Gaussian case. Linear Mean Squared Error Estimation. Quality of estimators.

Modern Theory of Detection and Estimation. Block-1: Estimation

Emilio Parrado-Hernández, emilio.parrado@uc3m.es

September 27, 2022

Universidad Carlos III de Madrid

Index

Estimation with Gaussian pdfs

- Estimation of **random variables** when the joint distribution of all the involved variables is a multivariate Gaussian.
- This is a very frequent case in real world problems.
- It is straightforward to proof that in this case all the marginals and all the conditionals will be Gaussian.
- Particularly, $p_{S|\mathbf{X}}(s|\mathbf{x})$ being Gaussian implies that the mean, the median and the mode of the posterior distribution coincide. Therefore $\hat{S}_{\text{MMSE}} = \hat{S}_{\text{MAD}} = \hat{S}_{\text{MAP}}$. We focus on $\hat{S}_{\text{MMSE}}$.

Index

Review of a Gaussian joint pdf for 1D variables

Assume $\mathbb{E}\{X\} = 0$ and $\mathbb{E}\{S\} = 0$. Then

$$p_{S,X}(s,x) \sim G\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} v_S & \rho \\ \rho & v_X \end{bmatrix} \right)$$

where v_S is the variance of S, v_X is the variance of X and $\rho = \mathbb{E}\{SX\} - \mathbb{E}\{S\}\mathbb{E}\{X\}$ is their covariance.
Then the posterior is given by

$$p_{S|X}(s|x) = \frac{p_{S,X}(s,x)}{p_X(x)}$$

$$= \frac{\frac{1}{2\pi\sqrt{v_X v_S - \rho^2}} \exp\left[-\frac{1}{2(v_X v_S - \rho^2)} \begin{bmatrix} s & x \end{bmatrix} \begin{bmatrix} v_X & -\rho \\ -\rho & v_S \end{bmatrix} \begin{bmatrix} s \\ x \end{bmatrix} \right]}{\frac{1}{\sqrt{2\pi v_X}} \exp\left[-\frac{x^2}{2v_X} \right]}$$

MMSE in 1D

We need the mean of $p_{S|X}(s|x)$. We exploit that $p_{S|X}(s|x)$ is Gaussian:

$$p_{S|X}(s|x) \sim G\left(m_{S|X}, v_{S|X}\right) = \frac{1}{\sqrt{2\pi v_{S|X}}} \exp\left[-\frac{(s - m_{S|X})^2}{2v_{S|X}}\right]$$

After developing both expressions for the same Gaussian and making term identification:

$$\frac{m_{S|X}^2}{v_{S|X}} = \frac{v_S x^2}{v_X v_S - \rho^2} - \frac{x^2}{v_X}$$

$$\frac{s\, m_{S|X}}{v_{S|X}} = \frac{\rho x s}{v_X v_S - \rho^2}$$

$$\frac{s^2}{v_{S|X}} = \frac{v_X s^2}{v_X v_S - \rho^2}$$

Therefore $\hat{s}_{\mathrm{MMSE}} = m_{S|X} = \frac{\rho}{v_X} x$. Notice it is a **linear function** of x.

Example: Estimation of a Gaussian signal with additive Gaussian noise

$X = S + N$, S is a Gaussian signal with zero mean and variance v_S. N is a Gaussian noise with zero mean and variance v_N independent of S.

We need to construct an estimator for S given X.

Example: Estimation of a Gaussian signal with additive Gaussian noise

$X = S + N$, S is a Gaussian signal with zero mean and variance v_S. N is a Gaussian noise with zero mean and variance v_N independent of S.

We need to construct an estimator for S given X.

According to the previous result $\hat{S} = \rho X / v_X$

$$\rho = \mathbb{E}\{(X - m_X)(S - m_S)\} = \mathbb{E}\{X\ S\} = \mathbb{E}\{(S + N)S\} = \mathbb{E}\{S^2\} + \mathbb{E}\{S\ N\} = v_S$$

$$v_X = v_S + v_N \text{ since they are independent}$$

Therefore $\hat{S} = \frac{v_S X}{v_S + v_N}$.

