TTK4250

Week 1

Probability and estimation

Edmund Førland Brekke

22. August 2024

Formalia

Instructors.

- Lecturer: Edmund Førland Brekke, `edmund.brekke@ntnu.no`, Room D335.
- Scientific assistant: Henrik Dobbe Flemmen, `henrik.d.flemmen@ntnu.no`.

Blackboard.

All information about the course will be made available on Blackboard, so it is recommended to visit this regularly. (Ensure that you receive notifications)

Reference group.

- Reference groups are an important part of NTNUs quality assurance of education
- Referansegruppen consists of three students, with the responsibility to
  - have three meetings during the semester with the lecturer.
  - have a dialogue with the class during the semester.
  - write a reference group report that sums up the students' opinions and recommendations for improvement of the course. This report is included in the final course report.

Expectations and prerequisites

About this course:

This will introduce you to the **probabilistic framework** that dominates in sensor fusion. The course will not cover very much of sensor fusion itself, but there will be a strong focus on **algorithms and methods** that play a central role in typical sensor fusion applications, such as **target tracking**, **inertial navigation** and **SLAM**.

Background and prerequisites that I expect you to have:

- You must have had linear systems theory or any other course where you have either became familiar with the **Kalman filter** or **multivariate Gaussians**.
- The course builds directly on TMA4240 Statistics, but key concepts that often are forgotten will be repeated in the first lectures.
- You are recommended to have additional background in statistics and probability. TTT4275 - Estimation, Detection and Classification is an ideal background, but any statistics course from IMF will also be very useful.
- The course will include 2 programming assignments to be solved in Python. Therefore you must be comfortable with programming.

Overview

Teaching material.

- Edmund Brekke: *Fundamentals of sensor fusion*, 2024.
- All lecture notes, assignments and solutions to the assignments.

Topics that we will study.

1. Probability and estimation.
2. The multivariate Gaussian.
3. The Kalman filter.
4. Nonlinear filters: EKF and particle filters.
5. The Interacting Multiple Models (IMM) method.
6. Single-target tracking.
7. Multi-target tracking.
8. Inertial navigation.
9. Probabilistic graphical models.
10. Simultaneous localization and mapping (SLAM).

Structure of the course

Lectures

- Thursdays 08.15-09.00 in H1.
- Fridays 08.15-10.00 in R10.

Textbook

Read as much as you can as early as possible.

Assignments

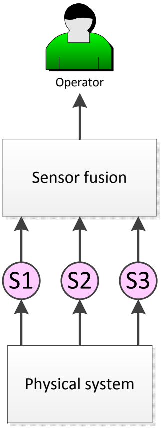
- 2 computer assignments, which both are compulsory.
- 6 written assignments: At least 4 of these must be be accepted. Hand-in through Blackboard.

Evaluation

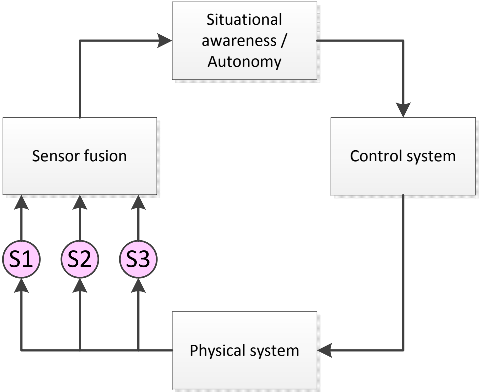
- 70% exam: Written or home exam, to be decided.
- 30% computer assignments and report based on these.

A detailed course schedule can be found on Blackboard.

Two roles of sensor fusion



Sensor fusion for surveillance and decision support

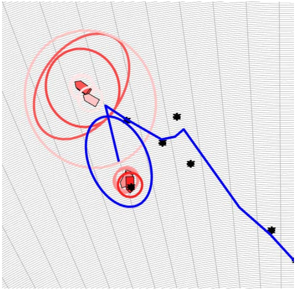


Sensor fusion in a closed-loop system

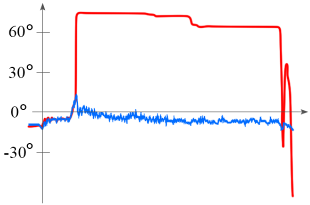
Why is sensor fusion important?

- Because challenges are non-trivial.
- Because poorly designed methods can have disastrous consequences.

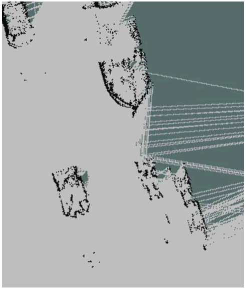
Competing data interpretations



Are estimates physically reasonable?



Quantify lack of knowledge



Why probability?

Most industry-standard methods in target tracking and navigation, and to a large extent also in SLAM, are based on probability theory.

- Probability theory is a language for quantifying uncertainty.

