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 $\mathbf{Y}$ is a random variable or a collection of random variables that contains all the information in $\mathbf{Y}$ 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 $\{f_{\mathbf{Y}|M}(\cdot|m)\}_{m\in\mathcal{M}}$ 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.

  $\Rightarrow$ 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 $\{\mathbb{P}\left[M = m | \boldsymbol{Y} = \boldsymbol{y}_{obs}\right]\}$ when fed with $T(\boldsymbol{y}_{obs})$ and any $\{\pi_m\}$.

Definition

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

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

- If $T(\cdot)$ is a sufficient statistic for $\{f_{\boldsymbol{Y}|M}(\cdot|m)\}_{m \in \mathcal{M}}$, then there exists an optimal decision rule based on $T(\boldsymbol{Y})$.
  - 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 $\boldsymbol{y}$
  - 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 $\{f_{\boldsymbol{Y}|M}(\cdot|m)\}_{m\in\mathcal{M}}$ iff there exist measurable functions $g_m : \mathbb{R}^{d'} \to [0,\infty)$, $m \in \mathcal{M}$, and $h : \mathbb{R}^d \to [0,\infty)$ such that for almost all $\boldsymbol{y} \in \mathbb{R}^d$ we have

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

Proof idea: $\psi_m(\{\pi_m\}, T(\boldsymbol{y}_{obs})) = \mathbb{P}\left[M = m|\boldsymbol{Y} = \boldsymbol{y}_{obs}\right] = \frac{\pi_m f_{\boldsymbol{Y}|M}(\boldsymbol{y}_{obs}|m)}{f_{\boldsymbol{Y}}(\boldsymbol{y}_{obs})}$.

"$\Rightarrow$" Identify the functions as $g_m(T(\boldsymbol{y})) = \frac{\psi_m(\{\pi_m\},T(\boldsymbol{y}_{obs}))}{\pi_m}$ and $h(\boldsymbol{y}) = f_{\boldsymbol{Y}}(\boldsymbol{y})$. "$\Leftarrow$" Use $f_{\boldsymbol{Y}|M}(\cdot|m) = g_m(T(\cdot))h(\cdot)$, $f_{\boldsymbol{Y}}(\cdot) = \sum_m \pi_m f_{\boldsymbol{Y}|M}(\cdot|m)$.

Markov Condition

- Tobias' favorite:

Markov condition

A measurable function $T : \mathbb{R}^d \to \mathbb{R}^{d'}$ forms a sufficient condition for $\{f_{Y|M}(\cdot|m)\}_{m\in\mathcal{M}}$ 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 $\{f_{Y|M}(\cdot|m)\}_{m\in\mathcal{M}}$, 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 $\{f_{Y|M}(\cdot|m)\}_{m\in\mathcal{M}}$.

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., $p_{\tilde{Y}|M} \stackrel{\mathscr{L}}{=} p_{Y|M}$. 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 Θ.
- 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 $T : \mathbb{R}^d \to \mathbb{R}^{d'}$ form a sufficient statistic for $\{f_{\boldsymbol{Y}|M}(\cdot|m)\}_{m\in\mathcal{M}}$. If $T(\cdot)$ can be written as $\phi \circ S$ with $\phi : \mathbb{R}^{d''} \to \mathbb{R}^{d'}$, then $S : \mathbb{R}^d \to \mathbb{R}^{d''}$ also forms a sufficient statistic.

- If $T(\boldsymbol{Y})$ is computable from $S(\boldsymbol{Y})$, then $S(\boldsymbol{Y})$ has to contain all information about M which are also in $T(\boldsymbol{Y})$, i.e., $\mathbb{P}\left[M = m|Y = \boldsymbol{y}_{obs}\right]$ is computable from $S(\boldsymbol{Y})$ as well.

Identify Sufficient Statistics

Two-step approach

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

Proof: $P_{M|S(T(\boldsymbol{Y}))} = P_{M|T(\boldsymbol{Y})} = P_{M|\boldsymbol{Y}}$ □

Conditionally Independent Observations

Let $T_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d'_i}$ form sufficient statistics for $\{f_{\boldsymbol{Y}_i|M}(\cdot|m)\}_{m\in\mathcal{M}}$, $i = 1, 2$ and $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are conditionally independent given M, then $(T_1(\boldsymbol{Y}_1), T_2(\boldsymbol{Y}_2))$ forms a sufficient statistic for $\{f_{\boldsymbol{Y}_1\boldsymbol{Y}_2|M}(\cdot|m)\}_{m\in\mathcal{M}}$.

Proof: Factorization theorem: $f_{\boldsymbol{Y}_1\boldsymbol{Y}_2|M} = f_{\boldsymbol{Y}_1|M} f_{\boldsymbol{Y}_2|M} = g_m^{(1)} h^{(1)} g_m^{(2)} h^{(2)}$ □

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 $\boldsymbol{Y}$ if $\boldsymbol{Y}$ forms a sufficient statistic based on $(\boldsymbol{Y}, R)$, i.e., $M - \boldsymbol{Y} - (\boldsymbol{Y}, R)$.

