Decision and Estimation in Data Processing

Chapter I. Random Signals

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.):

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

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

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

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

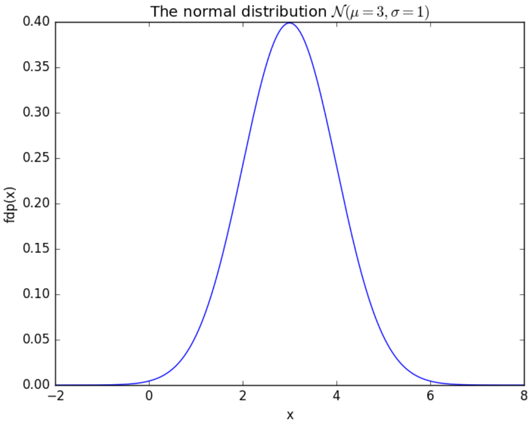
Properties of PDF/PMF/CDF

- 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

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



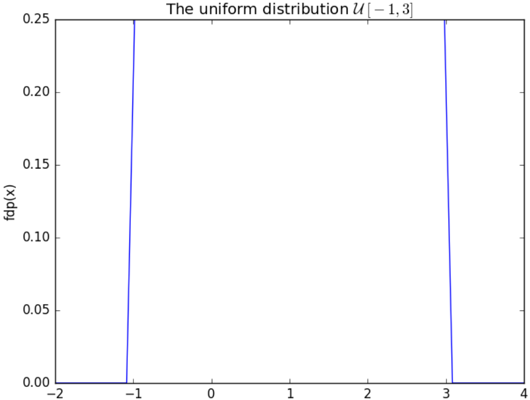
The normal distribution

- Has two parameters:
  - **Average value** $\mu =$ "center" of the function
  - **Standard deviation** $\sigma =$ "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 $\mathcal{U}\left[a, b\right]$

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

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.

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
  - $\sigma^2 =$ large: large spread around the average value
  - $\sigma^2 =$ 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: $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$

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 p.a.

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 p.a.

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 p.a.

Hic sunt leones

Statistical averages

We characterize random processes 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*)

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