

$$\mathbb{E}X(t_1)(X(t_2) - X(t_1)) = \mathbb{E}X(t_1).\mathbb{E}\{X(t_2) - X(t_1)\} = \lambda t_1 \cdot \lambda(t_2 - t_1)$$

since $(0,t_1)$ and (t_1,t_2) are two disjoint intervals. Therefore,

$$\mathbb{E}X(t_1)(X(t_2) - X(t_1)) = \mathbb{E}\{X(t_1)X(t_2)\} - \mathbb{E}X^2(t_1)) = R_{\mathbf{x}}(t_1,t_2) - R_{\mathbf{x}}(t_1,t_1)$$
$$\Leftrightarrow R_{\mathbf{x}}(t_1,t_2) = \lambda t_1 \cdot \lambda(t_2 - t_1) + R_{\mathbf{x}}(t_1,t_1) = \lambda t_1 \cdot \lambda(t_2 - t_1) + (\lambda t_1)^2 + (\lambda t_1) = \lambda^2 t_1 t_2 + \lambda t_1$$

More generally tho, if t_2 is not necessarily larger than t_1 : $R_{\mathbf{x}}(t_1,t_2) = \lambda^2 t_1 t_2 + \lambda \min(t_1,t_2)$

Autocorrelation properties. .

- $R_{\mathbf{x}}(t_1,t_2) = R^*_{\mathbf{x}}(t_2,t_1)$: hermitic symmetry
- Average power (at t fixed) : $R_{\mathbf{x}}(t,t) = \mathbb{E}|\mathbf{x}(t)|^2 \geq 0$
- The autocorrrelation function of a process is a definite positive function :

$$\forall a_i \text{ and } \forall a_j \quad \sum_{a_i,a_j} a_i a_j^* R_{\mathbf{x}}(t_i,t_j) \geq 0$$

(the converse is also true : given a positive definite function $R(t,s)$ we can find a process $\mathbf{x}$ with autocorrelation $R(t,s)$.

- In accordance with the correlation coefficient of two RV, $C(t_1,t_2) \leq \sqrt{C(t_1,t_1)C(t_2,t_2)}$.
- Linear system theorem.

  $$\text{If } Y(t) = \int g(t-u)X(u)\,du \;\; \text{then} \;\; R_{\mathbf{y}}(t_1,t_2) = \iint R_{\mathbf{x}}(t_1-u_1, t_2-u_2)\, g(u_1)\, g(u_2)\, du_1\, du_2$$

  Example. Let $\mathbf{z}(t) = \frac{d}{dt}\mathbf{x}(t) = \sum_i \delta(t-t_i)$ where $\mathbf{x}$ is a counting Poisson process and the t_i's the corresponding Poisson instants. Thus applying the Linear System property above, it follows that

  $$R_{\mathbf{z}}(t_1,t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} R_{\mathbf{x}}(t_1,t_2) = \frac{\partial}{\partial t_1}\lambda^2 t_1 + \lambda U(t_1-t_2) = \lambda^2 + \lambda\delta(t_1-t_2)$$

2.3.3 Stationary processes

An important class of processes is the class of processes whose statistics properties are time-shift invariant.

Strict-Sense Stationarity (SSS). A process $\mathbf{x}$ is strict-sense stationary if $\mathbf{x}(t)$ and $\mathbf{x}(t+c)$ are equal in distribution, i.e.

$$f_{\mathbf{x}}(x_1,\cdots,x_n;t_1,\cdots,t_n) = f_{\mathbf{x}}(x_1,\cdots,x_n;t_1+c,\cdots,t_n+c) \quad \forall c \;\text{ and }\; \forall n$$

In particular, for the second order statistic :

$$f_{\mathbf{x}}(x_1,x_2;t_1,t_2) = f_{\mathbf{x}}(x_1,x_2;t_1-t_1,t_2-t_1) = f_{\mathbf{x}}(x_1,x_2;t_2-t_1)$$
$$\Rightarrow \mathbb{E}\{\mathbf{x}(t_1)\mathbf{x}(t_2)\} = R_{\mathbf{x}}(t_2-t_1)$$

This property is in general very stringent on $\mathbf{x}$ and difficult to satisfy (and verify). Often then, the concept of stationarity is soften.

Wide-Sense Stationarity (WSS). A process $\mathbf{x}(t,\omega)$ is wide-sense stationary if its mean is constant :

$$\mathbb{E}\mathbf{x}(t) = m$$

and its autocorrelation depends only on $\tau = t_2 - t_1$:

$$\mathbb{E}\{\mathbf{x}(t+\tau)\mathbf{x}(t)\} = R(\tau)$$

In particular, the average power of $\mathbf{x}$, $\mathbb{E}|\mathbf{x}(t)|^2 = R(0)$ and is independent of t.

A SSS process is also WSS. While the converse is false in general. it applies if $\mathbf{x}$ is a normal process.

Conversely, non-stationarity is a non-property which in practice is very difficult to characterize. It comes to identifying any type of rupture, or smooth change in the statistics of a process, whereas in most situations only one realization (observation) of the random process is available.

Application example : Sketches. We describe a stationarity versus non-stationarity test applied to the detection of anomalies in Internet traffic flows[6]. The idea is to (artificially) generate, using hashing functions, *almost independent realizations* of the stochastic process defined by the number of new flows (or connections) arriving within a period Δ_t.

[6]"Identification d'anomalies statistiques dans le trafic internet par projections aléatoires multirésolutions", P. Borgnat, G. Dewaele, P. Abry. Proceeding of the XXIème colloque GRETSI, Troyes, France, 2007.

Analyzed signal – $\mathbf{x}(t)$: Number of starting flows (SYN) appearing on an aggregated link, within a time window of length Δ_t. Observation duration is $T \gg \Delta_t$. Each flow is identified by an IPsrc and an IPdest (plus possibly ports numbers).

Goal – To identify anomalous flows emitted by a specific source (IPsrc) or received by a specific destination (IPdest). Both cases corresponding respectively to :

1. an intrusive scan
2. a deny of service

Difficulties – An exhaustive search of a pathological source or destination lies in a high dimensional space (typically 2^{96}). Internet traffic is naturally non-stationary, which does not necessarily correspond to any type of attack.

