Elements of Probability and Random Processes

Arnaldo Spalvieri

Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano, ITALY

Probability

The well-known frequentistic notion of probability can be mapped onto a mathematical theory, the probability theory. Given the set S of all the possible outcomes of an experiment, called the *sample space*, the axioms of probability theory are

- $0 \leq P(A) \leq 1$, for any $A \in S$,
- $P(S) = 1$,
- If A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.

From the axioms it is possible to show that

- $P(\overline{A}) = 1 - P(A)$, where $\overline{A}$ is the *complement* of A: $\overline{A} \cap A$ is void, and $\overline{A} \cup A = S$,
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, where $P(A \cap B)$ is the *joint probability* of A and B, that is the probability that both A and B are outcomes of the same experiment,
- if $A_1, A_2, \cdots, A_n$ are mutually exclusive and their union is the sample space, then $\sum_{i=1}^{n} P(A_i) = 1$.

Conditional Probability

The conditional probability is the probability of an event given that another event has occurred. For instance, the probability that the outcome of the roll of the dice is odd given that the result is lower than four is 2/3, while the unconditional probability that the outcome of the roll of the dice is odd is $1/2$. The conditional probability of event B given event A in probability theory is defined as

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

In the above example $P(A \cap B)$ is the probability of an odd number lower than four, that is $1/3$, while $P(A)$ is the probability of a number lower than four, that is $1/2$, hence $P(B|A) = 2/3$. Note that, since $A \cap B \equiv B \cap A$,

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B),$$

leading to Bayes' rule

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

Independency

When

$$P(B|A) = P(B)$$

knowing that A occurred does not modify the probability of event B, hence we can say that A does not condition B, or, in other words, that A and B are *independent* events.

Mathematically speaking, independency is written

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

Since $A \cap B \equiv B \cap A$ one also has

$$P(A|B) = P(A).$$

Random Variables

For mathematical convenience, it is worth assigning a number to the outcome of a random experiment. The number is called *random variable*. Let X be a real random variable, let x be a pint on the real axis, and consider the probability of the event $X \leq x$. This probability depends on the random experiment and is a function of x. It is denoted

$$F_X(x) = P(X \leq x)$$

and it is called *cumulative distribution function* (cdf) of X. The properties of the cumulative distribution function are

- $0 \leq F_X(x) \leq 1$,
- $F_X(x_1) \leq F_X(x_2)$ if $x_1 \leq x_2$.

Probability Density Function

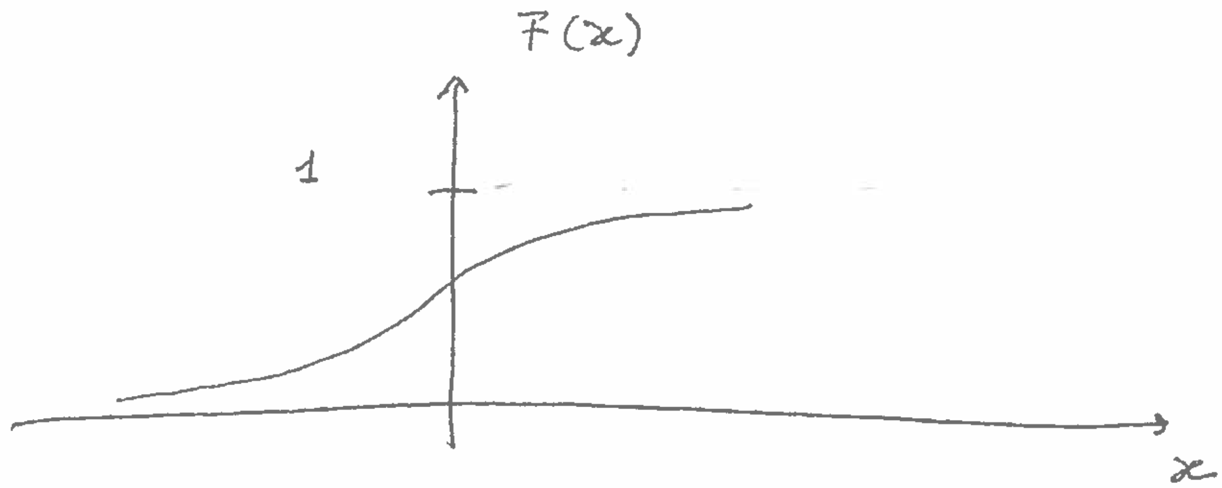
It is often useful to regard the cumulative distribution function as the following integral

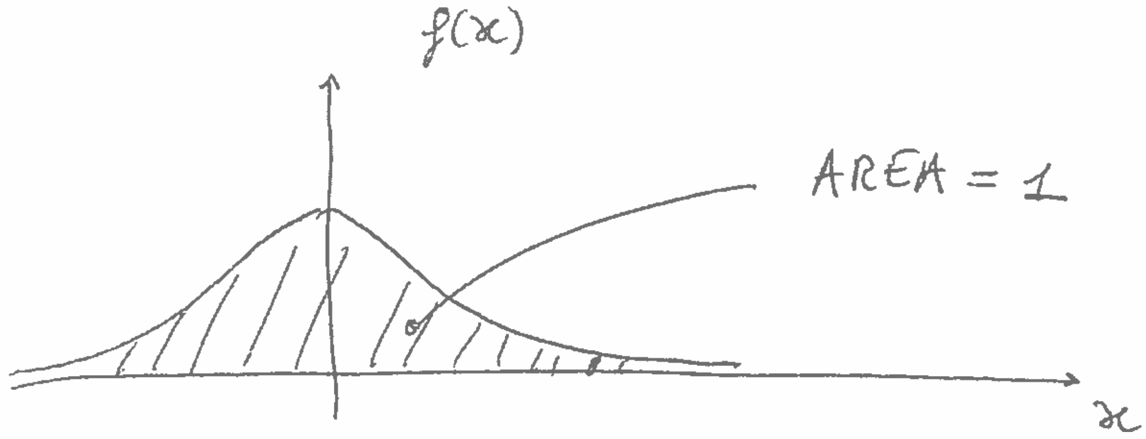
$$F_X(x) = \int_{-\infty}^{x} f_X(\xi) d\xi. \tag{1}$$

The function inside the integral takes the name of *probability density function* (pdf). From the properties of $F_X(x)$ one promptly realizes that

- $\int_{-\infty}^{\infty} f_X(\xi) d\xi = 1,$
- $f_X(x) \geq 0.$

A continuous random variable has smooth pdf, while the pdf of a discrete random variable is the weighted sum of delta functions, the sum of the weights being equal to one.


