Decision and Estimation in Data Processing

Chapter I. Random Signals

I.1 Random variables

Random variables

- A **random variable** is a variable that holds a value produced by a (partially) random phenomenon
  - basically it is *a name* attached to an arbitrary value
  - short notation: r.v.
- Typically denoted as X, Y etc..
- Examples:
  - The value of a dice
  - The value of the voltage in a circuit
- The opposite = a **constant value**

Realizations

- **A realization** = a single outcome of the random experiment
- **Sample space** Ω = the set of all values that can be taken by a random variable X
  - i.e. the set of all possible realizations
- Example: rolling a dice
  - The r.v. is denoted as X
  - We might get a realization $X = 3$
  - But we could have got any value from the sample space

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Discrete and continuous random variables

- **Discrete** random variable: if Ω is a discrete set
  - Example: value of a dice
- **Continuous** random variable: if Ω is a continuous set
  - Example: a voltage value

Discrete random variables

- Consider a discrete r.v. X
- The **cumulative distribution function (CDF)** = the probability that the value of X is smaller or equal than the argument x

$$F_X(x) = P\left\{X \leq x\right\}$$

- In Romanian: *"funcție de repartitie"*
- Example: CDF for a dice
- For discrete r.v., the CDF is "stairwise"

Discrete random variables

- The **probability mass function (PMF)** = the probability that X has value x
$$w_X(x) = P\{X = x\}$$
- Example: what is the PMF of a dice?
- Relation to CDF:
$$F(x) = \sum_{all \quad t \leq x} w(t)$$

Continuous random variables

- Consider a continuous r.v. X
- The CDF of a continuous r.v. is defined identically:

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

- The derivative of the CDF is the **probability density function (PDF)**

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

$$F_X(x) = \int_{-\infty}^{x} w_X(t)dt$$

Continuous random variables

- The PDF gives the probability that the value of X is in a small vicinity *epsilon* around x, divided by *epsilon*

$$
\begin{aligned}
w_X(x) = \frac{dF_X(x)}{dx} &= \lim_{\epsilon \to 0} \frac{F_X(x+\epsilon) - F_X(x-\epsilon)}{2\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{P(X \in [x-\epsilon, x+\epsilon])}{2\epsilon}
\end{aligned}
$$

Probability of an exact value

- The probability that a continuous r.v. X is **exactly** equal to a value x is **zero**
  - because there are an infinity of possibilities (continuous)
  - That's why we can't define a probability mass function like for discrete
- The PDF gives the probability of being **in a small vicinity** around some value x

Probability and distribution

- Compute probability based on PDF (continuous r.v.):

$$P\{A \leq X \leq B\} = \int_A^B w_X(x)dx$$

- Compute probability based on PMF (discrete r.v.):

$$P\{A \leq X \leq B\} = \sum_{x=A}^{B} w_X(x)$$

Graphical interpretation

- Probability that a r.v. X is between A and B is **the area below the PDF**
  - i.e. the integral from A to B
- Probability that X is exactly equal to a certain value is zero
  - the area below a single point is zero

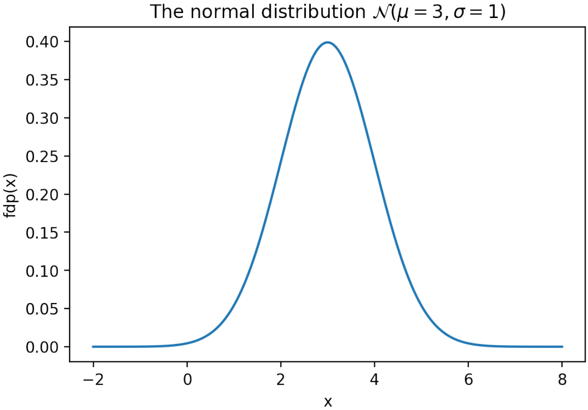
Properties of PDF/PMF/CDF

- The CDF is monotonously increasing (non-decreasing)
- The PDF/PMF are always ≥ 0
- The CDF starts from 0 and goes up to 1
- Integral/sum over all of the PDF/PMF $= 1$
- Some others, mention when needed