Overall idea – The approach works in two steps :

1. Partition the aggregated traffic into disjoint classes of flows. This allows for isolating the anomalous flow into a trace that gathers less flow heterogeneity, and thus accentuates the statistic discrepancy between normal traces (encompassing only regular flows) and irregular traces (containing at least one abnormal flow).
   Random projections (sketches) are used :

   $$x_m(t) = \mathcal{H}[\mathbf{x}](t), \quad \text{for } m = 1, \dots M$$

   The hashing function $\mathcal{H}$ is a projector that has m possible outputs, and it filters the aggregated trace $x(t)$ according to IPsrc or / and IPdest. Each of the outputs $x_m(t)$ represents a particular realization of a stochastic process $\mathbf{x}_m$, all being equal in law if no anomalous flow falls in the cluster $\mathcal{C}_m$. If one (or more) anomalous flow is projected in the class $\mathcal{C}_m$, statistics of this latter should deviate form the common statistics of the standard classes.
2. Multiply the number of independent realizations of each stochastic process $\mathbf{x}_m$. For this, use different orthogonal random projectors $\mathcal{H}_n,\ n = 1, \cdots, N$.

The statistical Gamma law of normal traffic :

$$\Gamma_{\alpha,\beta}(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$$

is inferred (i.e. parameters $\hat{\alpha}$ and $\hat{\beta}$ are estimated) by time average and sketches average (view as ensemble averages) of the $x_{m,n}(t)$ processes . Then, for each series $x_{m,n}(t)$, the following hypothesis test is performed :

$$\begin{cases} P_{\hat{\alpha},\hat{\beta}}\{\mathbf{x} > x_{m,n}(t)\} \leq \lambda & : \text{anomaly,} \\ P_{\hat{\alpha},\hat{\beta}}\{\mathbf{x} > x_{m,n}(t)\} > \lambda & : \text{normal,} \end{cases}$$

2.3.4 Power spectrum in the stationary case

In signal processing, a host of interesting properties are more easy to identify or to characterize with the spectral representation in the frequency domain. Spectra are associated with Fourier transform, which in the deterministic case, represents the signal as a superposition of complex exponential. We shall introduce the concept of spectrum for stationary stochastic processes, starting with a first approach which consists to apply Fourier transforms to average quantities (thus deterministic for WSS processes). Later on, we shall rely on the case of non-stationary processes to illustrate spectra as a superposition of complex exponentials.

Definition 7 (Power spectrum) *The power spectrum (or spectral density) of a WSS process* $\mathbf{x}(t)$ *is the Fourier transform* $S(\nu)$ *of its autocorrelation function* $R(\tau) = \mathbb{E}\{(x+\tau)\mathbf{x}(t)$

$$S(\nu) = \int_{-\infty}^{\infty} R(\tau)\, e^{-i2\pi\nu\tau}\, d\tau$$

Because $R(\tau) = R^*(-\tau)$ (hermitic property), $S(\nu)$ is a real function of frequency ν. Conversely,

$$R(\tau) = \int_{-\infty}^{\infty} S(\nu)\, e^{i2\pi\nu\tau}\, d\nu$$

Example. Let t_i be a set of Poisson points, and denote $\mathbf{n}(t_1, t_2)$ the corresponding Poisson RV within interval (t_1, t_2), for which we have $P\{\mathbf{n}(t_1, t_2) = k\} = e^{-\lambda(t_2-t_1)}(\lambda(t_2 - t_1))^k/k!$. Form the process

$$\mathbf{x}(t) = \begin{cases} +1 & \text{if } \mathbf{n}(0,t) \text{ is even} \\ -1 & \text{if } \mathbf{n}(0,t) \text{ is odd} \end{cases}$$

$\mathbf{x}(t)$ is a binary signal (digital communication) which commutes between $+1$ and -1, with probabilties

$$\begin{aligned} P\{\mathbf{x}(t) = 1\} &= P\{\mathbf{n}(0,t) = 0\} + P\{\mathbf{n}(0,t) = 2\} + \dots \\ &= e^{-\lambda t} + e^{-\lambda t}(\lambda t)^2/2! + \dots = e^{-\lambda t}\left(1 + (\lambda t)^2/2! + \cdots\right) \\ &= e^{-\lambda t}\cosh(\lambda t) \\ P\{\mathbf{x}(t) = -1\} &= P\{\mathbf{n}(0,t) = 1\} + P\{\mathbf{n}(0,t) = 3\} + \dots \\ &= e^{-\lambda t}\sinh(\lambda t) \end{aligned}$$

We want to determine $R_{\mathbf{x}}(t_1, t_2)$, which reads :

$$\begin{aligned} R_{\mathbf{x}}(t_1, t_2) &= \mathbb{E}\{\mathbf{x}(t_1)\mathbf{x}(t_2)\} = P\{\mathbf{x}(t_1) = 1, \mathbf{x}(t_2) = 1\} - P\{\mathbf{x}(t_1) = 1, \mathbf{x}(t_2) = -1\} \\ &\quad - P\{\mathbf{x}(t_1) = -1, \mathbf{x}(t_2) = 1\} + P\{\mathbf{x}(t_1) = -1, \mathbf{x}(t_2) = -1\} \end{aligned}$$

Let us assume $t_1 < t_2$, if $\mathbf{x}(t_1) = +1$, then $\mathbf{x}(t_2) = 1$ if and only if $\mathbf{n}(t_1, t_2)$ is even. Thus, by independence of RV $\mathbf{n}(0, t_1)$ and $\mathbf{n}(t_1, t_2)$, and posing $t = t_2 - t_1$,

$$\begin{aligned} &P\{\mathbf{x}(t_1) = 1, \mathbf{x}(t_2) = 1\} = P\{\mathbf{x}(t_1) = 1\} \cdot P\{\mathbf{x}(t) = 1\} = e^{-\lambda t_1}\cosh(\lambda t_1) e^{-\lambda t}\cosh(\lambda t) \\ &P\{\mathbf{x}(t_1) = 1, \mathbf{x}(t_2) = -1\} = P\{\mathbf{x}(t_1) = 1\} \cdot P\{\mathbf{x}(t) = -1\} = e^{-\lambda t_1}\cosh(\lambda t_1) e^{-\lambda t}\sinh(\lambda t) \\ &P\{\mathbf{x}(t_1) = -1, \mathbf{x}(t_2) = 1\} = P\{\mathbf{x}(t_1) = -1\} \cdot P\{\mathbf{x}(t) = -1\} = e^{-\lambda t_1}\sinh(\lambda t_1) e^{-\lambda t}\sinh(\lambda t) \\ &P\{\mathbf{x}(t_1) = -1, \mathbf{x}(t_2) = -1\} = P\{\mathbf{x}(t_1) = -1\} \cdot P\{\mathbf{x}(t) = 1\} = e^{-\lambda t_1}\sinh(\lambda t_1) e^{-\lambda t}\cosh(\lambda t) \end{aligned}$$

Substituting in $R(t_1, t_2)$, we get

$$R(t_1, t_2) = e^{-2\lambda|t_1 - t_2|}$$

The process $\mathbf{x}(t)$ is stationnary (at least WSS) with $R(\tau) = e^{-2\lambda|\tau|}$. Then, by Fourier transform, we obtain

$$S(\nu) = \frac{4\lambda}{4\lambda^2 + \nu^2}$$

which gives the necessary channel bandwidth to transfer a random binary signal at rate λ.