F(x)
1
x


f(x)
AREA = 1
x

Two Random Variables

Let us consider a random experiment that produces two results, for instance take randomly a man in a population and measure his weight and his height. Map the two results onto the two random variables X and Y. Define the joint cdf as

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

and the joint pdf as the function inside the following integral

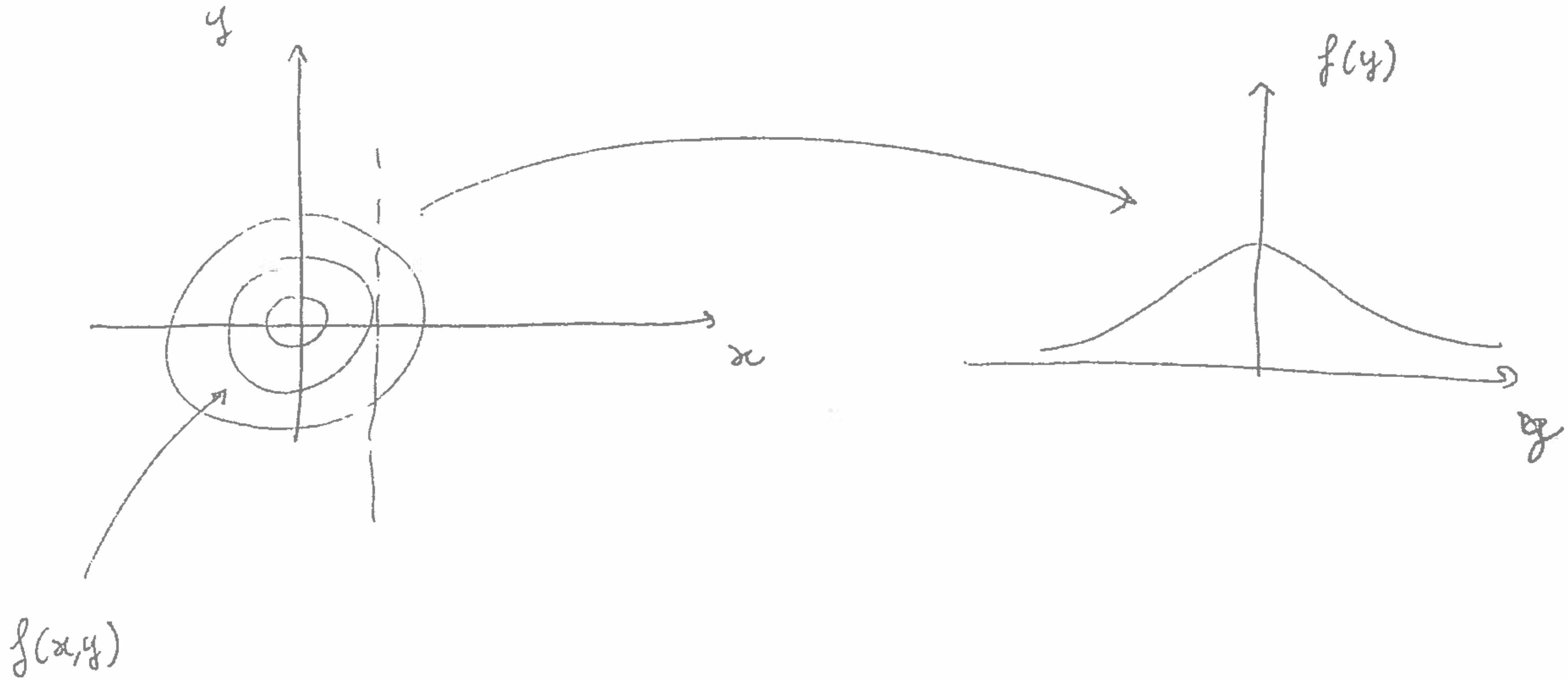
$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(\xi,\eta) d\xi d\eta. \tag{2}$$

The cdf of one of the two random variables is obtained by *saturation*, e.g. for X

$$F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(\xi,\eta) d\xi d\eta, \tag{3}$$

hence the *marginal* pdf of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,\eta) d\eta. \tag{4}$$



y
x
f(x,y)
f(y)
y

Conditional PDF and Independency

The conditional pdf of Y given X is defined as

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

If X and Y are statistically independent, then

$$f_{Y|X}(y|x) = f_Y(y), \tag{5}$$

that is

$$f_{X,Y}(x,y) = f_X(x) f_Y(y). \tag{6}$$

Expectation

Let $g(X)$ a real-valued transformation of random variable X. In the axiomatic probability theory, the expected value of $g(X)$ is defined as

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

The frequetistic interpretation of the expectation is meaningful:

$$E\{g(X)\} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} g(x_i),$$

where the random experiment is repeated N times and x_i is the i-th outcome of the random experiment. The term inside the limit in the right side of the above equation is called *sample average*.

Moments

Special cases of expectation are the *moments*, where $g(X) = X^n$, the most common one being the mean of X, which is obtained with $n = 1$ and is often denoted as μ_X:

$$E\{X\} = \mu_X = \int_{-\infty}^{\infty} x f_X(x)dx = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i.$$

The second-order moment is

$$E\{X^2\} = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i^2.$$

When the sequence of the N results of the random experiment are the samples of a discrete-time random signal, the mean is the value of the DC, while the second-order moment is the *power* of the signal.