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

$R \perp M$ & $\boldsymbol{Y} - M - R \quad \Rightarrow \quad R$ is *irrelevant* for guessing M given $\boldsymbol{Y}$

Proof: Factorization theorem

$$f_{\boldsymbol{Y}R|M}(\boldsymbol{y}, r|m) = f_{\boldsymbol{Y}|M}(\boldsymbol{y}|m) f_{R|M}(r|m) = f_{\boldsymbol{Y}|M}(\boldsymbol{y}|m) f_R(r) = g_m(\boldsymbol{y}) h(\boldsymbol{y}, r) \qquad \square$$

Let's take a break!

Some Results on Matrices

- Matrix $U \in \mathbb{R}^{n\times n}$ is **orthogonal** if $UU^T = I_n$ ($\Leftrightarrow U^TU = I_n$)
- 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 ϕ_ν which satisfy $\phi_\nu^T \phi_{\nu'} = \mathrm{I}\{\nu = \nu'\}$, $1 \le \nu \le n$.
  $\Rightarrow$ Spectral Theorem: $A = U\Sigma U^T$, with orthogonal U whose ν-th column is an eigenvector and diagonal matrix Σ with the ν-th eigenvalues on the ν-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 $\alpha^T K\alpha \ge 0$ 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: $K = U\Sigma U^T$.

Random Vectors

- n-dimensional random vector $\boldsymbol{X}$ defined over $(\Sigma, \mathcal{F}, P)$
  - mapping from experiment outcome Σ to $\mathbb{R}^n$
  - density is the joint density of the components
- **Expectation:** $\mathbb{E}[\boldsymbol{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])^T$
  - $\mathbb{E}[A\boldsymbol{X}] = A\mathbb{E}[\boldsymbol{X}]$, $A \in \mathbb{R}^{m\times n}$, and $\mathbb{E}[\boldsymbol{X}B] = \mathbb{E}[\boldsymbol{X}]B$, $B \in \mathbb{R}^{n\times m}$.
- **Covariance matrix:**

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

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

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 $\boldsymbol{W} \in \mathbb{R}^n$ the components $\{W_i\}$ are independent and $\mathcal{N}(0,1)$ distributed.

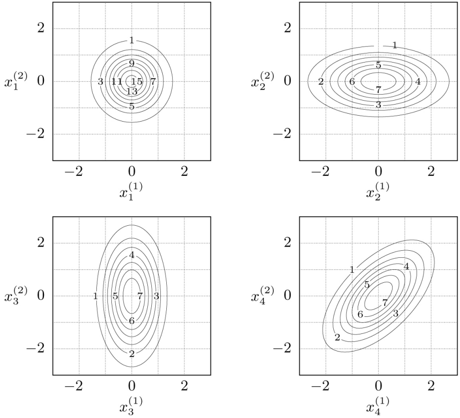
$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \prod_{\ell=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w_\ell^2}{2}\right) = \frac{1}{\sqrt{(2\pi)^n}} \exp\left(-\frac{\|\boldsymbol{w}\|^2}{2}\right)$$

2. Then a RV $\boldsymbol{X} \stackrel{\mathscr{L}}{=} A\boldsymbol{W}$ with matrix $A \in \mathbb{R}^{n\times m}$ is said to be **centered Gaussian**
3. Additionally with $\boldsymbol{\mu} \in \mathbb{R}^n$, the RV $\boldsymbol{X} \stackrel{\mathscr{L}}{=} A\boldsymbol{W} + \boldsymbol{\mu}$ is **Gaussian.**

Properties Gaussian Random Vectors

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

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 $\boldsymbol{X}$ and $\boldsymbol{Y}$ are **jointly Gaussian** if the stacked vector $(\boldsymbol{X}^T, \boldsymbol{Y}^T)^T$ 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 $\boldsymbol{X}$ and $\boldsymbol{Y}$ centered and jointly Gaussian with covariance matrices $K_{\boldsymbol{XX}}$ and $K_{\boldsymbol{YY}} \succ \boldsymbol{0}$. Then the conditional distribution of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ is a multivariate Gaussian with
   - mean $\mathbb{E}\left[\boldsymbol{X}\boldsymbol{Y}^T\right] K_{\boldsymbol{YY}}^{-1} \boldsymbol{y}$
   - covariance $K_{\boldsymbol{XX}} - \mathbb{E}\left[\boldsymbol{X}\boldsymbol{Y}^T\right] K_{\boldsymbol{YY}}^{-1} \mathbb{E}\left[\boldsymbol{Y}\boldsymbol{x}^T\right]$

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