Some properties.

- $S(\nu) \geq 0$ because $R(\tau)$ is a definite positive function. Moreover the average power $R(0) = \mathbb{E}|\mathbf{x}(t)|^2 = \int S(\nu)\, du$.
- Given an arbitrary positive function $S(\nu)$, we can always find a process $\mathbf{x}(t)$ with power spectrum $S(\nu)$.
- $\mathbf{x}(t)$ being a WSS process, the response $\mathbf{y}(t) = \int \mathbf{x}(t - \tau)h(\tau)\, d\tau$ of a linear system is such that

$$\begin{gathered} R_{\mathbf{y}}(\tau) = R_{\mathbf{x}}(\tau) \star h(\tau) \star h^*(-\tau) \\ S_{\mathbf{y}}(\nu) = S_{\mathbf{x}}(\nu)|H(\nu)|^2 \end{gathered}$$

– The integrated spectrum is defined as the integral

$$F(\nu) = \int_{-\infty}^{\nu} S(\nu)\, d\nu$$

$F(\nu)$ is necessarily a non decreasing function of ν, and formally, it allows to express the autocorrelation $R(\tau)$ as a Fourier-Stieltjes integral (integration on a measure)

$$R(\tau) = \int_{-\infty}^{\infty} e^{i2\pi\nu\tau}\, dF(\nu)$$

Moreover, $F(\nu)$ avoids the use of singularity fuctions in the spectral representation of $R(\tau)$. If $S(\nu)$ contains pulses of the type $\delta(\nu - \nu_i)$, then $F(\nu)$ is simply discontinuous at ν_i.

2.3.5 Harmonic analysis of stochastic processes

Given a stochastic process $\mathbf{x}(t)$, its spectral representation is defined as follows

$$\mathbf{x}(t) = \int_{-\infty}^{\infty} e^{-i2\pi\nu t}\mathbf{X}(\nu)\, d\nu \quad \text{and} \quad \mathbf{X}(\nu) = \int_{-\infty}^{\infty} e^{i2\pi\nu t}\mathbf{x}(t)\, dt$$

whith equalities in MS.

Loève condition : The spectral representation of $\mathbf{x}(t)$ hold, if and only if

$$\mathbb{E}\{\mathbf{X}(\nu_1)\mathbf{X}^*(\nu_2)\} := \Gamma(\nu_1, -\nu_2) \quad \text{is such that} \quad \int\int_{-\infty}^{\infty} |\Gamma(\nu_1, -\nu_2)| d\nu_1\, d\nu_2 < \infty$$

Theorem 5 (Wiener-Khintchine) . *If $\mathbf{X}(\nu)$ satisfies the Loeve condition, then*

$$R_{\mathbf{x}}(t_1, t_2) = \int\int e^{i2\pi\nu_1 t_1 + i2\pi\nu_2 t_2}\Gamma(\nu_1, \nu_2) d\nu_1 d\nu_2$$

In this relation, the spectral increment $\Gamma(\nu_1, \nu_2)$ plays the role of a spectrum, similar to that of a spectrum for WSS processes.

Formally, the equivalence holds, but in the general case, when no specific structure is imposed on $R(t_1, t_2)$ (such as stationarity), the notion of spectrum is of little use for non-stationary processes.

One solution to circumvent the difficulty arising with non-stationary processes, is to resort to a decomposition of x on a more adapted basis than the Fourier basis, e.g. wavelet bases.

We can show that if $\mathbf{x}$ is non-stationary with average power $\mathbb{E}|\mathbf{x}(t)|^2 = R(t, t) = p(t)$, then $\mathbf{X}(\nu)$ is a stationary process and its autocorrelation equals the Fourier transform $Q(\nu)$ of $q(t)$, i.e.

$$\mathbb{E}\{\mathbf{X}(\nu + \alpha)\mathbf{X}(\alpha)\} = \Gamma(\nu + \alpha, -\alpha) = Q(\nu)$$

Conversely, going back to stationary processes, we shown that $\Gamma(\nu_1, \nu_2)$ takes on a diagonal form, localized on the frequency axis, since then

$$\begin{aligned}
R(t_1, t_2) = R(t_2 - t_1) \quad &= \quad \int\int R(t_1 - t_2)\, e^{-i2\pi(\nu_1 t_1 + \nu_2 t_2)}\, dt_1\, dt_2 \\
&= \quad \int e^{-i2\pi(\nu_1 + \nu_2)t_2} \int R(\tau) e^{-i2\pi\nu_1\tau}\, d\tau\, dt_2 \\
\Rightarrow &\qquad \Gamma(\nu_1, \nu_2) = S(\nu_1) \int e^{-i2\pi(\nu_1 + \nu_2)t_2}\, dt_2 = S(\nu_1)\delta(\nu_1 + \nu_2) \\
\Rightarrow &\qquad \mathbb{E}\{\mathbf{X}(\nu_1)\mathbf{X}^*(\nu_2)\} = S(\nu_1)\delta(\nu_1 - \nu_2)
\end{aligned}$$

This shows that the Fourier transform of a stationary process is non-stationary white noise with average power $S(\nu)$. The converse is also true, that is, a process $\mathbf{x}$ in the spectral representation is WSS iff $\mathbb{E}\mathbf{X}(\nu) = 0$ for $\nu \neq 0$ and $\mathbb{E}\{\mathbf{X}(\nu_1)\mathbf{X}^*(\nu_2)\} = Q(\nu_1)\delta(\nu_1 - \nu_2)$

In order now to express the spectral representation of a WSS process in terms of a Fourier-Stieltjes integral, let us introduce the integrated process

$$\mathbf{Z}(\nu) = \int_0^\nu \mathbf{X}(\lambda)\, d\lambda \quad \Leftrightarrow \quad d\mathbf{Z}(\nu) = \mathbf{X}(\nu) d\nu$$