Physical interpretation when $v_S \gg v_N$ or $v_N \gg v_S$

Index

Joint multivariate Gaussian pdf

In a general multivariate case $\mathbf{S}$ and $\mathbf{X}$ are random vectors of dimensions N and M, respectively. Their joint pdf is

$$p_{\mathbf{S},\mathbf{X}}(\mathbf{s},\mathbf{x}) \sim G\left(\begin{bmatrix} \mathbf{m_S} \\ \mathbf{m_X} \end{bmatrix}, \begin{bmatrix} \mathbf{V_S} & \mathbf{V_{SX}} \\ \mathbf{V_{SX}^T} & \mathbf{V_X} \end{bmatrix}\right)$$

where $\mathbf{m_S}$ and $\mathbf{m_X}$ are the mean vectors and the covariances are

$$\mathbf{V_S} = \mathbb{E}\{(\mathbf{S}-\mathbf{m_S})(\mathbf{S}-\mathbf{m_S})^T\}$$
$$\mathbf{V_X} = \mathbb{E}\{(\mathbf{X}-\mathbf{m_X})(\mathbf{X}-\mathbf{m_X})^T\}$$
$$\mathbf{V_{SX}} = \mathbb{E}\{(\mathbf{S}-\mathbf{m_S})(\mathbf{X}-\mathbf{m_X})^T\}$$

Posterior distribution, multivariate case

The posterior $p_{\mathbf{S}|\mathbf{X}}(\mathbf{S}|\mathbf{X})$ is also a multivariate Gaussian with parameters

- mean:
$$\mathbf{m}_{\mathbf{S}|\mathbf{X}} = \mathbf{m}_{\mathbf{S}} + \mathbf{V}_{\mathbf{S}\mathbf{X}}\mathbf{V}_{\mathbf{X}}{}^{-1}(\mathbf{x} - \mathbf{m}_{\mathbf{X}})$$
- and covariance:
$$\mathbf{V}_{\mathbf{S}|\mathbf{X}} = \mathbf{V}_{\mathbf{S}} - \mathbf{V}_{\mathbf{S}\mathbf{X}}\mathbf{V}_{\mathbf{X}}{}^{-1}\mathbf{V}_{\mathbf{S}\mathbf{X}}^{T}$$

Therefore the MMSE estimator is given by

$$\hat{\mathbf{s}}_{\text{MMSE}} = \mathbb{E}\{\mathbf{s}|\mathbf{x}\} = \mathbf{m}_{\mathbf{S}} + \mathbf{V}_{\mathbf{S}\mathbf{X}}\mathbf{V}_{\mathbf{X}}{}^{-1}(\mathbf{x} - \mathbf{m}_{\mathbf{X}})$$

Index

Estimator with a fixed shape

- Sometimes you need to add a priori information about the estimation problem in the design of the estimator.
- Most of the times it means to fix a parametric shape for the estimation function $f_{\mathbf{w}}(\mathbf{X})$, with $\mathbf{w}$ a vector of parameters.
- Example: Observations in 2D and fix a shape for the estimator $\hat{S} = w_0 + w_1 X_1^2 + w_2 X_2^2$ (non-linear terms). The design task involves finding appropriate values for w_0, w_1 and w_2
- Minimize expected cost, but introducing the shape of the estimator as a constraint in the optimization

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} \ \mathbb{E}\{c(S, \hat{S})\} = \arg\min_{\mathbf{w}} \ \mathbb{E}\{c(S, f_{\mathbf{w}}(\mathbf{X}))\}$$

$$= \arg\min_{\mathbf{w}} \int_{\mathbf{x}} \int_{s} c(s, f_{\mathbf{w}}(\mathbf{x})) p_{S,\mathbf{X}}(s, \mathbf{x}) ds d\mathbf{x}$$

Example: Estimation with constraints

Two random variables S and X follow a joint pdf

$$p_{S,X}(s,x) = \left[\begin{array}{ll} \frac{1}{x}, & 0 < s < x < 1 \\ 0, & \text{otherwise} \end{array} \right.$$

Find the estimator of the form $\hat{s} = wx^2$ that minimizes the quadratic cost.

