Sufficient Statistics, Multivariate Gaussian Distribution
Course: Foundations in Digital Communications

Tobias Oechtering
presented by: Ragnar Thobaben

Royal Institute of Technology (KTH),
School of EE and ACCESS Center,
Communication Theory Lab
Stockholm, Sweden

5th lecture
Recapitulation

What did we do last lecture?
Outline - Motivation

- In practical systems we want/have to process our received data. What processing of the observed data does not reduce the possible detection performance?
  - Sufficient Statistics (chap 22)

- The most important multivariate distribution in Digital Communication:
  - Multivariate Gaussian Distribution (chap 23)
Introduction

“a sufficient statistic for guessing $M$ based on the observation $Y$ is a random variable or a collection of random variables that contains all the information in $Y$ that is relevant for guessing $M$”

- The idea of **sufficient statistics** ...
  - is a very deep concept with a strong impact;
  - provides fundamental intuition;
  - classifies processing which does not degrade performance;
  - is defined for $\{f_{Y|M}(\cdot|m)\}_{m \in M}$ and is unrelated to a prior.

- **Example**: In the 2-dimensional Gaussian 8-PSK detection problem, the decision is only based on the Euclidean distance between the observation and the symbols.
  - The scalar RV describing the distance is a sufficient statistic.
  ⇒ It summarizes the information needed for guessing $M$ optimally.
Definition and Main Consequences

- Roughly, $T(\cdot)$ is a sufficient statistic if there exists a black box that produces $\{\mathbb{P} \left[ M = m | Y = y_{obs} \right] \}$ when fed with $T(y_{obs})$ and any $\{\pi_m\}$.

**Definition**

A measurable mapping $T : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ forms a **sufficient statistic** for $\{f_{Y|M}(\cdot|m)\}_{m \in M}$ if there exist measurable functions $\psi_m : \mathbb{R}^{d'} \rightarrow [0, 1]$, $m \in M$, such that for every prior $\{\pi_m\}$ and almost all $y_{obs} \in \mathbb{R}^d$ where $\sum_m \pi_m f_{Y|M}(y_{obs}|m) > 0$ we have

$$\psi_m(\{\pi_m\}, T(y_{obs})) = \mathbb{P} \left[ M = m | Y = y_{obs} \right], \quad \forall m \in M.$$

- If $T(\cdot)$ is a sufficient statistic for $\{f_{Y|M}(\cdot|m)\}_{m \in M}$, then there exists an optimal decision rule based on $T(Y)$.
  - Note that $T(\cdot)$ does not have to be reversible.
Equivalent Conditions: Factorization Theorem

- Roughly, $T(\cdot)$ is a sufficient statistic if all densities can be written as a product of functions where
  - one does not depend on the message but possibly $y$
  - the other one depends on the message and $T(\cdot)$ only
- Useful in identifying sufficient statistics

### Factorization Theorem

$T(\cdot)$ denotes a sufficient statistic for $\{f_{Y|M}(\cdot|m)\}_{m \in M}$ iff there exist measurable functions $g_m : \mathbb{R}^{d'} \rightarrow [0, \infty)$, $m \in M$, and $h : \mathbb{R}^d \rightarrow [0, \infty)$ such that for almost all $y \in \mathbb{R}^d$ we have

$$f_{Y|M}(y|m) = g_m(T(y))h(y), \quad \forall m \in M.$$ 

**Proof idea:**

$\psi_m(\{\pi_m\}, T(y_{obs})) = \mathbb{P} \left[ M = m | Y = y_{obs} \right] = \frac{\pi_m f_{Y|M}(y_{obs}|m)}{f_Y(y_{obs})}$.