It comes from the specific diagonal form of $\Gamma(\nu_1, \nu_2)$, that the process $\mathbf{Z}(\nu)$ has orthogonal (uncorrelated) increment : For any $\nu_1 < \nu_2 < \nu_3 < \nu_4$:

$$\begin{aligned}
\mathbb{E}\{[\mathbf{Z}(\nu_2) - \mathbf{Z}(\nu_1)][\mathbf{Z}(\nu_4) - \mathbf{Z}(\nu_3)]^*\} &= \mathbb{E}\{\int_{\nu_1}^{\nu_2} \mathbf{X}(\lambda_1) d\lambda_1 \int_{\nu_3}^{\nu_4} \mathbf{X}^*(\lambda_2) d\lambda_2\} \\
&= \int_{\nu_1}^{\nu_2} \int_{\nu_3}^{\nu_4} \mathbb{E}\{\mathbf{X}(\lambda_1)\mathbf{X}^*(\lambda_2)\}\, d\lambda_1\, d\lambda_2 \\
&= \int_{\nu_1}^{\nu_2} \int_{\nu_3}^{\nu_4} S(\lambda_1)\delta(\lambda_1 - \lambda_2)\, d\lambda_1\, d\lambda_2 = 0 \quad \text{(disjoint intervals)}
\end{aligned}$$

and, similarly

$$\mathbb{E}\{|\mathbf{Z}(\nu_2) - \mathbf{Z}(\nu_1)|^2\} = \int_{\nu_1}^{\nu_2} S(\nu)\, d\nu$$

Taking limits $\nu_1 = \omega_1$, $\nu_2 = \omega_1 + d\omega_1$, $\nu_3 = \omega_2$, $\nu_4 = \omega_2 + d\omega_2$, we get

$$\begin{aligned}
\mathbb{E}\{d\mathbf{Z}(\omega_1) d\mathbf{Z}^*(\omega_2)\} &= \mathbb{E}\{\mathbf{X}(\omega_1)\mathbf{X}^*(\omega_2)\}\, d\omega_1 d\omega_2 = \Gamma(\omega_1, -\omega_2)\, d\omega_1 d\omega_2 \quad (= d^2\Gamma(\omega_1, -\omega_2)) \\
&= S(\omega_1)\delta(\omega_1 - \omega_2)\, d\omega_1 d\omega_2 \quad \text{(for WSS processes)} \\
&= dF(\omega_1)\delta(\omega_1 - \omega_2)\, d\omega_2
\end{aligned}$$

The spectral representation of $\mathbf{x}(t)$ can be written with the following Fourier-Stieltjes intergral :

$$\mathbf{x}(t) = \int_{-\infty}^{\infty} e^{i2\pi\nu t}\, d\mathbf{Z}(\nu)$$

A doubly orthogonal decomposition,

$$\begin{aligned}
&\langle e^{i2\pi\omega_1 t}, e^{i2\pi\omega_2 t}\rangle = \delta(\omega_1 - \omega_2) && \text{orthogonal basis} \\
\text{and} \quad &\mathbb{E}\{d\mathbf{Z}(\omega_1) d\mathbf{Z}^*(\omega_2)\} = 0, \quad \text{for all } \omega \neq \omega_2 && \text{orthogonal spectral increments}
\end{aligned}$$

and with

$$\mathbb{E}\{|d\mathbf{Z}(\omega)|^2\} = dF(\omega) d\omega$$

It is this last relation that permits to define the spectrum $S(\omega)$ of WSS process $\mathbf{x}(t)$, in terms of the increment process $d\mathbf{Z}(\omega)$.

2.3.6 Symmetry properties of stochastic processes

Stationarity. Stationarity (strict- or wide-sense) is a type of symmetry property that compels time invariance to the statistics (of all or restricted orders) of a process.

Processes with independent increments.

Definition 8 $\mathbf{x}(t)$ *is a stochastic process with independent increments* $\mathbf{z}(t,\tau) = \mathbf{x}(t+\tau) - \mathbf{x}(t)$ *if*

$$\forall t \text{ and } \forall \tau > 0, \quad \mathbf{z}(t,\tau) \text{ is independent of } \mathbf{x}(u), \ \forall u \leq t.$$

Example 1. The counting Poisson process $\mathbf{x}(t) := \mathbf{n}(0,t)$ has independent increments.

Example 2. A process $\mathbf{M}(t)$ is a martingale if

$$\mathbb{E}\{\mathbf{M}(t)|\{\mathbf{M}(u), \ u \leq s\}\} = \mathbf{M}(s) \ \forall s \leq t \quad \Rightarrow \mathbb{E}\{\mathbf{M}(t) - \mathbf{M}(s)|\{\mathbf{M}(u), \ u \leq s\}\} = 0, \ \forall s \leq t$$

Consequently, a martingale has uncorrelated increments (and independent if M is moreover gaussian).

Example 3. A process $\mathbf{x}(t)$ has infinitely divisible law if

$$\mathbf{x}(t) = \sum_{i=0}^{N} \mathbf{Z}(u_i, \tau_i)$$

where $\mathbf{Z}(u_i, \tau_i)$ are independent random variables satisfying :

$$\forall \varepsilon > 0, \ \forall \eta > 0, \quad P\{|\mathbf{Z}(u_i, \tau_i)| > \varepsilon\} < \eta \ \ \text{(continuity in probability)}$$

$$\text{and } \begin{cases} \tau_i = u_{i+1} - u_i < \mu \text{ (sufficiently small)} \\ 0 = u_0 < u_1 < \cdots < u_N = t \end{cases}$$

Examples of infinitely divisible laws are : Gaussian, $\alpha-$stable laws, Poisson, ...

Theorem 6 (Lévy) *Any continuous time process* $\mathbf{x}(t)$ *can be decomposed as follows*

$$\mathbf{x}(t) = f(t) + \sum_{n|t_n \leq t} \mathbf{y}_{t_n} + \mathbf{x}_2(t)$$

where $f(t)$ *is a deterministic function,* $\mathbf{y}_{t_n}$ *a discrete process and* $\mathbf{x_2}(t)$ *a process with infinitely divisible law.*

Processes with stationary increments.