Example: Estimation with constraints

Two random variables S and X follow a joint pdf

$$p_{S,X}(s,x) = \left[\begin{array}{ll} \frac{1}{x}, & 0 < s < x < 1 \\ 0, & \text{otherwise} \end{array} \right.$$

Find the estimator of the form $\hat{s} = wx^2$ that minimizes the quadratic cost.

$$\hat{s} = \arg\min_{w} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (s - wx^2)^2 p_{S,X}(s,x) ds dx$$

Taking derivatives and making them equal to zero

$$\hat{s} = \frac{5}{8} x^2$$

Index

LMMSE Estimator motivation

- Bayesian estimation: minimize expected cost. Leads to
  - MSE: $\mathbb{E}\{s|\mathbf{X}\}$. You need $p_{S|\mathbf{X}}(s|\mathbf{x})$ and compute an integral
  - MAP: $\arg\max_s p_{S|\mathbf{X}}(s|\mathbf{x})$. You need also $p_{S|\mathbf{X}}(s|\mathbf{x})$ and maximize.
- Under Gaussian joint pdfs MAP and MSE estimators coincide and they are **linear**
- What if we can't access the complete $p_{S|\mathbf{X}}(s|\mathbf{x})$?

LMMSE essence

- Assume linearity $\hat{s} = \mathbf{w}^T\mathbf{x} + w_0$
- Minimize MSE

LMMSE properties

- Depends only on first and second order statistics
- Easy to evaluate
- In general LMMSE is suboptimal
- ... but optimal in the ubiquitous Gaussian case

Index

1D case

1. MMSE:
$$\hat{s} = \arg\min_{\hat{s}} \mathbb{E}\{(s-\hat{s})^2\}$$
2. Linearity
$$\hat{s} = w_0 + w_1 x \rightarrow \hat{s} = \arg\min_{w_0,w_1} \mathbb{E}\{(s - w_0 - w_1 x)^2\}$$
3. Optimizing
   - $\frac{\partial}{\partial w_0}\mathbb{E}\{(s-w_0-w_1x)^2\} = 0 \rightarrow \mathbb{E}\{(s-w_0-w_1x)\} = \mathbb{E}\{e\} = 0$: **Error with zero mean**
   - $\frac{\partial}{\partial w_1}\mathbb{E}\{(s-w_0-w_1x)^2\} = 0 \rightarrow \mathbb{E}\{(xe)\} = 0$ **Error orthogonal to observations**
4. Solution
$$w_1 = \frac{\text{Covariance}(x,s)}{\text{Variance}(x)}, \qquad w_0 = \mathbb{E}\{s\} - \frac{\text{Covariance}(x,s)}{\text{Variance}(x)}\mathbb{E}\{x\}$$
   Notice $w_0 = 0$ if x and s are zero mean.
5. MSE: $\mathbb{E}\{e^2\} = \mathbb{E}\{s^2\} - w_0\mathbb{E}\{s\} - w_1\mathbb{E}\{sx\}$

Multivariate case

1. $\hat{s} = w_0 + \sum_{j=1}^{d} w_j x_j = w_0 + \mathbf{w}^T\mathbf{x}$
2. $\hat{s} = \arg\min_{w_0,\mathbf{w}} \mathbb{E}\{(s - w_0 - \mathbf{w}^T\mathbf{x})^2\}$
3. Optimization
   - $\frac{\partial}{\partial w_0}\mathbb{E}\{(s - w_0 - \mathbf{w}^T\mathbf{x})^2\} = 0 \rightarrow \mathbb{E}\{e\} = 0$: **Error with zero mean** $w_0 = \mathbb{E}\{s\} - \mathbf{w}^T\mathbb{E}\{\mathbf{x}\}$
   - $\nabla_w \mathbb{E}\{(s - w_0 - \mathbf{w}^T\mathbf{x})^2\} = \mathbf{0} \rightarrow \mathbb{E}\{(\mathbf{x}e)\} = \mathbf{0}$ **Error orthogonal to each of the observed variables**
   - Consequently: **Error orthogonal to the estimator** $\mathbb{E}\{\hat{s}e\} = 0$
4. **Normal Equations** We develop the principle of orthogonality for the ith variable x_i:

$$\mathbb{E}\{(x_i e)\} = 0 \Rightarrow \mathbb{E}\{x_i s\} - w_0\mathbb{E}\{x_i\} - \sum_j w_j \mathbb{E}\{x_i x_j\} = 0$$

Substituting w_0

$$\mathbb{E}\{x_i s\} - \mathbb{E}\{s\}\mathbb{E}\{x_i\} + \sum_j w_j \mathbb{E}\{x_j\}\mathbb{E}\{x_i\} - \sum_j w_j \mathbb{E}\{x_i x_j\} = 0$$

Multivariate case

Remember

$$\mathrm{cov}(s, x_i) = \mathbb{E}\{x_i s\} - \mathbb{E}\{x_i\}\mathbb{E}\{s\}; \quad \mathrm{cov}(x_i, x_j) = \mathbb{E}\{x_i x_j\} - \mathbb{E}\{x_i\}\mathbb{E}\{x_j\}$$

Substituting covariances yields

$$\mathrm{cov}(s, x_i) - \sum_j w_j \mathrm{cov}(x_i, x_j) = 0, \quad \forall i = 1, \ldots, d$$

Stack the d equations (one for each x_i) and write in matrix form:

$$\mathbf{c}_{\mathbf{x},s} = C_{\mathbf{x},\mathbf{x}}\mathbf{w}$$

5. Solution:

$$\mathbf{w} = C_{\mathbf{x},\mathbf{x}}^{-1}\mathbf{c}_{\mathbf{x},s}$$

$$w_0 = w_0 = \mathbb{E}\{s\} - \mathbf{c}_{\mathbf{x},s}^T C_{\mathbf{x},\mathbf{x}}^{-1}\mathbb{E}\{\mathbf{x}\}$$

6. MSE: $\mathbb{E}\{e^2\} = \mathrm{var}(s) - \mathbf{c}_{\mathbf{x},s}^T C_{\mathbf{x},\mathbf{x}}^{-1}\mathbf{c}_{\mathbf{x},s}$

Example

The statistical relationship between S and the observed variables X_1 and X_2 is given by:

$$\begin{array}{lll} \mathbb{E}\{S\} = 1/2 & \mathbb{E}\{X_1\} = 1 & \mathbb{E}\{X_2\} = 0 \\ \mathbb{E}\{S^2\} = 4 & \mathbb{E}\{X_1^2\} = 3/2 & \mathbb{E}\{X_2^2\} = 2 \\ \mathbb{E}\{SX_1\} = 1 & \mathbb{E}\{SX_2\} = 2 & \mathbb{E}\{X_1X_2\} = 1/2 \end{array}$$

Determine the LMMSE estimator of S given the observations.

Example

The statistical relationship between S and the observed variables X_1 and X_2 is given by:

$$\begin{array}{lll} \mathbb{E}\{S\} = 1/2 & \mathbb{E}\{X_1\} = 1 & \mathbb{E}\{X_2\} = 0 \\ \mathbb{E}\{S^2\} = 4 & \mathbb{E}\{X_1^2\} = 3/2 & \mathbb{E}\{X_2^2\} = 2 \\ \mathbb{E}\{SX_1\} = 1 & \mathbb{E}\{SX_2\} = 2 & \mathbb{E}\{X_1X_2\} = 1/2 \end{array}$$

Determine the LMMSE estimator of S given the observations.