$\longrightarrow$ Enable the algorithms to **hedge** on different possibilities in the same way as a hedge fund manager.

$\longrightarrow$ The uncertainty of the algorithms can have implications for what decisions are rational.

- Correlations, Bayes' rule and other probabilistic tools can help us make inference when limited/noisy data is available.

The axioms of probability

- A probability $\Pr\{\cdot\}$ is **something that we assign to an event E**.
- The union of all possible events is called the outcome space $\mathbb{O}$.
- The probability obeys the following three axioms:

1. $\Pr\{E\} \in \mathbb{R}$, $\Pr\{E\} \geq 0$.
2. $\Pr\{\mathbb{O}\} = 1$.
3. For any sequence of disjoint events $E_1, \ldots, E_n$ we have

$$\Pr\left\{\bigcup_{i=1}^{n} E_i\right\} = \sum_{i=1}^{n} \Pr\{E_i\}.$$

Bread-and-butter definitions and rules

Definition: Conditional probability

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

Definition: Independence

The events A and B are independent if $\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}$.

The total probability theorem

$$\Pr\{A\} = \sum_n \Pr\{A|B_n\}\Pr\{B_n\}$$

Bayes' rule

$$\Pr\{A|B\} = \frac{\Pr\{B|A\}\Pr\{A\}}{\Pr\{B\}}$$

Random variables: Core concepts

Random variables

A quantity X is a random variable when it is uncertain and our knowledge about it is given in terms of probabilities.

Realization

A particular value x that the random variable X can attain.

Outcome space

The space $\mathbb{O}$ of all realizations for a random variable.

Probability measure

A function from subsets of the outcome space to the interval $[0, 1]$, that returns the probability that X is within a subset.

Random variables: Core concepts

Example: Speed of an airplane

- Possible realizations of X are $50\,\mathrm{m/\,s}$ or $100\,\mathrm{m/\,s}$.
- A reasonable outcome space could be $[0, \infty)$.
- The event that $X \in [50\,\mathrm{m/\,s}, 100\,\mathrm{m/\,s}]$ has some probability, which cannot be smaller than the probability of the event $X \in [50\,\mathrm{m/\,s}, 60\,\mathrm{m/\,s}]$.

Example: Throwing a dice

- Possible realizations of X are 1, 2, . . ., 6.
- The outcome space is $\{1, 2, 3, 4, 5, 6\}$.
- The event that $X \in \{1, 2\}$ has some probability, which cannot be smaller than the probability of the event $X = 1$.

What is the probability of the event that $X \in \{1, 2, 3, 4, 5, 6\}$?

Probability distributions

Definition: Cumulative distribution function (cdf)

The cdf, denoted $P(x)$, of $X \in \mathbb{R}$ is the probability $\Pr\{X \leq x\}$.

Definition: Probability density function (pdf)

The pdf of the scalar random variable X is the derivative

$$p(x) = \frac{\partial P(x)}{\partial x}.$$

Pdf's and probabilities are two different things. We get probabilities when we integrate a pdf over a subset of the outcome space.

The definitions are easily extended to $n = 2, 3, \ldots$ for $\mathbb{R}^n$:

$$P(x, y) = \Pr\{X \leq x, Y \leq y\}$$
$$p(x, y) = \frac{\partial^2}{\partial x \partial y} P(x, y) = \frac{\partial^2}{\partial y \partial x} P(x, y).$$

Examples of discrete probability distributions

- For a discrete random variable we can again define the cdf as $P(x) = \Pr\{X < x\}$.
- The pdf will then contain δ-spikes at all possible values in the outcome space.
- Since the outcome space is discrete, we can just as well work directly with the probabilities for different realizations.

The randomness of a discrete random variable is quantified by the probability mass function (pmf) $p(x)$.

The Bernoulli distribution

A Bernoulli random variable with parameter r has a binary outcome space: $E = \{0, 1\}$, and its pmf is given by

$$p(x) = \begin{cases} 1 - r & \text{if } x = 0 \\ r & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

We write $p(x) = \mathrm{Bernoulli}(x; r)$ to signify this distribution.

Examples of discrete probability distributions

The Binomial distribution

A binomial random variable with parameters $r \in [0, 1]$ and $n \in \mathbb{N}$ has outcome space $\{0, \ldots, n\}$ and its probability distribution is given by

$$p(x) = \binom{n}{x} r^x (1-r)^{n-x}.$$

The Poisson distribution

A Poisson random variable with parameter λ has the countable (that means infinite, but discrete) outcome space $\{0, 1, 2, 3, \ldots\}$ and its probability distribution is given by

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

We write $p(x) = \mathrm{Poisson}(x; \lambda)$ to signify this distribution.