Central Moments

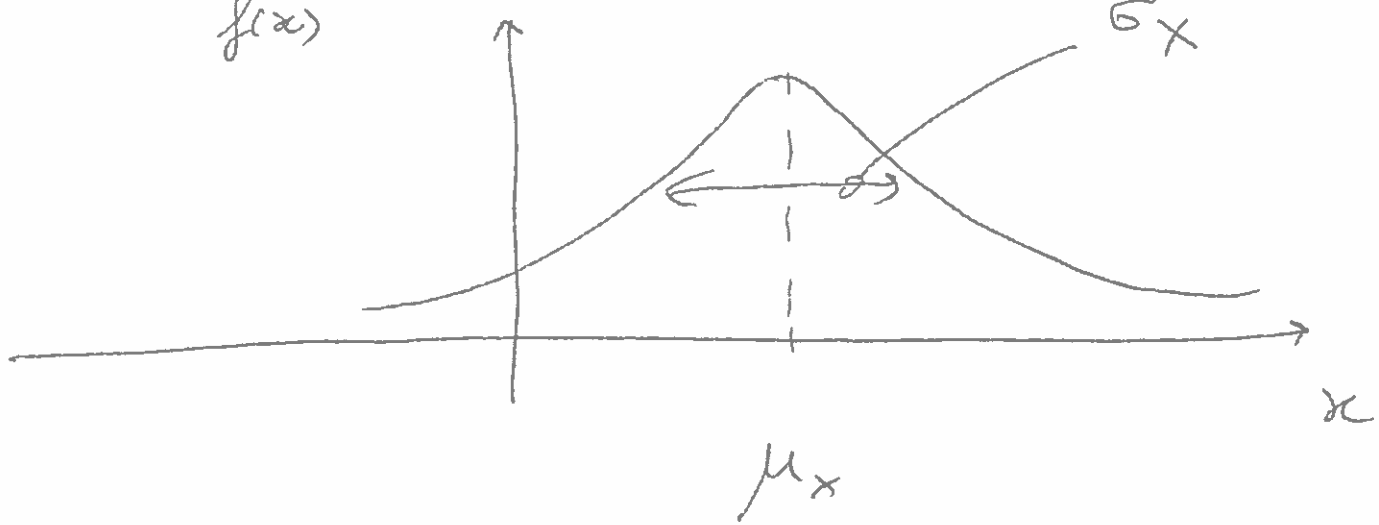
Central moments are defined as

$$E\{(X-\mu_X)^n\}.$$

The second-order central moment is called *variance*, and is denoted

$$\sigma_X^2 = E\{(X-\mu_X)^2\}.$$

The square root of the variance, that is σ_X, is called *standard deviation*, or, in the language of signals, RMS (root mean square) value. The variance measures how much the results of the random experiment are spread around the mean. When the random variable has zero mean the variance is equal to the second-order moment.


f(x)
σx
x
μx

Joint Moments

The concept of moments can be extended in a straightforward manner to joint random variables. For instance, the *correlation* between X and Y is the following joint moment

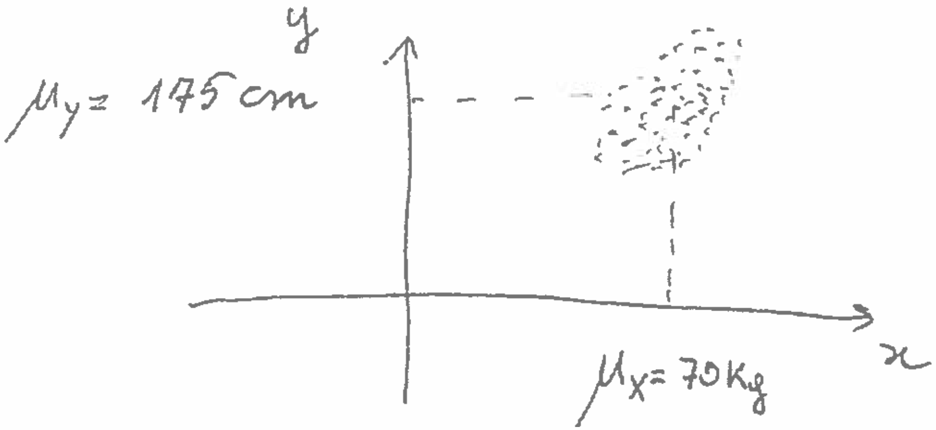
$$corr_{XY} = E\{XY\}.$$

The most important among central joint moments is the *covariance*,

$$cov_{XY} = E\{(X - \mu_X)(Y - \mu_Y)\} = E\{XY\} - \mu_X \mu_Y.$$

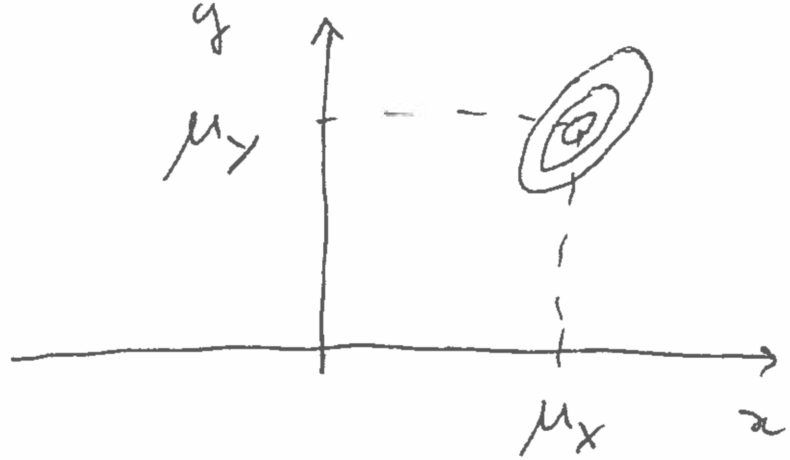
When $cov_{XY} = 0$, X and Y are said to be *uncorrelated*. Independent random variables are always uncorrelated, but the converse is not always true. The covariance can be normalized by dividing it by the product $\sigma_X \sigma_Y$, getting the so-called *correlation coefficient*

