Probability

- Probability theory: branch of mathematics for description and modelling of random events
- Modern probability theory - the axiomatic definition of probability - introduced by Kolmogorov in [Kolmogorov, 1933]
Definition of Probability

- Experiment with an uncertain outcome: *random experiment*
- Union of all possible *outcomes* $\zeta$ of the random experiment: *certain event* or *sample space* of the random experiment - $\mathcal{O}$
- *Event*: subset $\mathcal{A} \subseteq \mathcal{O}$
- *Probability*: measure $P(\mathcal{A})$ assigned to $\mathcal{A}$ satisfying the following three axioms
  - Probabilities are non-negative real numbers: $P(\mathcal{A}) \geq 0$, $\forall \mathcal{A} \subseteq \mathcal{O}$
  - Probability of the certain event: $P(\mathcal{O}) = 1$
  - If $\{\mathcal{A}_i : i = 0, 1, \cdots \}$ is a countable set of events such that $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$ for $i \neq j$, then
    \[
    P\left(\bigcup_i \mathcal{A}_i\right) = \sum_i P(\mathcal{A}_i) \tag{1}
    \]
Independence and Conditional Probability

- Two events $A_i$ and $A_j$ are **independent** if
  \[ P(A_i \cap A_j) = P(A_i)P(A_j) \]  
  (2)

- The **conditional probability** of an event $A_i$ given another event $A_j$, with $P(A_j) > 0$ is
  \[ P(A_i | A_j) = \frac{P(A_i \cap A_j)}{P(A_j)} \]  
  (3)

- With direct consequence: **Bayes’ theorem**
  \[ P(A_i | A_j) = P(A_j | A_i) \frac{P(A_i)}{P(A_j)} \text{ with } P(A_i), P(A_j) > 0 \]  
  (4)

- Definitions (2) and (3) also imply that, if $A_i$ and $A_j$ are independent and $P(A_j) > 0$, then
  \[ P(A_i | A_j) = P(A_i) \]  
  (5)
Random Variables

- **Random variable** $S$: function of the sample space $\mathcal{O}$ that assigns a real value $S(\zeta)$ to each outcome $\zeta \in \mathcal{O}$ of a random experiment.
- **Cumulative distribution function (cdf)** of a random variable $S$:

\[
F_S(s) = P(S \leq s) = P(\{\zeta: S(\zeta) \leq s\})
\]

(6)

- Properties of cdf: non-decreasing, $F_S(-\infty) = 0$, and $F_S(\infty) = 1$
- **$N$-dimensional cdf, joint cdf, or joint distribution**:

\[
F_S(s) = P(S \leq s) = P(S_0 \leq s_0, \cdots, S_{N-1} \leq s_{N-1})
\]

(7)

with $S = \{S_0, \cdots, S_{N-1}\}^T$ being a random vector.

- Joint cdf of two random vectors $X$ and $Y$

\[
F_{XY}(x, y) = P(X \leq x, Y \leq y)
\]

(8)
Conditional Cdfs

- **Conditional cdf** of random variable $S$ given event $\mathcal{B}$ with $P(\mathcal{B}) > 0$

  \[
  F_{S|\mathcal{B}}(s | \mathcal{B}) = P(S \leq s | \mathcal{B}) = \frac{P(\{S \leq s\} \cap \mathcal{B})}{P(\mathcal{B})} \tag{9}
  \]

- Conditional cdf of a random variable $X$ given another random variable $Y$

  \[
  F_{X|Y}(x|y) = \frac{F_{XY}(x,y)}{F_Y(y)} = \frac{P(X \leq x, Y \leq y)}{P(Y \leq y)} \tag{10}
  \]

- Conditional cdf of a random vector $X$ given another random vector $Y$ is given by $F_{X|Y}(x|y) = F_{XY}(x,y)/F_Y(y)$
Continuous Random Variables

- If cdf $F_S(s)$ is a continuous function: *probability density function* (pdf)

\[
F_S(s) = \frac{dF_S(s)}{ds} \Leftrightarrow F_S(s) = \int_{-\infty}^{s} f_S(t) \, dt \tag{11}
\]

- Properties of pdfs: $f_S(s) \geq 0$, $\int_{-\infty}^{\infty} f_S(t) \, dt = 1$

  - **Uniform pdf:**
    \[
f_S(s) = \frac{1}{A} \quad \text{for} \quad -A/2 \leq s \leq A/2 \quad A > 0 \tag{12}
    \]

  - **Laplacian pdf:**
    \[
f_S(s) = \frac{1}{\sigma_S \sqrt{2}} e^{-|s-\mu_S|\sqrt{2}/\sigma_S} \quad \sigma_S > 0 \tag{13}
    \]

  - **Gaussian pdf:**
    \[
f_S(s) = \frac{1}{\sigma_S \sqrt{2\pi}} e^{-(s-\mu_S)^2/(2\sigma^2)} \quad \sigma_S > 0 \tag{14}
    \]
Generalized Gaussian Distribution

\[ f_S(s) = \frac{\beta}{2\alpha \Gamma(1/\beta)} \cdot e^{-|x-\mu|/\alpha}^{\beta} \]

\[ \Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt \]  