$$\hat{s} = \mathbf{c}_{S,\mathbf{X}}^T C_{\mathbf{x},\mathbf{x}}^{-1}\mathbf{x} + \mathbb{E}\{s\} - \mathbf{c}_{\mathbf{x},s}^T C_{\mathbf{x},\mathbf{x}}^{-1}\mathbb{E}\{\mathbf{x}\}$$

$$\hat{s} = \left[\begin{array}{cc} .5 & 2 \end{array}\right] \left[\begin{array}{cc} 0.5 & 0.5 \\ 0.5 & 2 \end{array}\right]^{-1} \left[\begin{array}{c} X_1 \\ X_2 \end{array}\right] + 0.5 - \left[\begin{array}{cc} .5 & 2 \end{array}\right] \left[\begin{array}{cc} 0.5 & 0.5 \\ 0.5 & 2 \end{array}\right]^{-1} \left[\begin{array}{c} 1 \\ 0 \end{array}\right]$$

Index

Index

Assessing the quality of an estimator

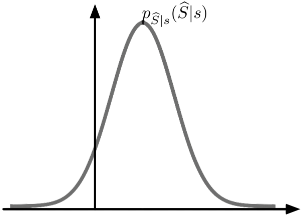
- One can design different estimators to work in a same scenario (problem).
- Fix **criteria** that enable a **fair comparison** between estimators.
- **Expected cost** for a determined cost function leads to always choose the Bayesian estimator as optimum.
- **Other measures** that can be of interest in different scenarios you may come accross.
  - **Bias** ($\sim$ systematic error)
  - **Variance** (concentration of the estimations around their expected value)

  Important! Bias and variance depend on if the variable to be estimated is random or deterministic.

Index

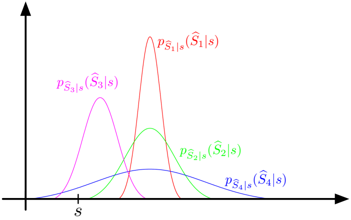
Probability density of the estimator

- $p_{\widehat{S}|s}(\widehat{s}|s)$ provides a complete characterization of the behaviour of the estimator.
  (Since $\widehat{S}$ is a function of the observations, $\widehat{S} = f(\mathbf{X})$, $p_{\widehat{S}|s}(\widehat{s}|s)$ can be obtained from $p_{\mathbf{X}|s}(\mathbf{X}|s)$ with a transformation of random variable.



Motivation

Imagine a case of estimation of a parameter s with 4 different estimators ($\widehat{S}_1 = f_1(\mathbf{X})$, $\widehat{S}_2 = f_2(\mathbf{X})$, $\widehat{S}_3 = f_3(\mathbf{X})$ and $\widehat{S}_4 = f_4(\mathbf{X})$) Examining their pdfs, which estimator seems more appropriate?



Perhaps $\widehat{S}_3$ as the probability of getting estimations close to the true s is significantly larger.

Bias and variance when s is deterministic

- **Bias**:
  - Expresses how far is the mean of $p_{\widehat{S}|s}(\widehat{s}|s)$ from the true value of s

$$\text{Bias}\left\{\widehat{S}|s\right\} = \mathbb{E}\left\{s - \widehat{S}|s\right\} = \mathbb{E}\{s|s\} - \mathbb{E}\left\{\widehat{S}|s\right\} = s - \mathbb{E}\left\{\widehat{S}|s\right\}$$

- **Variance**:

$$\text{Variance}\left\{\widehat{S}|s\right\} = \mathbb{E}\left\{\left(\widehat{S} - \mathbb{E}\left\{\widehat{S}\right\}\right)^2 \Big| s\right\} = \mathbb{E}\left\{\widehat{S}^2|s\right\} - \mathbb{E}^2\left\{\widehat{S}|s\right\}$$

If s is a deterministic parameter, the variance of the estimator coincides with the estimation error.

$$\text{Variance}\left\{\widehat{S} - s\Big|s\right\} = \text{Variance}\left\{\widehat{S}|s\right\} - \text{Variance}\{s|s\} = \text{Variance}\left\{\widehat{S}|s\right\}$$

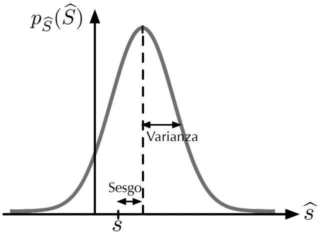
Important! The bias and the variance of the estimators of a deterministic parameter are a function of its true value (s).