$$\rho_{XY} = \frac{cov_{XY}}{\sigma_X \sigma_Y}.$$

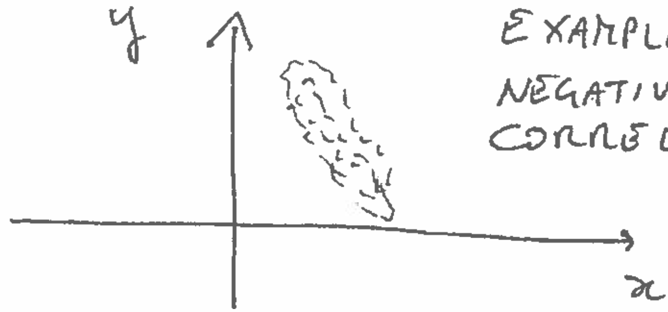


SAMPLE ; WEIGHT $= x$
HEIGHT $= y$

POSITIVE CORRELATION



REPRESENTATION OF $f(x,y)$ BY EQUAL-LEVEL LINES



EXAMPLE OF NEGATIVE CORRELATION

Sum of Two Random Variables

Let X and Y be two random variables and let the mean values, variances, and correlation coefficient between them be known. Also, let

$$Z = X + Y.$$

One has

$$\begin{aligned}
E\{Z\} &= E\{X+Y\} = E\{X\} + E\{Y\}, \\
E\{Z^2\} &= E\{(X+Y)^2\} = E\{X^2\} + E\{Y^2\} + 2E\{XY\} \\
&= \sigma_X^2 + \mu_X^2 + \sigma_Y^2 + \mu_Y^2 + 2(\rho_{XY}\sigma_X\sigma_Y + \mu_X\mu_Y), \\
E\{(Z-\mu_Z)^2\} &= E\{Z^2\} - \mu_Z^2 \\
&= \sigma_X^2 + \mu_X^2 + \sigma_Y^2 + \mu_Y^2 + 2(\rho_{XY}\sigma_X\sigma_Y + \mu_X\mu_Y) - (\mu_X + \mu_Y)^2 \\
&= \sigma_X^2 + \sigma_Y^2 + 2\rho_{XY}\sigma_X\sigma_Y,
\end{aligned}$$

hence with $\rho_{XY} = 0$ the variance of the sum is the sum of the variances, while with $\rho_{XY} = 1$ the variance of the sum is the square of the sum of the standard deviations.

Gaussian Random Variable

Let

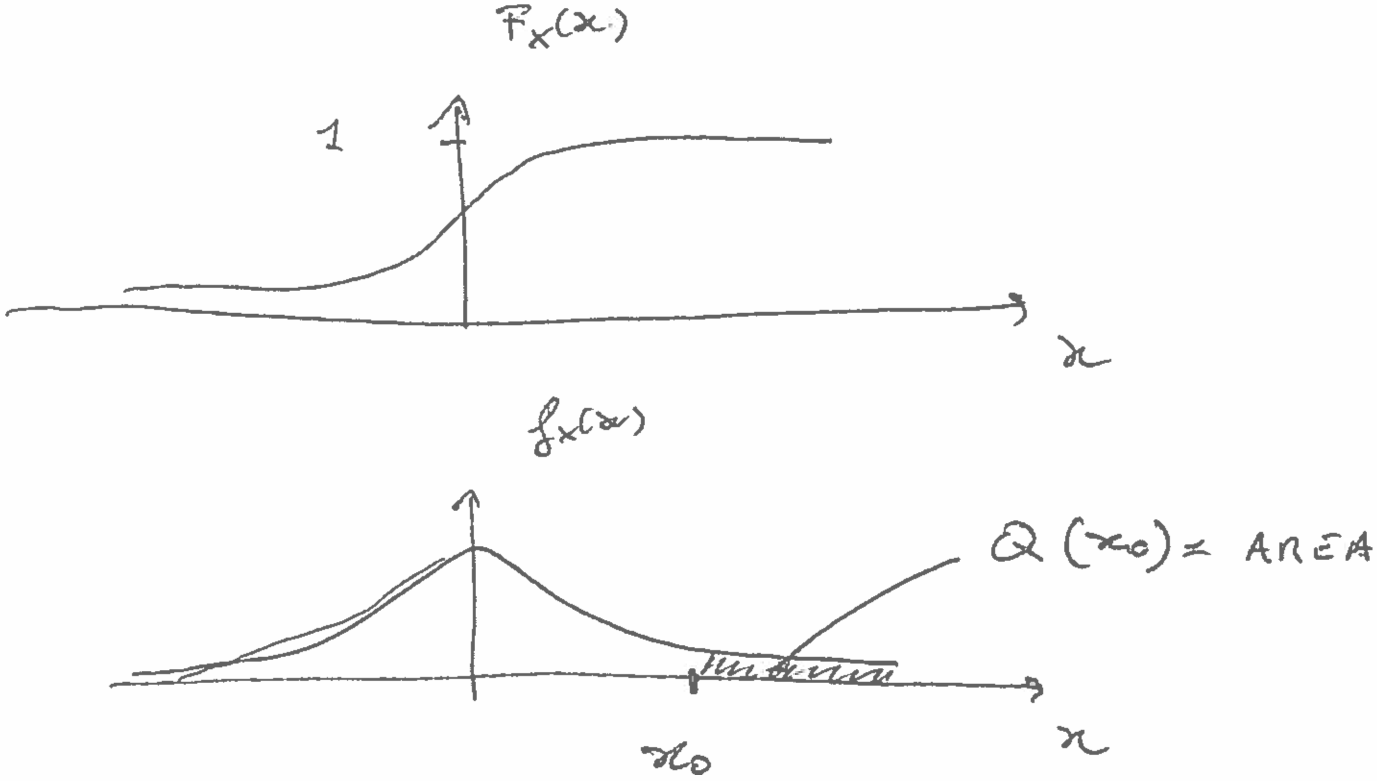
$$X = \lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Y_i,$$

where Y_i are i.i.d. random variables with zero mean and unit variance. The central limit theorem says that