The normal distribution

- Probability density function

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



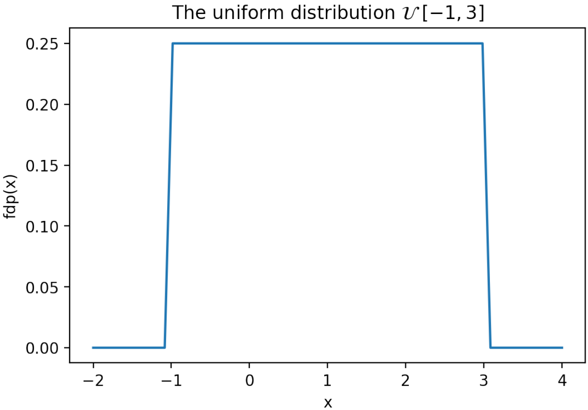
The normal distribution

- Has two parameters:
  - **Average value** μ = “center” of the function
  - **Standard deviation** σ = “width” is the function
- The front constant is just for normalization (ensures that integral $= 1$)
- Extremely often encountered in real life
- Any real value is possible $(w(x) > 0, \forall x \in \mathbb{R})$
- Usually denoted as $\mathcal{N}(\mu, \sigma)$

The uniform distribution

- The probability density function = constant, between two endpoints

$$w(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & elsewhere \end{cases}$$



The uniform distribution

- Has two parameters: the limits a and b of the interval
- The "height" of the function is $\frac{1}{b-a}$, for normalization
- Very simple
- Only values from the interval $[a, b]$ are possible
- Denoted as $\mathcal{U}\left[a, b\right]$

Other distributions

- Many other distributions exist, relevant for particular applications

R.v. as functions of other r.v.

- A function applied to a r.v. produces another r.v.
- Examples: if X is a r.v. with distribution $\mathcal{U}\,[0, 10]$, then
  - $Y = 5 + X$ is another r.v., with distribution $\mathcal{U}\,[5, 15]$
  - $Z = X^2$ is also another r.v.
  - $T = cos(X)$ is also another r.v.
- Reason: since X is random, the values Y, Z, T are also random
- X, Y, Z, T are not independent
  - A certain value of one of them automatically implies the value of the others

Exercise

Exercise:

- If X is a r.v. with distribution $\mathcal{U}\,[0, \pi]$, compute the probability density of a r.v. Y defined as

$$Y = cos(X)$$

Computing probabilities for the normal distribution

- How to compute $\int_a^b$ for a normal distribution?
  - Can't be done with algebraic formula, non-elementary function
- Use *the error function*:

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$$

- The CDF of a normal distribution $\mathcal{N}(\mu, \sigma^2)$

$$F(X) = \frac{1}{2}(1 + erf(\frac{x - \mu}{\sigma\sqrt{2}}))$$

- The values of *erf()* are available / are computed numerically
  - e.g. on GOogle, search for $erf(0.5)$
- Other useful values:
  - $erf(-\infty) = -1$
  - $erf(\infty) = 1$

Exercise

Exercise:

- Let X be a r.v. with distribution $\mathcal{N}(3, 2)$. Compute the probability that $X \in [2, 4]$