"⇒" Identify the functions as $g_m(T(y)) = \frac{\psi_m(\{\pi_m\}, T(y_{obs}))}{\pi_m}$ and $h(y) = f_Y(y)$. "⇐" Use $f_{Y|M}(\cdot|m) = g_m(T(\cdot))h(\cdot)$, $f_Y(\cdot) = \sum_m \pi_m f_{Y|M}(\cdot|m)$.
Markov Condition

- Tobias’ favorite:

### Markov condition

A measurable function \( T : \mathbb{R}^d \to \mathbb{R}^{d'} \) forms a sufficient condition for \( \{f_{Y|M}(\cdot|m)\}_{m \in M} \) iff for any prior \( \{\pi_m\}_m \) we have

\[
M - T(Y) - Y
\]

- \( M - T(Y) - Y \) means
  - \( M \) and \( Y \) are conditionally independent given \( T(Y) \)
  - equalities \( P_{M|T(Y)Y} = P_{M|T(Y)} \) and \( P_{Y|T(Y)M} = P_{Y|T(Y)} \)
- Since \( P_{M|Y} = P_{M|T(Y)Y} \) (\( T(Y) \) fct of \( Y \)), the last implies \( P_{M|Y} = P_{M|T(Y)} \).
  - The conditional distribution of \( M \) given \( Y \) follows from the conditional distribution of \( M \) given \( T(Y) \).
Pairwise Sufficiency and Simulating Observables

Pairwise Sufficiency

Consider \( \{f_{Y|M}(\cdot|m)\}_{m \in M} \), assume \( T(\cdot) \) forms a sufficient statistic for every pair \( f_{Y|M}(\cdot|m) \) and \( f_{Y|M}(\cdot|m') \) where \( m \neq m' \). Then \( T(\cdot) \) is a sufficient statistic for \( \{f_{Y|M}(\cdot|m)\}_{m \in M} \).

Simulating Observables (roughly statement)

Since sufficient statistic \( T(Y) \) contains all information about \( M \) which are in \( Y \), i.e., \( p_{M|T(Y)} = p_{M|Y} \), it is possible to generate a RV \( \tilde{Y} \) using \( T(Y) \) that appears statistically like \( Y \) given \( M \), i.e., \( p_{\tilde{Y}|M} \overset{\text{L}}{=} p_{Y|M} \). The opposite direction is also true, if such a function \( T(Y) \) exists, then it forms a sufficient statistic.

- This requires a local random number generator \( \Theta \).
- Anything learned about \( M \) from \( Y \) can be learned from \( \tilde{Y} \).
Identify Sufficient Statistics

- A not helpful result in terms of “summarizing” but still relevant:

**5-minute exercise**

Show that any reversible transformation $T(\cdot)$ forms a sufficient statistic.
Identify Sufficient Statistics

- A not helpful result in terms of “summarizing” but still relevant:

**5-minute exercise**

Show that any reversible transformation $T(\cdot)$ forms a sufficient statistic.

**Computable from the Statistic**

Let $T : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ form a sufficient statistic for $\{f_{Y|M}(\cdot|m)\}_{m \in M}$. If $T(\cdot)$ can be written as $\phi \circ S$ with $\phi : \mathbb{R}^{d''} \rightarrow \mathbb{R}^{d'}$, then $S : \mathbb{R}^d \rightarrow \mathbb{R}^{d''}$ also forms a sufficient statistic.

- If $T(Y)$ is computable from $S(Y)$, then $S(Y)$ has to contain all information about $M$ which are also in $T(Y)$, i.e.,

$$\mathbb{P} \left[ M = m | Y = y_{obs} \right]$$

is computable from $S(Y)$ as well.
Identify Sufficient Statistics

Two-step approach

If \( T : \mathbb{R}^d \to \mathbb{R}^{d'} \) forms a sufficient statistic for \( \{f_{Y|M}(\cdot|m)\}_{m \in M} \) and if \( S : \mathbb{R}^{d'} \to \mathbb{R}^{d''} \) forms a sufficient statistic for the corresponding densities of \( T(Y) \), then the composition \( S \circ T \) forms a sufficient statistic for \( \{f_{Y|M}(\cdot|m)\}_{m \in M} \).