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

that is a Gaussian distribution with zero mean and unit variance. The cdf cannot be computed in closed form. It can be tabulated and it is expressed as

$$F_X(x) = 1 - Q(x) = 1 - \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi.$$



$F_X(x)$
1
x
$f_X(x)$
x_0
x
$Q(x_0) = $ AREA

Gaussian Random Variable

For $x > 3$, the following approximation is tight:

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi \approx \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

More generally, the pdf and the cdf of a Gaussian random variable X with mean μ_X and variance σ_X^2 are

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}, \quad F_X(x) = 1 - Q\left(\frac{x-\mu_X}{\sigma_X}\right).$$

Jointly Gaussian Random Variables

The multivariate pdf of jointly Gaussian random variables, or Gaussian vector, has the form

$$f_{X_1,X_2,\cdots,X_n}(x_1,x_2,\cdots,x_n) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

where $\mathbf{x}$ is the column vector $(x_1, x_2, \cdots, x_n)^T$, $\boldsymbol{\mu}$ is the column vector of the mean values, Σ is the covariance matrix, $|\Sigma|$ is the determinant of Σ. When the Gaussian variables are uncorrelated, Σ is diagonal, therefore its inverse is diagonal with entries that are the inverse of the entries of Σ. In this case, the joint pdf can be factored:

$$f_{X_1,X_2,\cdots,X_n}(x_1,x_2,\cdots,x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right),$$

hence uncorrelated Gaussian random variables are also independent.

Random Process

A random process is a random experiment where the outcome of the experiment, called *realization*, is a waveform in time domain. The random process is denoted $X(t)$, while its generic realization is denoted $x(t)$. The mean of $X(t)$ is

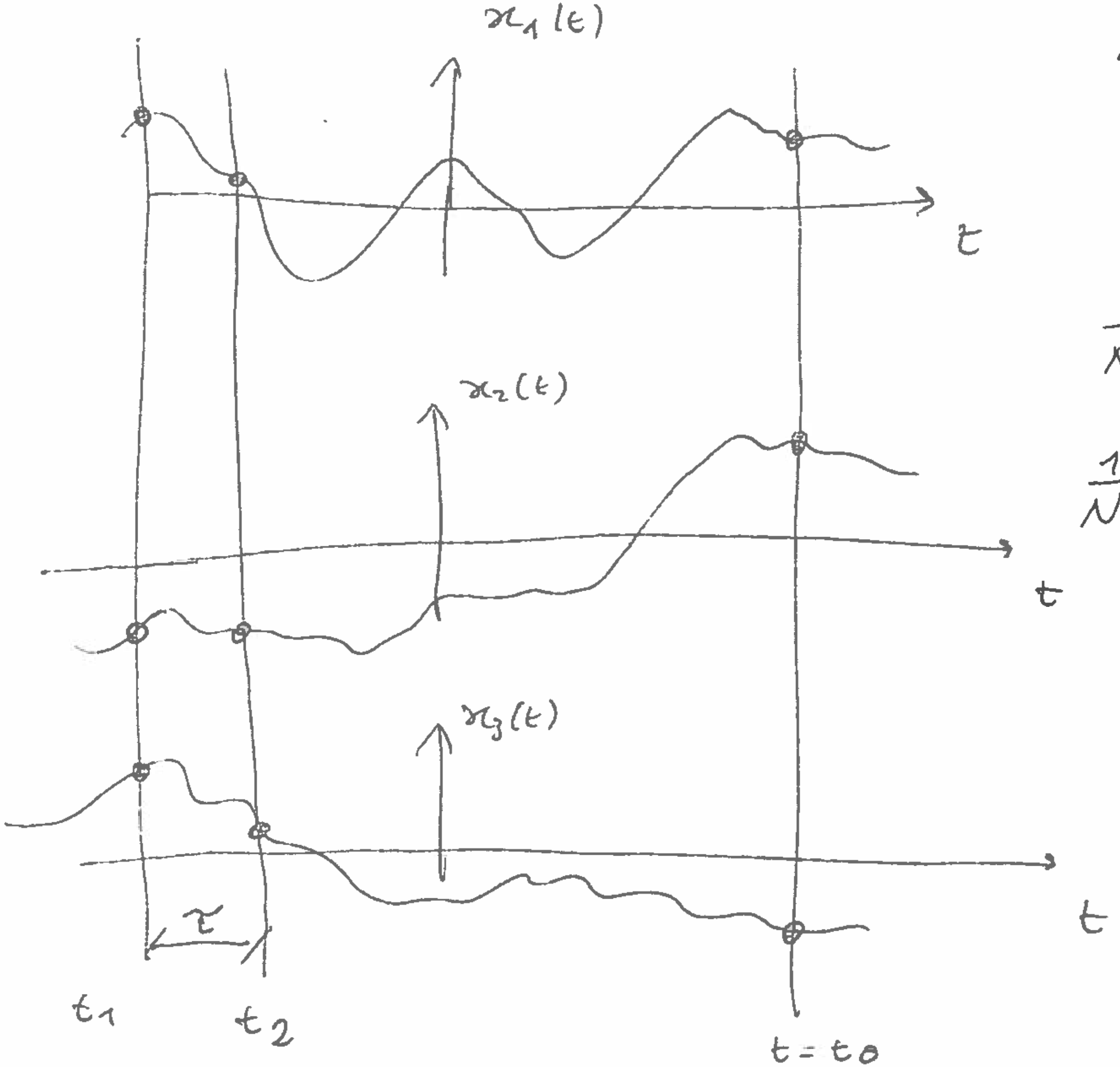
$$\mu_X(t) = E\{X(t)\} = \int_{-\infty}^{\infty} x f_{X(t)}(x) dx = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i(t).$$