(15)
Joint and Conditional pdfs

- **$N$-dimensional pdf, joint pdf, or joint density**

\[
f_S(s) = \frac{\partial^N F_S(s)}{\partial s_0 \cdots \partial s_{N-1}}
\] (16)

- **Conditional pdf or conditional density** $f_{S|B}(s|B)$ of a random variable $S$ given an event $B$

\[
f_{S|B}(s|B) = \frac{dF_{S|B}(s|B)}{ds}
\] (17)

- Conditional density of a random vector $X$ given another random vector $Y$

\[
f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}
\] (18)
Discrete Random Variables

- **Discrete random variable** $S$: if its cdf $F_S(s)$ represents a staircase function
- $S$ takes values of countable set $\mathcal{A} = \{a_0, a_1, \ldots\}$
- **Probability mass function** (pmf):
  \[
  p_S(a) = P(S = a) = P(\{\zeta: S(\zeta) = a\}) \tag{19}
  \]
- Cdf of discrete random variable
  \[
  F_S(s) = \sum_{a \leq s} p(a) \tag{20}
  \]
- Binary pmf:
  \[
  \mathcal{A} = \{a_0, a_1\} \quad p_S(a_0) = p, \quad p_S(a_1) = 1 - p \tag{21}
  \]
- Uniform pmf:
  \[
  \mathcal{A} = \{a_0, a_1, \cdots, a_{M-1}\} \quad p_S(a_i) = 1/M \quad \forall a_i \in \mathcal{A} \tag{22}
  \]
- Geometric pmf:
  \[
  \mathcal{A} = \{a_0, a_1, \cdots\} \quad p_S(a_i) = (1 - p) p^i \quad \forall a_i \in \mathcal{A} \tag{23}
  \]
Joint and Conditional pmfs

- \emph{N-dimensional pmf} or \emph{joint pmf} for a random vector \( S = (S_0, \cdots, S_{N-1})^T \)

\[
p_S(a) = P(S = a) = P(S_0 = a_0, \cdots, S_{N-1} = a_{N-1})
\]  

(24)

- Joint pmf of two random vectors \( X \) and \( Y \): \( p_{XY}(a_x, a_y) \)

- \emph{Conditional pmf} \( p_{S|B}(a | B) \) of a random variable \( S \) given an event \( B \), with \( P(B) > 0 \)

\[
p_{S|B}(a | B) = P(S = a | B)
\]

(25)

- Conditional pmf of a random vector \( X \) given another random vector \( Y \)

\[
p_{X|Y}(a_x | a_y) = \frac{p_{XY}(a_x, a_y)}{p_Y(a_y)}
\]

(26)
Example for Joint pmf

- For example, samples in picture and video signals typically show strong statistical dependencies.
- Below: histogram of two horizontally adjacent samples for the picture 'Lena'.
Expectation

- **Expectation values or expected values**
  of continuous random variables $S$
  \[
  E \{g(S)\} = \int_{-\infty}^{\infty} g(s) \, f_S(s) \, ds \tag{27}
  \]
  of discrete random variables $S$
  \[
  E \{g(S)\} = \sum_{a \in A} g(a) \, p_S(a) \tag{28}
  \]

- Important expectation values are **mean** $\mu_S$ and **variance** $\sigma^2_S$
  \[
  \mu_S = E \{S\} \quad \text{and} \quad \sigma^2_S = E \{(S - \mu_s)^2\} \tag{29}
  \]

- Expectation value of a function $g(S)$ of a set $N$ random variables
  $S = \{S_0, \cdots, S_{N-1}\}$
  \[
  E \{g(S)\} = \int_{R^N} g(s) \, f_S(s) \, ds \tag{30}
  \]
Conditional Expectation

- **Conditional expectation value** of function $g(S)$ given an event $\mathcal{B}$, with $P(\mathcal{B}) > 0$

\[
E \{g(S) \mid \mathcal{B}\} = \int_{-\infty}^{\infty} g(s) f_{S \mid \mathcal{B}}(s \mid \mathcal{B}) \, ds
\]  

(31)

- Conditional expectation value of function $g(X)$ given a particular value $y$ for another random variable $Y$

\[
E \{g(X) \mid y\} = E \{g(X) \mid Y = y\} = \int_{-\infty}^{\infty} g(x) f_{X \mid Y}(x, y) \, dx
\]  

(32)
Random Processes

- Series of random experiments at time instants $t_n$, with $n = 0, 1, 2, \ldots$
- Outcome of experiment: random variable $S_n = S(t_n)$
- **Discrete-time random process**: series of random variables $S = \{S_n\}$
- Statistical properties of discrete-time random process $S$: $N$-th order joint cdf

\[
F_{S_k}(s) = P(S_k^{(N)} \leq s) = P(S_k \leq s_0, \cdots, S_{k+N-1} \leq s_{N-1})
\]  
(33)

- **Continuous random process**

\[
f_{S_k}(s) = \frac{\partial^N}{\partial s_0 \cdots \partial s_{N-1}} F_{S_k}(s)
\]  
(34)