**Proof:** \( P_{M|S(T(Y))} = P_{M|T(Y)} = P_{M|Y} \)

Conditionally Independent Observations

Let \( T_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d'_i} \) form sufficient statistics for \( \{f_{Y_i|M}(\cdot|m)\}_{m \in M}, i = 1, 2 \) and \( Y_1 \) and \( Y_2 \) are conditionally independent given \( M \), then \( (T_1(Y_1), T_2(Y_2)) \) forms a sufficient statistic for \( \{f_{Y_1Y_2|M}(\cdot|m)\}_{m \in M} \).

**Proof:** Factorization theorem: \( f_{Y_1Y_2|M} = f_{Y_1|M}f_{Y_2|M} = g_m^{(1)}h^{(1)}g_m^{(2)}h^{(2)} \)
Irrelevant Data

- Roughly, the “part” of the observation which is not part in a sufficient statistic is *irrelevant* for the purpose of detection.

**Definition**

*R* is said to be *irrelevant* for guessing *M* given *Y* if *Y* forms a sufficient statistic based on (*Y*, *R*), i.e., *M* − *Y* − (*Y*, *R*).

- A RV can be irrelevant, but still depend on the RV we wish to guess.

\[ R \perp M \land Y - M - R \implies R \text{ is irrelevant for guessing } M \text{ given } Y \]

**Proof:** Factorization theorem

\[ f_{YR|M}(y, r|m) = f_{Y|M}(y|m) f_{R|M}(r|m) = f_{Y|M}(y|m) f_R(r) = g_m(y) h(y, r) \]

\[ \square \]
Let’s take a break!
Some Results on Matrices

- Matrix $U \in \mathbb{R}^{n \times n}$ is orthogonal if $UU^T = I_n$ ($\Leftrightarrow U^T U = I_n$)
- Matrix $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$
- If $A \in \mathbb{R}^{n \times n}$ is symmetric, then $A$ has $n$ real eigenvalues with eigenvectors $\phi_v$ which satisfy $\phi_v^T \phi_{v'} = I \{v = v', 1 \leq v \leq n\}$.
  \[ \Rightarrow \text{Spectral Theorem: } A = U \Sigma U^T, \text{ with orthogonal } U \text{ whose } v\text{-th column is an eigenvector and diagonal matrix } \Sigma \text{ with the } v\text{-th eigenvalues on the } v\text{-th position on the diagonal.} \]
- A symmetric matrix $K \in \mathbb{R}^{n \times n}$ is called positive semidefinite or non-negative definite ($K \succeq 0$) if $\alpha^T K \alpha \geq 0$ for all $\alpha \in \mathbb{R}^n$ and is called positive definite ($K \succ 0$) if $\alpha^T K \alpha > 0$ for all $\alpha \in \mathbb{R}^n \setminus \{0\}$.
  - $K \succeq 0$ ($K \succ 0$)
    $\Leftrightarrow \exists$ (non-singular) $S \in \mathbb{R}^{n \times n}$: $K = S^T S$
    $\Leftrightarrow K$ symmetric and all eigenvalues are non-negative (positive)
    $\Leftrightarrow \exists$ orthogonal $U \in \mathbb{R}^{n \times n}$ and diagonal matrix $\Sigma \in \mathbb{R}^{n \times n}$ with non-negative (positive) diagonal entries: $K = U \Sigma U^T$. 

KTH course: Foundations in Digital Communications ©Tobias Oechtering
Random Vectors

- $n$-dimensional random vector $X$ defined over $(\Sigma, \mathcal{F}, P)$
  - mapping from experiment outcome $\Sigma$ to $\mathbb{R}^n$
  - density is the joint density of the components

**Expectation:** $\mathbb{E} [X] = (\mathbb{E} [X_1], \ldots, \mathbb{E} [X_n])^T$
  - $\mathbb{E} [AX] = A\mathbb{E} [X], A \in \mathbb{R}^{m \times n}$, and $\mathbb{E} [XB] = \mathbb{E} [X]B, B \in \mathbb{R}^{n \times m}$.