The *autocorrelation* of $X(t)$ is

$$\begin{aligned} R_X(t_1, t_2) = E\{X^*(t_1)X(t_2)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^* x_2 f_{X(t_1),X(t_2)}(x_1 x_2) dx_1 dx_2 \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i^*(t_1) x_i(t_2). \end{aligned} \tag{7}$$

A process is stationary to mean and autocorrelation when

$$\mu_X(t) = \mu_X, \;\; R_X(t_1, t_2) = R_X(t_2 - t_1).$$



~~ENSEMBLE AVERAGE~~

SAMPLE AVERAGE

$$\frac{1}{N}\sum_{i=1}^{N} x_i(t_1)\, x_i(t_2)$$

$$\frac{1}{N}\sum_{i=1}^{N} x_i(t_0)$$

The Concept of Autocorrelation

Consider a stationary process and, for convenience, let $t = t_1$, $\tau = t_2 - t_1$, and denote the autocorrelation as

$$R_X(\tau) = E\{X^*(t)X(t+\tau)\}, \quad \forall t.$$

The autocorrelation says us how much the random signal is similar to itself in the average after τ seconds. In other words, the autocorrelation brings us information about the memory of the signal. Signals with long memory have autocorrelation with long duration, while signals with short memory have an autocorrelation with short duration.

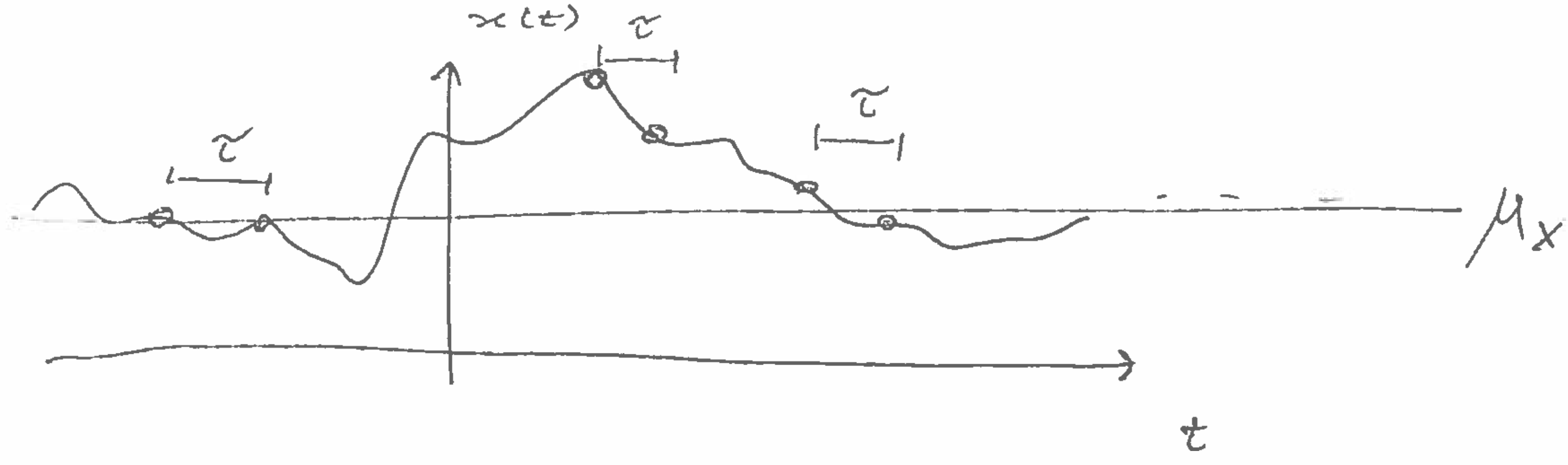
Ergodic Process

A process is said to be ergodic when anyone of the realizations of the process can represent the entire process. In this case the expectation is, in the limit, equal to the *time average*. For a stationary and ergodic signal, given any realization $x(t)$ and any time instant t one has

$$\mu_X = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i(t) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)dt,$$

$$R_X(\tau) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i^*(t)x_i(t+\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x^*(t)x(t+\tau)dt. \tag{8}$$

With ergodic signals, the concept of mean and autocorrelation is therefore exactly the one that we found with non-random signals.



TIME AVERAGE

Autocorrelation and Power Spectral Density

The *power spectral density* (psd) of the random process is the Fourier transform of the autocorrelation

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j2\pi f \tau} d\tau.$$

Given an input process with psd $S_X(f)$ to a filter with frequency response $H(f)$, the psd of the output process is

$$S_Y(f) = S_X(f)|H(f)|^2.$$

Given two random processes such that $E\{X^*(t)Y(t+\tau)\} = 0$, then for $Z(t) = X(t) + Y(t)$, from the definition of autocorrelation one has

$$R_Z(\tau) = R_X(\tau) + R_Y(\tau),$$

hence

$$S_Z(f) = S_X(f) + S_Y(f).$$

Gaussian Process

A random process is said to be a Gaussian process if the distribution of any set of samples $X(t_1), X(t_2), \cdots, X(t_n)$ is a multivariate Gaussian distribution. An important property of a Gaussian process is that when it is filtered the result is again a Gaussian process. Another important property is that, if the samples of the Gaussian process are uncorrelated, then they are also independent.