Definition 9 $\mathbf{x}(t)$ *is a stochastic process with stationary increments* $\mathbf{z}(t,\tau) = \mathbf{x}(t+\tau) - \mathbf{x}(t)$ *if*

$\forall t$ and $\forall \tau > 0$, $\mathbf{z}(t,\tau)$ is a (strict- or wide-sense) stationary process (with respect to t).

Example. Random Walk.

$$\mathbf{x}_n = \sum_{i=1}^{n} \mathbf{z}_i \ \text{ where the increments } \mathbf{z}_i\text{'s are i.i.d., } \mathbf{z}_i = \begin{cases} +s \text{ with probabilit } p \\ -s \text{ with probabilit } q = 1-p \end{cases}$$

We then have $\mathbf{x}_n = ks - (n-k)s = (2k-n)s = ms$, and

$$P\{\mathbf{x}_n = ms\} = C_n^k p^k \, q^{n-k} = C_n^k \frac{1}{2^n} \text{ if p=q=1/2}$$

and

$$\mathbb{E}\mathbf{x}_n = \sum_{i=1}^{n} \mathbb{E}\mathbf{z}_i = 0 \quad \text{and} \quad \mathbb{E}|\mathbf{x}_n|^2 = \sum_{i=1}^{n} |\mathbf{z}_i|^2 = ns^2.$$

When $n \to 0$, $C_n^k p^k q^{n-k} \simeq (2\pi npq)^{-\frac{1}{2}}\, e^{-(k-np)^2/2npq}$ (DeMoivre-Laplace thm), then

$$P\{\mathbf{x}_n = ms\} \simeq \frac{2}{\sqrt{n}\sqrt{2\pi}}\, e^{-\frac{m^2}{2n}} \sim \mathcal{N}(0, \sqrt{n})$$

Now, posing $t = nT$ (continuous time), $\mathbf{x_n} \rightsquigarrow \mathbf{x}(t)$, such that

$$\begin{cases} \mathbb{E}\mathbf{x}(t) = 0, \ \forall t \\ \mathbb{E}|\mathbf{x}(t)|^2 = ns^2 = \frac{s^2}{T} t \ \forall t \quad \text{(non-stationary)} \end{cases}$$

The limit process $\mathbf{x}(t)$ when $s \to 0$, $T \to 0$, with $s^2/T \to 2D^2$ finite, is called **Wiener process** or **(ordinary) Brownian motion**, and

$$\begin{aligned} P\{\mathbf{x}(t) \leq x = ms\} &= \int_{-\infty}^{x/s} \frac{2}{\sqrt{2\pi n}}\, e^{-\frac{m^2}{2n}}\, dm \\ &= \int_{-\infty}^{x} \frac{2}{\sqrt{2\pi s^2 t/T}}\, e^{-\frac{y^2}{2s^2 t/T}}\, dy \quad \text{(posing } y = ms\text{)} \\ &= \int_{-\infty}^{x} \frac{2}{\sqrt{2\pi}\sqrt{2D^2 t}}\, e^{-\frac{y^2}{2(2D^2 t)}}\, dy \end{aligned}$$

Finally,

$$p_{\mathbf{x}(t)}(x) \sim \mathcal{N}(0, D\sqrt{2t}).$$

Regarding autocorrelation of the process, considering that increments are independent (by construction) :

$$\mathbb{E}\{(\mathbf{x}(t_2) - \mathbf{x}(t_1))\, \mathbf{x}(t_1)\} = \mathbb{E}\,(\mathbf{x}(t_2) - \mathbf{x}(t_1))\, \underbrace{\mathbb{E}\mathbf{x}(t_1)}_{=0} = 0,$$

it comes that

$$R(t_2, t_1) = \mathbb{E}\{\mathbf{x}(t_2)\mathbf{x}(t_1)\} = \mathbb{E}|\mathbf{x}(t_1)|^2 = 2D^2 t_1$$

and conversely, if $t_2 < t_1$,

$$R(t_1, t_2) = 2D^2 \min(t_1, t_2).$$

3 Wavelet decompositions

3.1 Gabor transform : an impossible paradigm for orthogonal basis

3.2 Continuous wavelet transform and frames

3.3 Discrete wavelet transform

3.3.1 The Haar system : an introductory illustration

3.3.2 Multiresolution analysis

A multiresolution analysis consists of a sequence of successive approximation spaces V_j satisfying the embedding relation

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset \cdots \subset V_j \subset V_{j+1} \subset \cdots \tag{5}$$

with

(i) $\displaystyle\bigcap_{m=-\infty}^{m=\infty} V_m = \{0\}$

(ii) $\displaystyle\bigcup_{m=-\infty}^{m=\infty} V_m$ is dense in $L^2(\mathbb{R})$

(iii) $x(t) \in V_m \;\Leftrightarrow\; x(2t) \in V_{m+1}$

(iv) there exists a function $\phi(t)$ such that the collection $\{\phi(t-n),\, n \in \mathbb{Z}\}$ constitutes a basis for V_0.

$$\forall x \in V_0, \quad \int |x(t)|^2\, dt = \sum_k \left| \int x(t)\, \phi(t-k)\, dt \right|^2$$

From *(iii)* and *(iv)*, we deduce that $\{\phi_{mn}(t) = 2^{m/2}\phi(2^m t - n),\, \forall n \in \mathbb{Z}\}$ is a basis for V_m. If we denote P_j the orthogonal projection operator onto V_j,

$$P_j x(t) = \sum_k \langle x, \phi_{jk} \rangle \phi_{jk}(t),$$

then, *(ii)* ensures that $\lim_{j\to\infty} P_j x = x$, for all $x \in L^2(\mathbb{R})$.

Let us know consider the orthogonal complement space W_m of V_m in V_{m+1},

$$V_{m+1} = V_m \oplus W_m.$$

The basic principle of multiresolution analysis is that there exists a function ψ such that

$$P_{m+1}x = P_m x + \sum_k \langle x, \psi_{m,k} \rangle \psi_{m,k}(t),$$

which means that the set $\{\psi_{m,k}(t) = 2^{m/2}\psi(2^m t - k),\, k \in \mathbb{Z}\}$ constitutes an orthogonal basis of space W_m. Directly inherited from the properties *(i)–(iv)*, we get :

$$(i) \qquad W_j \perp W_{j'} \text{ for any } j \neq j'$$

$$(ii) \qquad V_j = V_J + \bigoplus_{m=0}^{m=J-j+1} W_{J-m}$$

$$L^2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j$$

$$(iii) \qquad x(t) \in W_m \;\Leftrightarrow\; x(2t) \in W_{m+1}$$

$$x(t) \in W_m \;\Leftrightarrow\; x(2^{-m}t) \in W_0$$

(iv) there exists a function $\psi(t)$ such that the collection $\{2^{m/2}\psi(2^m t - n),\, m \in \mathbb{Z},\, n \in \mathbb{Z}\}$ constitutes a basis for $L^2(\mathbb{R}$.