**Covariance matrix:**

$$K_{XX} = \mathbb{E} [(X - \mathbb{E} [X])(X - \mathbb{E} [X])^T]$$

- Let $Y = AX$, then $K_{YY} = AK_{XX}A^T$.
- Covariance matrix is non-negative definite, i.e., $K_{XX} \succeq 0$. 
Multivariate Gaussian Distribution

- Most important multi-variate distribution in Digital Communications
- Straightforward extension from univariate Gaussian

**Definition: Gaussian distribution**

1. For a **standard Gaussian** RV $W \in \mathbb{R}^n$ the components $\{W_i\}$ are independent and $\mathcal{N}(0, 1)$ distributed.

   \[
   f_W(w) = \prod_{\ell=1}^{n} \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{w_\ell^2}{2} \right) = \frac{1}{\sqrt{(2\pi)^n}} \exp \left( -\frac{||w||^2}{2} \right)
   \]

2. Then a RV $X \overset{\mathcal{L}}{=} AW$ with matrix $A \in \mathbb{R}^{n \times m}$ is said to be **centered Gaussian**

3. Additionally with $\mu \in \mathbb{R}^n$, the RV $X \overset{\mathcal{L}}{=} AW + \mu$ is **Gaussian**.
Properties Gaussian Random Vectors

- \( X \sim L A W + \mu \) and \( W \) standard \( \Rightarrow (E[X] = \mu \) and \( K_{XX} = AA^T) \)
- If the components of a Gaussian RV \( X \) are uncorrelated, the covariance matrix \( K_{XX} \) is diagonal and the components of \( X \) are independent.
- If the components of a Gaussian RV are pairwise independent, then they are independent.
- If \( W \) is standard Gaussian, and \( U \) is orthogonal matrix, then \( UW \) is also standard Gaussian RV.
- **Canonical Representation** of a centered Gaussian RV \( X \) with \( K_{XX} = U\Sigma U^T \), then \( X \sim L U\sigma^{1/2}W \) with \( W \) standard Gaussian.
  - From Gaussian to standard Gaussian: \( \Sigma^{1/2}U^T(X - \mu) \sim N(0, I_n) \).
Canonical Representation of a Centered Gaussian

Contour plot of centered Gaussian distributions

\[ X_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} W \quad X_2 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} W \quad X_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} W \quad X_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \sqrt{2} \\ -\frac{1}{\sqrt{2}} & \sqrt{2} \end{bmatrix} W \]
Jointly Gaussian Vectors

- Two RV $X$ and $Y$ are **jointly Gaussian** if the stacked vector $(X^T, Y^T)^T$ is Gaussian.

We have the following amazing results:

1. Independent Gaussian vectors are jointly Gaussian.
2. If two jointly Gaussian vectors are uncorrelated, then they are independent.
3. Let $X$ and $Y$ centered and jointly Gaussian with covariance matrices $K_{XX}$ and $K_{YY} > 0$. Then the conditional distribution of $X$ given $Y = y$ is a multivariate Gaussian with
   - mean $\mathbb{E}[XY^T]K_{YY}^{-1}y$
   - covariance $K_{XX} - \mathbb{E}[XY^T]K_{YY}^{-1}\mathbb{E}[Yx^T]$
Outlook - Assignment

- Sufficient Statistics
- Multivariate Gaussian Distributions

Next lecture

Complex Gaussian and Circular Symmetry, Continuous-Time Stochastic Processes

- Reading Assignment: Chap 24-25
- Homework: (please check with the official assignment on the webpage)
  - Problems in textbook: Exercise 22.2, 22.4, 22.5, 22.7, 22.9, 23.8, 23.11, and 23.14
  - Deadline: Dec 7