Thermal Noise

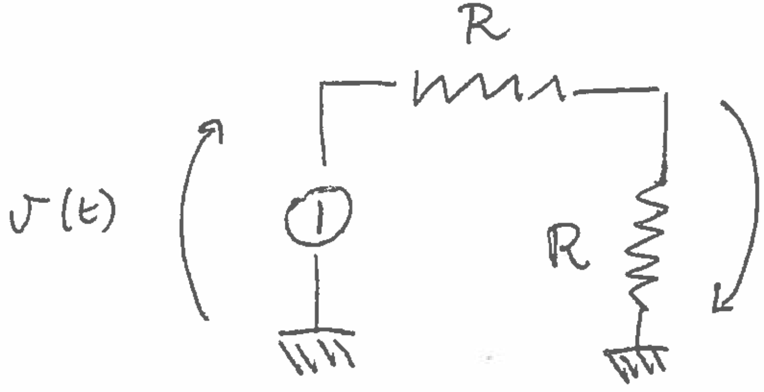
The random motion of electrons inside a resistor caused by temperature produces a random flow of electrons from the resistor to the load. This current is called *thermal noise.* The model is that of a voltage generator $v(t)$ connected in series with a resistance R. The voltage is a stationary and ergodic Gaussian process with zero mean and white psd up to the frequency of KT/h Hz, where $K = 1.38 \cdot 10^{-23}$ is Boltzmann's constant, $h = 6.63 \cdot 10^{-34}$ is Plank's constant, and T is the temperature in Kelvin degrees. Measuring the variance of the voltage in a bandwidth of B Hz one finds

$$E\{v^2(t)\} = 4KTBR.$$

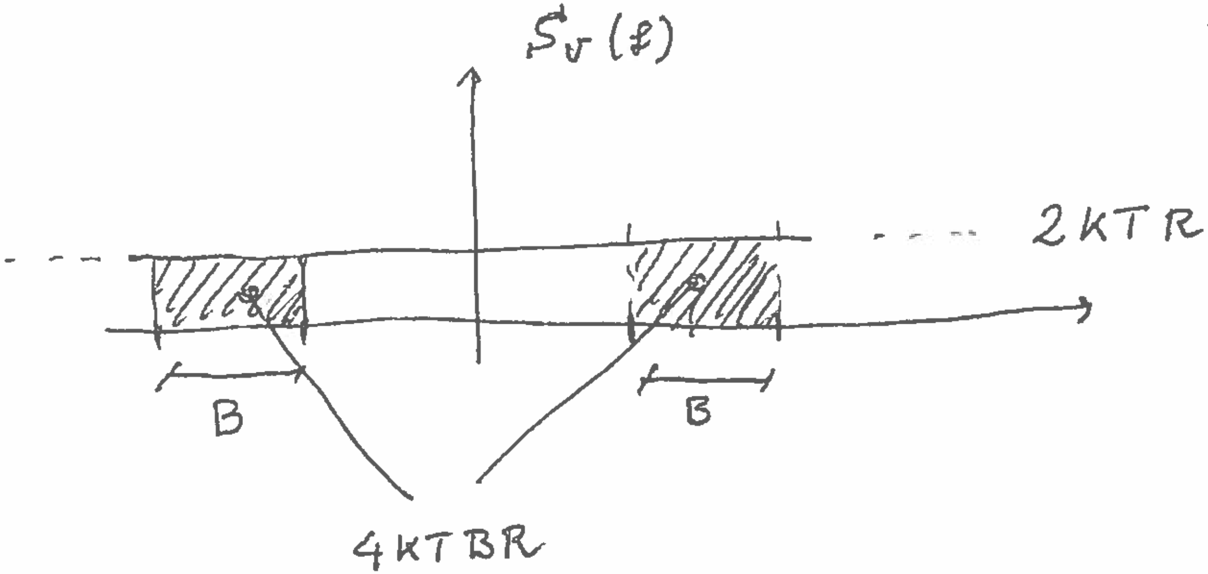
Assume that the voltage generator and the resistor are connected to a matched load, that is a resistor of resistance R, leading to maximum power transfer. The voltage on the load is $v(t)/2$, the current on the load is $i(t) = v(t)/(2R)$, and the power transferred to the load is

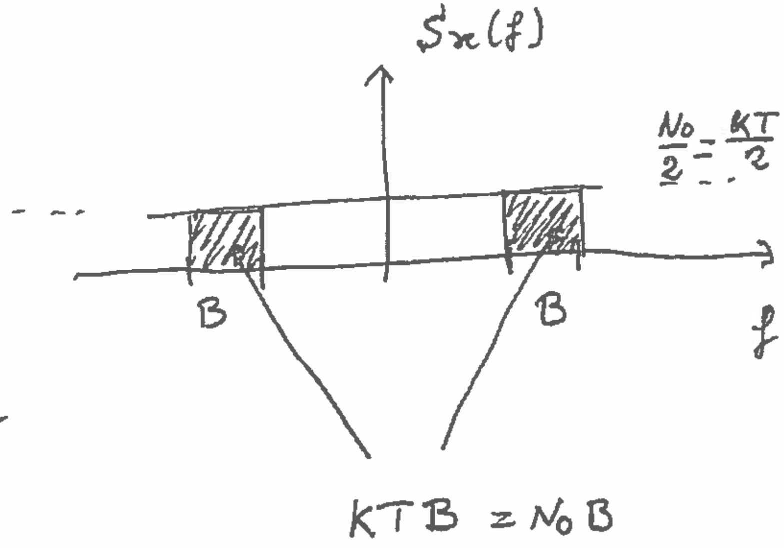
$$P = E\{i(t)v(t)/2\} = R^{-1}E\{v^2(t)/4\} = KTB = N_0B.$$

The two-sided power spectral density of thermal noise is therefore $N_0/2$.



$$\frac{v(t)}{2}\ , \quad i(t) = \frac{v(t)}{2R}\ , \quad x^2(t) = \frac{v(t)}{2}\, i(t) = \frac{v^2(t)}{4R}$$