The scaling functions $\{2^{j/2}\phi(2^j - k),\ f \in \mathbb{Z}\}$ is an orthonormal basis of V_j, the approximation space at scale (or resolution) j, and the wavelet functions $\{2^{j/2}\psi(2^j - k),\ f \in \mathbb{Z}\}$ form an orthonormal basis of W_j, the detail space at scale j.
We shall now give the generic construction of a wavelet system $\{\phi, \psi\}$ generating a multiresolution analysis.

Let $\{\phi_{0,n}\}_{n\in\mathbb{Z}}$ a basis of $V_0 \subset V_1$, then there exists a series of coefficients $h[n]$ such that

$$h[n] = \int \phi(t)\, 2^{1/2}\phi(2t-n)\, dt \quad ; \quad \sum_n |h[n]|^2 = 1 \ \ (\text{since } \phi \text{ is of norm } 1)$$

leading to the so-called *two-scales relation* :

$$\phi(t) = 2^{1/2} \sum_n h[n]\phi(2t-n) \ \ (\text{as } \{\phi_{1,n}\}_{n\in\mathbb{Z}} \text{ is a basis of } V_1).$$

Taking the Fourier transform, we get

$$\Phi(\nu) = 2^{-1/2} \sum_n h[n] \Phi\left(\frac{\nu}{2}\right) e^{i2\pi\frac{\nu}{2}n}$$

where we pose $H(\nu) = 2^{-1/2}\sum_n h[n]e^{i2\pi\nu n}$ the discrete Fourier transform of the series $2^{-1/2}h[n]$, a naturally one-periodic function. Then,

$$\Phi(\nu) = H\left(\frac{\nu}{2}\right) \Phi\left(\frac{\nu}{2}\right). \tag{6}$$

and moreover

$$\begin{aligned}
\delta_{0,k} &= \int \phi(t)\,\phi(t-k)\,dt \ \ (\text{orthogonality}) \\
&= \int |\Phi(\nu)|^2\, e^{i2\pi\nu k}\, d\nu \ \ (\text{Moyal isometry relation}) \\
&= \int_0^1 \left(\sum_{n=-\infty}^{\infty} |\Phi(\nu+n)|^2\right) e^{i2\pi\nu k}\, d\nu \ \ (\text{partition of the real axis})
\end{aligned}$$

$$\Longleftrightarrow \qquad \sum_{n=-\infty}^{\infty} |\Phi(\nu+n)|^2 \;=\; 1, \ \forall\nu \ \ (\text{because } \textstyle\int_0^1 e^{i2\pi\nu k}\, d\nu = \delta_k). \tag{7}$$

Therefore,

$$(6) \Leftrightarrow \sum_n |\Phi(\nu+n)|^2 = \sum_n \left|\Phi\left(\frac{\nu+n}{2}\right)\right|^2 \left|H\left(\frac{\nu+n}{2}\right)\right|^2 = 1, \quad \forall\nu$$

in particular for $\xi = \nu/2$ and splitting the sum in the RHS into odd and even integers, we obtain

$$\begin{aligned} \textstyle\sum_k |H(\xi+k)|^2|\Phi(\xi+k)|^2 + \sum_k |H(\xi+k+\frac{1}{2})|^2|\Phi(\xi+k+\frac{1}{2})|^2 = 1 \\ |H(\xi)|^2 \underbrace{\sum_k |\Phi(\xi+k)|^2}_{=1\ (Eq.(7))} + |H(\xi+\tfrac{1}{2})|^2 \underbrace{\sum_k |\Phi(\xi+k+\frac{1}{2})|^2}_{=1\ (Eq.(7))} = 1 \\ \Longleftrightarrow \quad |H(\xi)|^2 + |H(\xi+\tfrac{1}{2})|^2 = 1, \ \forall\xi \end{aligned} \tag{8}$$

Similarly for $\psi(t) \in W_0 \subset V_1$, we have

$$\psi(t) = 2^{1/2}\sum_n g[n]\phi(2t-n), \text{ where we have posed } g[n] = \int \psi(t)\, 2^{1/2}\phi(2t-n)\, dt.$$

Analogously, taking Fourier transforms of both sides of this relation,

$$\Psi(\nu) = G\left(\frac{\nu}{2}\right)\Phi\left(\frac{\nu}{2}\right). \tag{9}$$

As for $H(\nu)$, the frequency response $G(\nu) = \sum_n \sqrt{2} g[n] e^{i2\pi\nu n}$ is a one-periodic function.
As, $W_0 \perp V_0$,

$$\begin{aligned} 0 &= \int \psi(t)\phi(t-k)\, dt = \int \Psi(\nu)\Phi^*(\nu)\, e^{i2\pi\nu k}\, d\nu \\ &= \int_0^1 \left(\sum_{n=-\infty}^{\infty} \Psi(\nu+n)\Phi^*(\nu+n)\right) e^{i2\pi\nu k}\, d\nu \\ \Longleftrightarrow \quad & \sum_{n=-\infty}^{\infty} \Psi(\nu+n)\Phi^*(\nu+n) = 0 \end{aligned} \tag{10}$$

Substituting relations (6) and (9) into this sum, yields :

$$\sum_n \Phi\left(\frac{\nu+n}{2}\right) G\left(\frac{\nu+n}{2}\right) \Phi^*\left(\frac{\nu+n}{2}\right) H^*\left(\frac{\nu+n}{2}\right) = 0$$

Posing $\xi = \nu/2$ and splitting the sum according to odd and even indices,

$$\begin{aligned} \textstyle\sum_k |\Phi(\xi+k)|^2 G(\xi+k) H^*(\xi+k) + \sum_k \left|\Phi\left(\xi+k+\frac{1}{2}\right)\right|^2 G\left(\xi+k+\frac{1}{2}\right) H^*\left(\xi+k+\frac{1}{2}\right) = 0 \\ G(\xi)H^*(\xi) \underbrace{\sum_k |\Phi(\xi+k)|^2}_{=1\ (Eq.7))} + G\left(\xi+\tfrac{1}{2}\right) H^*\left(\xi+\tfrac{1}{2}\right) \underbrace{\sum_k \left|\Phi\left(\xi+k+\frac{1}{2}\right)\right|^2}_{=1\ (Eq.7))} = 0 \\ \Longleftrightarrow \quad G(\xi)H^*(\xi) + G\left(\xi+\tfrac{1}{2}\right) H^*\left(\xi+\tfrac{1}{2}\right) = 0, \ \forall\xi \end{aligned} \tag{11}$$