Multiple random variables

- Consider a system with two continuous r.v. X and Y
- Joint cumulative distribution function:

$$F_{XY}(x_i, y_j) = P\left\{X \leq x_i \cap Y \leq y_i\right\}$$

- Joint probability density function:

$$w_{XY}(x_i, y_j) = \frac{\partial^2 P_{XY}(x_i, y_j)}{\partial x \partial y}$$

- The joint PDF gives the probability that the values of the two r.v. X and Y are in a vicinity of x_i and y_i **simultaneously**
- Similar for discrete r.v.: the joint PMF

$$w_{XY}(x, y) = P\left\{X = x \cap Y = y\right\}$$

Independent random variables

- Two v.a. X and Y are **independent** if the value of one of them does not influence in any way the value of the other
- For independent r.v., the probability that $X = x$ and $Y = y$ is the product of the two probabilities
- Discrete r.v.:

$$w_{XY}(x, y) = w_X(x) \cdot w_Y(y)$$

$$P\{X = x \cap Y = y\} = P\{X = x\} \cdot P\{Y = y\}$$

- Relation holds for CDF / PDF / PMF, continuous or discrete r.v.
- Same for more than two r.v.

Independent random variables

Exercise:

- Compute the probability that three r.v. X, Y and Z i.i.d. $\mathcal{N}(-1, 1)$ are all positive simultaneously
  - ***i.i.d*** $=$ "independent and identically distributed"

Statistical averages

- R.v. are described by statistical averages ("*moments*")
- The average value (moment of order 1)
- Continuous r.v.:

$$\overline{X} = E\{X\} = \int_{-\infty}^{\infty} x \cdot w_X(x) dx$$

- Discrete r.v.:

$$\overline{X} = E\{X\} = \sum_{x=-\infty}^{\infty} x \cdot w_X(x) dx$$

- (Example: the entropy of H(X) = the average value of the information)
- Usual notation: μ

Properties of the average value

- Computing the average value is a **linear** operation
  - because the underlying integral / sum is a linear operation
- Linearity

$$E\{aX + bY\} = aE\{X\} + bE\{Y\}$$

- Or:

$$E\{aX\} = aE\{X\}, \forall a \in \mathbb{R}$$

$$E\{X + Y\} = E\{X\} + E\{Y\}$$

- No proof given here

Average squared value

- Average squared value = average value of the squared values
- Moment of order 2
- Continuous r.v.:

$$\overline{X^2} = E\{X^2\} = \int_{-\infty}^{\infty} x^2 \cdot w_X(x) dx$$

- Discrete r.v.:

$$\overline{X^2} = E\{X^2\} = \sum_{-\infty}^{\infty} x^2 \cdot w_X(x) dx$$

- Interpretation: average of squared values = average energy of a signal

Dispersion (variance)

- Dispersion (variance) = average squared value of the difference to the average value
- Continuous r.v.:

$$\sigma^2 = \overline{\{X-\mu\}^2} = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot w_X(x)dx$$

- Discrete r.v.:

$$\sigma^2 = \overline{\{X-\mu\}^2} = \sum_{-\infty}^{\infty} (x-\mu)^2 \cdot w_X(x)dx$$

- Interpretation: how much do the values vary around the average value
  - σ^2 = large: large spread around the average value
  - σ^2 = small: values are concentrated around the average value

Relation between the three values

- Relation between the average value, the average squared value, and the dispersion:

$$
\begin{aligned}
\sigma^2 &= \overline{\{X - \mu\}^2} \\
&= \overline{X^2 - 2 \cdot X \cdot \mu + \mu^2} \\
&= \overline{X^2} - 2\mu\overline{X} + \mu^2 \\
&= \overline{X^2} - \mu^2
\end{aligned}
$$

Sum of random variables

- Sum of two or more **independent** r.v. is also a r.v.
- Its distribution = the **convolution** of the distributions of the two r.v.
- If $Z = X + Y$
$$w(z) = w(x) \star w(y)$$
- Particular case: if X and Y are normal r.v., with $\mathcal{N}(\mu_X, \sigma_X^2)$ and $\mathcal{N}(\mu_Y, \sigma_Y^2)$, then:
  - Z is also a normal r.v., with $\mathcal{N}(\mu_Z, \sigma_Z^2)$, having:
  - average = sum of the two averages: $\mu_Z = \mu_X + \mu_Y$
  - dispersion = sum of the two dispersions: $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$

I.2 Random processes

Random process

- A **random process** = a sequence of random variables indexed in time
- **Discrete-time** random process $f[n]$ = a sequence of random variables at discrete moments of time
  - e.g.: a sequence 50 of throws of a dice, the daily price on the stock market
- **Continuous-time** random process $f(t)$ = a continuous sequence of random variables at every moment
  - e.g.: a noise voltage signal, a speech signal
- Every sample from a random process is a (different) random variable!
  - e.g. $f(t_0)$ = value at time t_0 is a r.v.