Baseband Equivalent of Passband Noise

Suppose that thermal noise is filtered by a passband filter with frequency response of unit amplitude and bandwidth B around f_0:

$$n_{pb}(t) = \sqrt{2}n_c(t)cos(2\pi f_0 t) - \sqrt{2}n_s(t)sin(2\pi f_0 t),$$

where $n_c(t)$ and $n_s(t)$ are i.i.d. baseband Gaussian processes of bandwidth $B/2$. Representing the passaband noise trough its complex envelope one writes

$$n(t) = n_c(t) + jn_s(t).$$

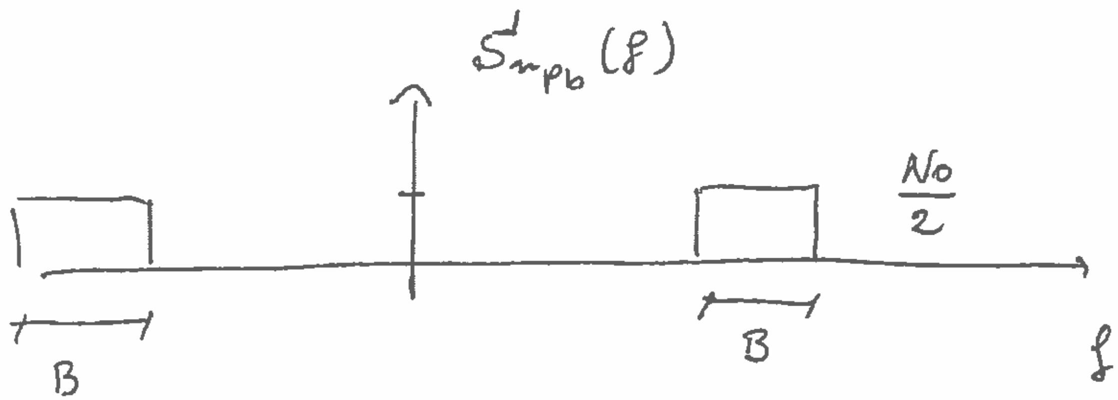
The psd of $n(t)$ is obtained by translating to the baseband the passband psd $N_0/2$ and multiplying it by two, hence

$$S_N(f) = N_0, \quad -B/2 < f < B/2.$$

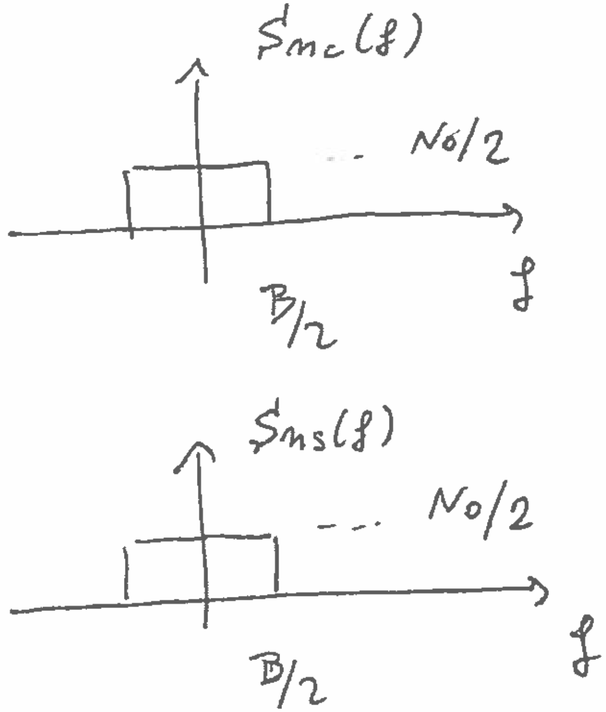
The power of the complex baseband noise process is

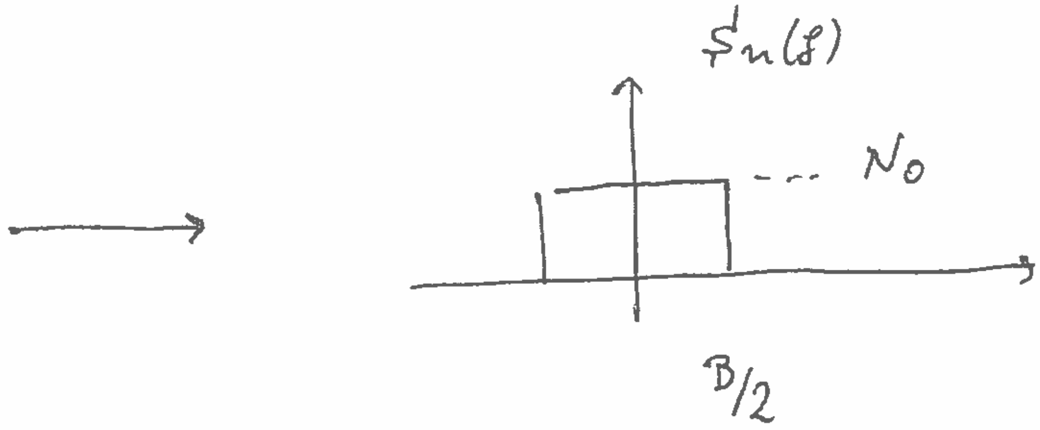
$$P_N = N_0 B,$$

that is equal to the power of the passband noise, as it should be.



$P = N_0 B$





$P = N_0 B$

Data Signal

Consider the data sequence

$$a(t) = \sum_{i=-\infty}^{\infty} a_i \delta(t - iT),$$

where a_i are complex i.i.d. random variables with zero mean and variance σ_a^2 and T is the symbol repetition interval. The data sequence is not stationary in strict sense, it is *cyclostationary* of period T. The expectation of a cyclostationary signal is obtained by integrating the time-varying expectation over the period and dividing by the period:

$$E\{g(X)\} = \frac{1}{T} \int_0^T E\{g(X(t))\} dt.$$

After straightforward calculations one gets

$$R_A(\tau) = \frac{\sigma_a^2}{T} \delta(\tau), \;\; S_A(f) = \frac{\sigma_a^2}{T}.$$

Filtered Data Signal

When the data sequence is filtered by a shaping filter with impulse response $h(t)$ one has

$$s(t) = \sum_k a_k h(t - kT).$$

The power spectral density and the power of $s(t)$ are

$$S_S(f) = S_A(f)|H(f)|^2 = \frac{\sigma_a^2}{T}|H(f)|^2,$$

$$P_S = \frac{\sigma_a^2}{T} E_h.$$