From this last relation, we conclude that, for $\{\phi, \psi\}$ to generate a multiresolution analysis scheme, H and G have to form a pair of quadrature mirror filters (QMF).
Given H, the solution G of (11) is of the following general form

$$G(\nu) = \lambda(\nu)H^*\left(\nu+\frac{1}{2}\right),$$

where λ is an arbitrary one-periodic function verifying the condition

$$\lambda(\nu) + \lambda\left(\nu + \frac{1}{2}\right) = 0.$$

A particularly interesting solution is given by the specific choice :

$$\lambda(\nu) = -e^{i2\pi\nu}, \tag{12}$$

leading to :

$$\begin{aligned}
G(\nu) &= e^{i2\pi\left(\nu+\frac{1}{2}\right)} H^*\left(\nu + \frac{1}{2}\right) \\
g[n] &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi\left(\nu+\frac{1}{2}\right)} H^*\left(\nu + \frac{1}{2}\right) e^{-i2\pi\nu n}\, d\nu \\
&= e^{i\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} H^*\left(\nu + \frac{1}{2}\right) e^{i2\pi\nu(1-n)}\, d\nu \;=\; -\int_0^1 H^*(\xi)\, e^{i2\pi\left(\xi-\frac{1}{2}\right)(1-n)}\, d\xi \\
&= -e^{-i\pi(1-n)} \int_0^1 H^*(\xi)\, e^{i2\pi\xi(1-n)}\, d\xi \;=\; (-1)^{2-n} \int_0^1 H^*(\xi)\, e^{i2\pi\xi(1-n)}\, d\xi \\
g[n] &= (-1)^n h^*[1-n] \qquad (13)
\end{aligned}$$

Under these conditions, let us finally demonstrate that the resulting system $\{\psi(t-n),\, n \in \mathbb{Z}\}$ is an orthogonal basis of W_0.
We start with :

$$\begin{aligned}
\sum_{n=-\infty}^{\infty} |\Psi(\nu+n)|^2 &= \sum_n \left|G\left(\frac{\nu+n}{2}\right)\right|^2 \left|\Phi\left(\frac{\nu+n}{2}\right)\right|^2 \quad \text{(see relation (9))} \\
&= \sum_n \left|H\left(\frac{\nu+n+1}{2}\right)\right|^2 \left|\Phi\left(\frac{\nu+n}{2}\right)\right|^2 \quad \text{(see 1st Eq. of relation (13))} \\
&= \sum_k \left|H\left(\frac{\nu}{2}+\frac{2k}{2}+\frac{1}{2}\right)\right|^2 \left|\Phi\left(\frac{\nu}{2}+\frac{2k}{2}\right)\right|^2 + \sum_k \left|H\left(\frac{\nu}{2}+\frac{2k+1}{2}+\frac{1}{2}\right)\right|^2 \left|\Phi\left(\frac{\nu}{2}+\frac{2k+1}{2}\right)\right|^2 \\
&= \left|H\left(\frac{\nu}{2}+\frac{1}{2}\right)\right|^2 \underbrace{\sum_k \left|\Phi\left(\frac{\nu}{2}+k\right)\right|^2}_{=1 \text{ (cf. (7))}} + \left|H\left(\frac{\nu}{2}\right)\right|^2 \underbrace{\sum_k \left|\Phi\left(\frac{\nu}{2}+k+\frac{1}{2}\right)\right|^2}_{=1 \text{ (cf. (7))}} \\
\sum_{n=-\infty}^{\infty} |\Psi(\nu+n)|^2 &= \left|H\left(\frac{\nu}{2}+\frac{1}{2}\right)\right|^2 + \left|H\left(\frac{\nu}{2}\right)\right|^2 \;=\; 1 \quad \text{(see Eq. (8))}
\end{aligned}$$

Therefore

$$\begin{aligned}
\int \psi(t)\psi(t-k)\, dt &= \int |\Psi(\xi)|^2\, e^{i2\pi\xi k}\, d\xi \\
&= \int_0^1 \underbrace{\left(\sum_{n=-\infty}^{\infty} |\Psi(\xi+n)|^2\right)}_{=1 \text{ (see equality above)}} e^{i2\pi\xi k}\, d\xi \\
\int \psi(t)\psi(t-k)\, dt &= \delta_{k0}.
\end{aligned}$$

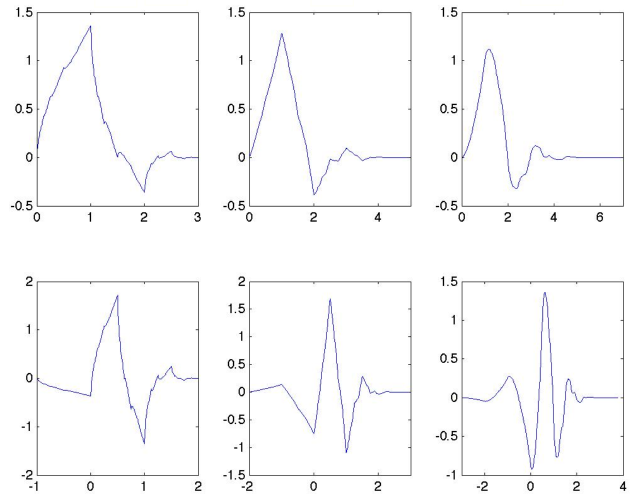


FIG. 4 – Plots of the scaling functions $\phi(t)$ (top row) and wavelets $\psi(t)$ (bottom row) for the compactly supported Daubechies wavelets. Each column corresponds to a specific number of vanishing moments : left $N_\psi = 2$, center $N_\psi = 3$ and right $N_\psi = 4$.

Figure 4 displays some examples of compactly supported scaling functions and wavelets with different numbers of vanishing moments (different regularities). Each system generates a multiresolution analysis scheme.