- **Discrete random process**

\[
F_{S_k}(s) = \sum_{\mathbf{a} \in \mathcal{A}^N} p_{S_k}(\mathbf{a})
\]  
(35)

$\mathcal{A}^N$ product space of the alphabets $\mathcal{A}_n$ and

\[
p_{S_k}(\mathbf{a}) = P(S_k = a_0, \cdots, S_{k+N-1} = a_{N-1})
\]  
(36)
**Stationary Random Process**

- *Stationary random process*: statistical properties invariant to a shift in time
  - $N$-th order joint cdf $F_{S_k}(s)$, pdf $f_{S_k}(s)$, and pmf $p_{S_k}(a)$ are independent of time instant $t_k$ and are denoted by $F_S(s)$, $f_S(s)$, and $p_S(a)$, respectively
- $N$-th order autocovariance matrix
  \[
  C_N = E\{ (S^{(N)} - \mu_N)(S^{(N)} - \mu_N)^T \} 
  \]
  is a symmetric *Toeplitz matrix*
  \[
  C_N = \sigma_S^2 \begin{pmatrix}
  1 & \rho_1 & \rho_2 & \cdots & \rho_{N-1} \\
  \rho_1 & 1 & \rho_1 & \cdots & \rho_{N-2} \\
  \rho_2 & \rho_1 & 1 & \cdots & \rho_{N-3} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  \rho_{N-1} & \rho_{N-2} & \rho_{N-3} & \cdots & 1
  \end{pmatrix}
  \]
- For Toeplitz matrices, see the standard reference [Grenander and Szegö, 1958] and the tutorial [Gray, 2005]
Memoryless and i.i.d. Random Processes

- **Memoryless random process**: random process $\mathbf{S} = \{S_n\}$ for which the random variables $S_n$ are independent
- **Independent and identical distributed** (iid) random process: stationary and memoryless random process
- $N$-th order cdf $F_{\mathbf{S}}(s)$, pdf $f_{\mathbf{S}}(s)$, and pmf $p_{\mathbf{S}}(a)$ for iid processes, with $s = (s_0, \cdots, s_{N-1})^T$ and $a = (a_0, \cdots, a_{N-1})^T$

$$F_{\mathbf{S}}(s) = \prod_{k=0}^{N-1} F_S(s_k), \quad f_{\mathbf{S}}(s) = \prod_{k=0}^{N-1} f_S(s_k), \quad p_{\mathbf{S}}(a) = \prod_{k=0}^{N-1} p_S(a_k) \quad (39)$$

$F_{\mathbf{S}}(s)$, $f_{\mathbf{S}}(s)$, and $p_{\mathbf{S}}(a)$ are the marginal cdf, pdf, and pmf, respectively
Markov Processes

- **Markov process**: future outcomes do not depend on past outcomes, but only on the present outcome,

\[ P(S_n \leq s_n \mid S_{n-1} = s_{n-1}, \cdots) = P(S_n \leq s_n \mid S_{n-1} = s_{n-1}) \quad (40) \]

- Discrete processes

\[ p_{S_n}(a_n \mid a_{n-1}, \cdots) = p_{S_n}(a_n \mid a_{n-1}) \quad (41) \]

- Example for a discrete Markov process (calculate \( p(a) \))

<table>
<thead>
<tr>
<th>( a )</th>
<th>( a_0 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(a \mid a_0) )</td>
<td>0.90</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>( p(a \mid a_1) )</td>
<td>0.15</td>
<td>0.80</td>
<td>0.05</td>
</tr>
<tr>
<td>( p(a \mid a_2) )</td>
<td>0.25</td>
<td>0.15</td>
<td>0.60</td>
</tr>
<tr>
<td>( p(a) )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Example for a discrete Markov process (calculate \( p(a) \))

<table>
<thead>
<tr>
<th>( a )</th>
<th>( a_0 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p(a \mid a_0) )</td>
<td>0.90</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>( p(a \mid a_1) )</td>
<td>0.15</td>
<td>0.80</td>
<td>0.05</td>
</tr>
<tr>
<td>( p(a \mid a_2) )</td>
<td>0.25</td>
<td>0.15</td>
<td>0.60</td>
</tr>
<tr>
<td>( p(a) )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Markov Processes II

- Continuous Markov processes
  \[ f_{S_n}(s_n \mid s_{n-1}, \cdots) = f_{S_n}(s_n \mid s_{n-1}) \] (42)

- Given zero-mean iid process \( Z = \{Z_n\} \), continuous Markov process \( S = \{S_n\} \) with mean \( \mu_S \)
  \[ S_n = Z_n + \rho (S_{n-1} - \mu_S) + \mu_S, \quad \text{with} \quad |\rho| < 1 \] (43)

- Variance \( \sigma^2_S \) of stationary Markov process \( S \)
  \[ \sigma^2_S = E \{(S_n - \mu_S)^2\} = E \{(Z_n - \rho (S_{n-1} - \mu_S))^2\} = \frac{\sigma^2_Z}{1 - \rho^2} \] (44)

- Gauss-Markov Process, \( \rho = 0.9, \mu_S = 0 \)