Realizations of random processes

- A **realization** of the random process = a particular sequence of realizations of the underlying r.v.
  - e.g. we see a given noise signal on the oscilloscope, but *we could have seen any other realization just as well*
- When we consider a random process = we consider the set of all possible realizations

Distributions of order 1 of random processes

- Every sample $f(t_1)$ from a random process is a random variable
  - with CDF $F_1(x; t_1)$
  - with PDF $w_1(x; t_1) = \frac{dF_1(x;t_1)}{dx}$
- The sample at time t_2 is a different random variable with **possibly different** functions
  - with CDF $F_1(x; t_2)$
  - with PDF $w_1(x; t_2) = \frac{dF_1(x;t_2)}{dx}$
- These functions specify how the value of one sample is distributed
- The index w_1 indicates we consider a single random variable from the process (distributions of order 1)
- Same for discrete-time random processes

Distributions of order 2

- A pair of random variables $f(t_1)$ and $f(t_2)$ sampled from the random process $f(t)$ have
  - joint CDF $F_2(x_i, x_j; t_1, t_2)$
  - joint PDF $w_2(x_i, x_j; t_1, t_2) = \frac{\partial^2 F_2(x_i, x_j; t_1, t_2)}{\partial x_i \partial x_j}$
- These functions specify how the pair of values is distributed (distributions of order 2)
- Same for discrete-time random processes

Distributions of order n

- Generalize to n samples of the random process
- A set of n random variables $f(t_1), ...f(t_n)$ sampled from the random process $f(t)$ have
  - joint CDF $F_n(x_1, ...x_n; t_1, ...t_n)$
  - joint PDF $w_n(x_1, ...x_n; t_1, ...t_n) = \frac{\partial^2 F_n(x_1, ...x_n; t_1, ...t_n)}{\partial x_1 ... \partial x_n}$
- These functions specify how the whole set of n values is distributed (distributions of order n)
- Same for discrete-time random processes

Statistical averages

Random processes are characterized using statistical / temporal averages (*moments*)

1. Average value

$$\overline{f(t_1)} = \mu(t_1) = \int_{-\infty}^{\infty} x \cdot w_1(x; t_1) dx$$

2. Average squared value (*valoarea patratica medie*)

$$\overline{f^2(t_1)} = \int_{-\infty}^{\infty} x^2 \cdot w_1(x; t_1) dx$$

Statistical averages - variance

3. Variance ($=$ *dispersia*)

$$\sigma^2(t_1) = \overline{\{f(t_1) - \mu(t_1)\}^2} = \int_{-\infty}^{\infty} (x - \mu(t_1)^2 \cdot w_1(x; t_1) dx$$

- The variance can be computed as:

$$\begin{aligned} \sigma^2(t_1) =& \overline{\{f(t_1) - \mu(t_1)\}^2} \\ =& \overline{f(t_1)^2 - 2f(t_1)\mu(t_1) + \mu(t_1)^2} \\ =& \overline{f^2(t_1)} - \mu(t_1)^2 \end{aligned}$$

- Note:
  - these three values are calculated across all realizations, at time t_1
  - they characterize only the sample at time t_1
  - at a different time t_2, the r.v. $f(t_2)$ is different so all average values might be different

Statistical averages - autocorrelation

4. The autocorrelation function

$$R_{ff}(t_1, t_2) = \overline{f(t_1)f(t_2)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 w_2(x_1, x_2; t_1, t_2) dx_1 dx_2$$