Properties

Bias and variance of the estimador of a deterministic paramter

$\text{Bias}\{\widehat{S}|s\} = s - \mathbb{E}\{\widehat{S}|s\}$

$\text{Variance}\left\{\widehat{S}|s\right\} = \mathbb{E}\{\widehat{S}^2|s\} - \mathbb{E}^2\{\widehat{S}|s\}$



- The estimators with **zero bias** are known as **unbiased estimators**.
- If the estimator operates on a number K of observations of a random variable and $\text{Variance}(\widehat{S}) \to 0$ if $K \to \infty$, the estimator is **consistent in variance**.

Mean squared error

Mean squared error of the estimator of a deterministic parameter

$$\begin{aligned}\mathbb{E}\Big\{(s-\widehat{S})^2\big|s\Big\} &= \text{Variance}\Big\{\widehat{S}-s\big|s\Big\} + \mathbb{E}^2\{s-\widehat{S}|s\} \\ &= \text{Variance}\{\widehat{S}|s\} + \Big[\text{Bias}(\widehat{S}|s)\Big]^2\end{aligned}$$

Example

Calculate the bias and variance of the sample estimation of the mean of a random variable.

Index

General vision

- As in the deterministic case, one could directly use $p_{\widehat{S}|S}(\widehat{s}|s)$ to study the performance of an estimador.
- However, when S is a random variable, the true value $(S = s)$ changes every time we repeat the experiment of drawing the observations and running the estimator.

Bias and variance

Bias and variance of the estimator of a random variable

$$\text{Bias}\{\widehat{S}\} = \mathbb{E}\{S - \widehat{S}\} = \mathbb{E}\{S\} - \mathbb{E}\{\widehat{S}\}$$
$$\text{Variance}\left\{\widehat{S}\right\} = \mathbb{E}\left\{\left(\widehat{S} - \mathbb{E}\{\widehat{S}\}\right)^2\right\} = \mathbb{E}\{\widehat{S}^2\} - \mathbb{E}^2\{\widehat{S}\}$$

The calculation of these expectations uses the joint pdf of S and $\mathbf{X}$.

Important! The bias and variance of the estimators of a random variable **are not** a **function** of the true value of S (s).

Mean squared error

The mean squared error of the estimator of a deterministic parameter was deterministic

$$\begin{aligned}\mathbb{E}\left\{(S-\widehat{S})^2\right\} &= \text{Variance}\left\{\widehat{S}-S\right\} + \mathbb{E}^2\{S-\widehat{S}\} \\ &= \text{Variance}\{\widehat{E}\} + \left[\text{Bias}(\widehat{S})\right]^2\end{aligned}$$

If $(\widehat{S})$ is random, the variance of the error in general will not be equal to the variance of the estimator.

Properties

- The **unconstrained minimum mean square error estimator is unbiased**:

$$\begin{aligned}
\mathrm{Bias}\{\widehat{S}_{\mathrm{MMSE}}\} &= \mathbb{E}\Big\{S - \widehat{S}_{\mathrm{MMSE}}\Big\} = \mathbb{E}\{S\} - \mathbb{E}\Big\{\widehat{S}_{\mathrm{MMSE}}\Big\} \\
&= \mathbb{E}\{S\} - \int \mathbb{E}\ \{S|\mathbf{X} = \mathbf{x}\} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \mathbb{E}\{S\} - \mathbb{E}\{S\} \\
&= 0
\end{aligned}$$

- The **linear MMSE** is **unbiased**.
  $(\mathbb{E}(E^*) = 0)$