3.3.3 Discrete wavelets as a pyramidal filter bank

Analysis. The approximation coefficients at scale m read

$$\begin{aligned}
a_x[n,m] &= \int x(t)\,\phi_{nm}(t)\,dt \;=\; \int x(t)\,2^{m/2}\phi(2^m t - n)\,dt \\
&= \int x(t) 2^{m/2} \underbrace{\left[\sqrt{2}\sum_{k=-\infty}^{\infty} h[k]\,\phi(2(2^m - t) - k)\right]}_{\text{decomp. of } \phi_{nm} \text{ onto } V_{m+1}} dt \\
&= \sum_{k=-\infty}^{\infty} h[k] \int x(t) 2^{(m+1)/2}\phi(2^{m+1}t - (k+2n))\,dt \\
&= \sum_{k=-\infty}^{\infty} h[k] a_x[k+2n, m+1] \;=\; \sum_{k=-\infty}^{\infty} a_x[k, m+1] h[k-2n]
\end{aligned}$$

or equivalently

$$a_x[n,m] = h[\cdot] \underset{2n}{*} a_x[\cdot, m+1]. \tag{14}$$

Equation (14) must be interpreted as the convolution of the approximation series at finer scale $m+1$ with the low-pass filter coefficients $h[n]$, followed by a decimation of a factor 2 (i.e. retain only one over two output samples).
Similarly, the detail coefficients at scale m are obtained by convolution of the approximation series at finer scale $m+1$ with the high-pass filter coefficients $g[n]$, followed by a decimation of a factor 2,

$$d_x[n,m] \quad = \quad g[\cdot] \underset{2n}{*} a_x[\cdot, m+1]. \tag{15}$$

In practice, assuming that we start with the signal projection onto V_0 (the finest resolution space corresponding to observational scale fixed by the sampling rate), approximations and details at coarser scales, $j < 0$, are obtained by iteratively applying the QMF pair (14) and (15) to the preceding approximation sequence. The resulting pyramidal algorithm is depicted in figure 5.
Remark. Whenever the continuous version of the analyzed signal $x(t)$ is available, the finest approximation $a_x[0,n]$ can be obtained by projecting this latter onto the space V_0. In most of the case though, only a sampled version $x[n] = x(nT)$ is observed, which can directly and reasonably be considered as the approximation $P_0 x$ itself[7].

Synthesis As we shall see, the pyramidal algorithm, with the (almost) same pair of quadrature filters h and g, is perfectly reversible, allowing for reconstructing the initial series $x[n]$ from the approximation $a_x[m,n]$ at any scale $m < 0$ and its associated detail sequences $\{d[k,n],\, 0 > k \geq m\}$. Let us then recall that

$$P_m x(t) = \sum_k a_x[m,k]\phi_{m,k}(t)$$

and

$$P_{m+1}x(t) = P_m x(t) + \sum_k d_x[m,k]\psi_{m,k}(t).$$

Then, the approximation coefficients in V_{m+1} read

$$a_x[m+1,n] = \langle P_{m+1}x, \phi_{(m+1),n}\rangle = \sum_k a_x[m,k]\langle \phi_{m,k}, \phi_{(m+1),n}\rangle + \sum_k d_x[m,k]\langle \psi_{m,k}, \phi_{(m+1),n}\rangle$$

and because

$$\begin{cases} \langle \phi_{m,k}, \phi_{m+1,n}\rangle = \int 2^{\frac{m}{2}}\phi(2^m t - k)2^{\frac{m+1}{2}}\phi(2^{m+1}t - n)\,dt = \sqrt{2}\int \phi(t)\phi(2t-(n-2k))\,dt = h[n-2k] \\ \langle \psi_{m,k}, \phi_{m+1,n}\rangle = \int 2^{\frac{m}{2}}\psi(2^m t - k)2^{\frac{m+1}{2}}\phi(2^{m+1}t - n)\,dt = \sqrt{2}\int \psi(t)\phi(2t-(n-2k))\,dt = g[n-2k] \end{cases}$$

then

$$\begin{aligned} a_x[m+1,n] &= \sum_k h[n-2k]a_x[m,k] + \sum_k g[n-2k]d_x[m,k] \\ &= \sum_l h[n-l]\, a_x\left[m, \frac{l}{2}\right] + \sum_l g[n-l]\, d_x\left[m, \frac{l}{2}\right] \end{aligned} \tag{16}$$

Eq. (16) must be interpreted as follows (see figure 5) :

1. insert a zero valued sample between each two consecutive samples of $a_x[m,n]$ (resp. of $d_x[m,n]$)
2. convolve the resulting zero-added sequence $a_x[m,l/2]$ (resp. $d_x[m,l/2]$) with the low-pass filter $\tilde{h}[n] = h[-n]$ (resp. with the high-pass filter $\tilde{g}[n] = g[-n]$)
3. sum the two filters' outputs to get $a_x[m+1,n]$.

[7] A rigorous procedure to compute $P_0 x$ from $x[n]$ exists (see e.g. [**?**] and references therein).

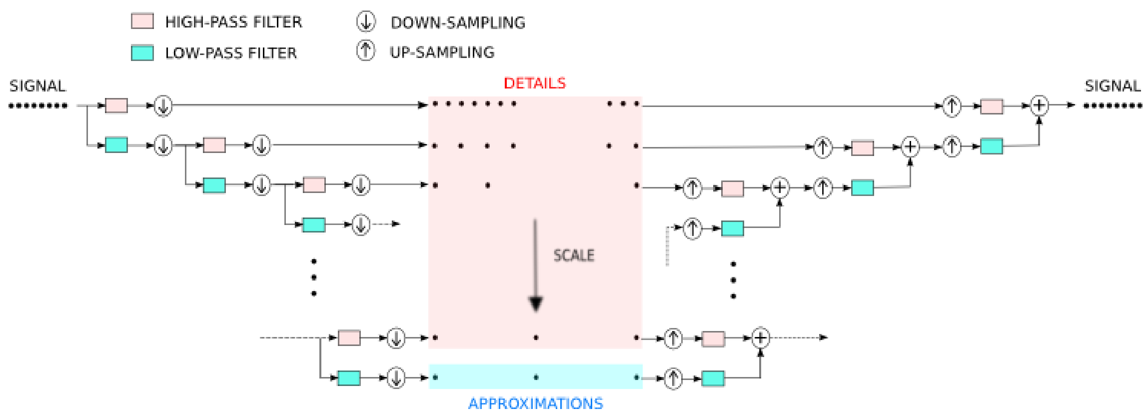


FIG. 5 – Pyramidal QMF filter bank corresponding to a wavelet multiresolution analysis.

[5, 4, 7, 1, 2, 6, 3]