5. The correlation function (for different random processes $f(t)$ and $g(t)$)

$$R_{fg}(t_1, t_2) = \overline{f(t_1)g(t_2)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 y_2 w_2(x_1, y_2; t_1, t_2) dx_1 dy_2$$

- Note:
  - these functions may have different values for a different pair of values (t_1, t_2)

Temporal averages

- What to do when we only have access to a single realization $f^{(k)}(t)$?
- Compute values **for a single realization $f^{(k)}(t)$, across all time moments**

1. Temporal average value

$$\overline{f^{(k)}(t)} = \mu^{(k)} = \lim_{T\to\infty} \frac{1}{T} \int_{T/2}^{T/2} f^{(k)}(t) dt$$

2. Temporal average squared value

$$\overline{[f^{(k)}(t)]^2} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} [f^{(k)}(t)]^2 dt$$

Temporal variance

3. Temporal variance

$$\sigma^2 = \overline{\left\{f^{(k)}(t) - \mu^{(k)}\right\}^2} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} (f^{(k)}(t) - \mu^{(k)})^2 dt$$

- The variance can be computed as:

$$\sigma^2 = \overline{[f^{(k)}(t)]^2} - [\mu^{(k)}]^2$$

- Note:
  - these values do not depend anymore on time t (integrated)

Temporal autocorrelation

4. The temporal autocorrelation function

$$R_{ff}(t_1, t_2) = \overline{f^{(k)}(t_1 + t) f^{(k)}(t_2 + t)}$$
$$= \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} f^{(k)}(t_1 + t) f^{(k)}(t_2 + t) dt$$

5. The temporal correlation function (for different random processes $f(t)$ and $g(t)$)

$$R_{fg}(t_1, t_2) = \overline{f^{(k)}(t_1 + t) g^{(k)}(t_2 + t)}$$
$$= \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} f^{(k)}(t_1 + t) g^{(k)}(t_2 + t) dt$$

Statistical and temporal averages

- Statistical averages are usually the relevant values
- But in real life, we can only compute the temporal values
- Fortunately, in many cases they are the same (ergodicity, see later)

Stationary random processes

- All the statistical averages are dependent on the time
  - i.e. they might be different for a sample at t_2
- **Stationary** random process = when all statistical averages are **identical if we shift the time origin** (e.g. delay the signal)
- Equivalent definition: if all the PDF are identical when shifting the time origin

$$w_n(x_1, ...x_n; t_1, ...t_n) = w_n(x_1, ...x_n; t_1 + \tau, ...t_n + \tau)$$

- Basically, nothing should depend on the time t

Strict-sense and wide-sense stationary

- Strictly stationary / strongly stationary / strict-sense stationary:
  - relation holds for every n
- Weakly stationary / wide-sense stationary:
  - relation holds only for $n = 1$ and $n = 2$ (the most used)

Consequences of stationarity

- For $n = 1$:

$$w_1(x_i; t_1) = w_1(x_i; t_2) = w_1(x_i)$$

- The average value, average squared value, variance of a sample are all **identical** for any time t

$$\overline{f(t)} = constant, \forall t$$

$$\overline{f^2(t)} = constant, \forall t$$

$$\sigma^2(t) = constant, \forall t$$

Consequences of stationarity

- For $n = 2$:

$$w_2(x_i, x_j; t_1, t_2) = w_2(x_i, x_j; 0, t_2 - t_1) = w_2(x_i, x_j; t_2 - t_1)$$

- The autocorrelation function depends only on the **time difference** $\tau = t_2 - t_1$ between the samples

$$R_{ff}(t_1, t_2) = R_{ff}(0, t_2 - t_1) = R_{ff}(\tau) = \overline{f(t)f(t+\tau)}$$

- Is the average value of a product of two samples time τ apart
- Depends on a single value τ = time difference of the two samples

Consequences of stationarity

- Same for correlation function between two different r.p
- Depends only on the **time difference** $\tau = t_2 - t_1$ between the samples

$$R_{fg}(t_1, t_2) = R_{fg}(0, t_2 - t_1) = R_{fg}(\tau) = \overline{f(t)g(t+\tau)}$$

- Is the average value of a product of two samples time τ apart

Ergodic random processes

- In practice, we have access to a single realization
- **Ergodic** random process = the temporal averages on any realization are equal to the statistical averages
- We can compute / estimate all averages from a single realization (any)
  - the realization must be very long (length $\to \infty$) for precise results
- Realizations are all similar to the others, statistically
  - a single realization is characteristic of the whole process

Ergodic random processes

- Most random processes we care about are ergodic and stationary
  - e.g. noises
- Example of non-ergodic process:
  - throw a dice, then the next 50 values are identical to the first
  - a single realization is not characteristic

I.3 More on autocorrelation

The Power Spectral Density of a random process

- The Power Spectral Density (PSD) $S_{ff}(\omega)$ is the power of the random process at every frequency f $(\omega = 2\pi f)$
- The PSD describes how the power of a signal is distributed in frequency
  - e.g. some random processes have more power at low frequency, others at high frequency etc.
- The power in the frequency band $[f_1, f_2]$ is equal to $\int_{f_1}^{f_2} S_{ff}(\omega) d\omega$
- The whole power of the signal is $\int_{-\infty}^{\infty} S_{ff}(\omega) d\omega$
- The PSD is basically a measurable quantity
  - it can be determined experimentally
  - it is important in practical (engineering) applications

The Wiener-Khinchin theorem

- *Rom: teorema Wiener-Hincin*

Theorem:

- **The Power Spectral Density = the Fourier transform of the autocorrelation function**

$$S_{ff}(\omega) = \int_{-\infty}^{\infty} R_{ff}(\tau) e^{-j\omega\tau} d\tau$$

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

- No proof
- Makes a relation between two rather different domains
  - autocorrelation function: a *statistical* property
  - PSD function: a *physical* property (relevant for engineering purposes)

White noise

- White noise = a random process with autocorrelation function equal to a Dirac function
$$R_{ff}(\tau) = \delta(\tau)$$
- Any two different samples ($\tau \neq 0$) have zero correlation (are uncorrelated)
  - they do not vary similarly
- Power spectral density = Fourier transform of a Dirac = a constant
  - has equal power at all frequencies up to ∞
- In real life, power goes to 0 for very high frequencies
  - "*band-limited white noise*"
  - Samples which are very close are necessarily somewhat correlated
- White noise can have almost any distribution
  - normal, uniform etc.

Properties of the autocorrelation function

1. Is even

$$R_{ff}(\tau) = R_{ff}(-\tau)$$

- Proof: change variable in definition

2. At infinite it goes to a constant

$$R_{ff}(\infty) = \overline{f(t)}^2 = const$$

- Proof: two samples separated by ∞ are independent

3. Is maximum in 0

$$R_{ff}(0) \geq R_{ff}(\tau)$$

- Proof: start from $\overline{(f(t) - f(t+\tau))^2} \geq 0$
- Interpretation: different samples might vary differently, but a sample always varies identically with itself

Properties of the autocorrelation function

4. Value in 0 = the power of the random process

$$R_{ff}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{ff}(\omega) d\omega$$

- Proof: Put $\tau = 0$ in inverse Fourier transform of Wiener-Khinchin theorem

5. Variance = difference between values at 0 and ∞

$$\sigma^2 = R_{ff}(0) - R_{ff}(\infty)$$

- Proof: $R_{ff}(0) = \overline{f(t)^2}$, $R_{ff}(\infty) = \overline{f(t)}^2$

Autocorrelation of filtered random processes

- Consider a random process applied as input to a system
  - either continuous-time: input $x(t)$, system $H(s)$, output $y(t)$
  - or discrete-time: input $x[n]$, system $H(z)$, output $y[n]$
- How does the autocorrelation of y depend on that of the input x?