Examples of continuous probability distributions

The uniform distribution

A uniformly distributed random variable X on the interval $[a, b]$ has the pdf

$$p(x) = \mathrm{Uniform}(x\,;\,[a, b]) = \frac{1}{b-a}\chi_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

The Gaussian distribution

A Gaussian random variable with expectation μ and variance σ^2 has the pdf

$$p(x) = \mathcal{N}(x\,;\,\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

The cdf of the Gaussian does not exist in closed form. It is typically expressed in terms of the so-called error function according to

$$P(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \quad \text{where} \quad \mathrm{erf}(x) = \frac{1}{\sqrt{\pi}}\int_0^x \exp(-t^2)\mathrm{d}t$$

Examples of continuous probability distributions

The Gamma distribution

A Gamma random variable with shape parameter k and scale parameter θ has the pdf

$$p(x) = \mathrm{Gamma}(x\,;\,k, \theta) = \frac{x^{k-1}\exp(-x/\theta)}{\theta^k\Gamma(k)}$$

Several special cases are of importance.

- The exponential distribution results if $k = 1$ and $\theta = 1/\lambda$:

$$p(x) = \lambda e^{-\lambda x}$$

- The χ^2 distribution results if $k = \frac{n}{2}$ and $\theta = 2$:

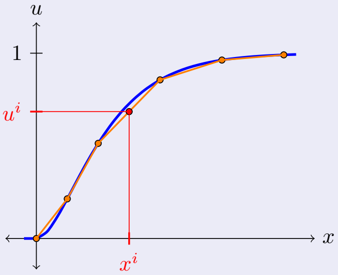
$$p(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} \exp\left(-\frac{x}{2}\right)$$

Sampling from continuous probability distributions

Option 1

Use inbuilt functions such as `randn(4,1)` and `rand(4,1)`, etc.

Option 2



Use cdf inversion, also known as the Smirnov transform or inverse transform sampling.

Inverse transform sampling is easy if the cdf is invertible in closed form, but can also be used by means of interpolation if this is not the case.

More complicated random variables can often be simulated by exploiting their relationships to other random variables.

Moments

The most important moments are the expectation and the covariance, which for random vectors are

$$E[X] = \int \mathbf{x} p(\mathbf{x}) \mathrm{d}\mathbf{x}$$

$$\mathrm{Var}[X] = E[(X - E[X])(X - E[X])^\top] = \int (\mathbf{x} - E[X])(\mathbf{x} - E[X])^\top p(\mathbf{x}) \mathrm{d}\mathbf{x}.$$

These are related to, but not the same as, the sample mean and the sample covariance

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$$

$$\mathbf{P} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$$

- Moments summarize information **about the pdf** in a single number (vector, matric, etc.).
- The sample mean and sample covariance say something **about a sample** that may more may not be drawn from some pdf.

Useful rules for expectation and variance

- Expectation of linear combinations

$$E[\mathbf{A}X + \mathbf{B}Y] = \mathbf{A}E[X] + \mathbf{B}E[Y].$$

- Variance of linear combinations

$$\mathrm{Var}[\mathbf{A}X + \mathbf{B}Y] = \mathbf{A}\mathrm{Var}(X)\mathbf{A}^{\mathsf{T}} + \mathbf{B}\mathrm{Var}(Y)\mathbf{B}^{\mathsf{T}} + \mathbf{A}\mathrm{Cov}(X, Y)\mathbf{B}^{\mathsf{T}} + \mathbf{B}\mathrm{Cov}(Y, X)\mathbf{A}^{\mathsf{T}}.$$

- Independence implies that ...

$$E_{X,Y}[XY] = E_X[E_Y[XY]] = E_Y[E_X[X|Y]Y] = E_X[X]E_Y[Y].$$

- Jensen's inequality: For any convex function $g(\cdot)$ the following must hold:

$$g(E[X]) \leq E[g(X)].$$

Higher-order moments and heavytailedness

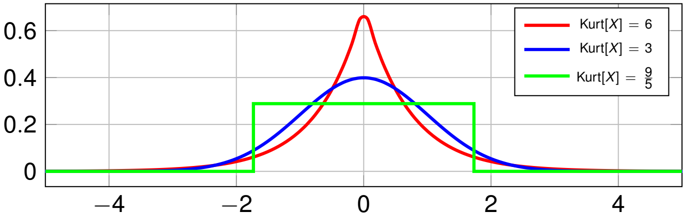
Third and fourth order moments

- The skewness tells us whether the distribution is more spread out on one side of the expectation:

$$\gamma = \frac{E[(X-\mu)^3]}{\sigma^3}.$$

- The kurtosis tells us whether the distribution has heavier or lighter tails than a Gaussian (reference value 3):

$$\mathrm{Kurt}[X] = \frac{E[(X-\mu)^4]}{\sigma^4}$$



Generating functions

Moment-generating function

For a continuous random variable we use $M_X(s) = E_X[e^{sx}] = \int_{-\infty}^{\infty} p(x)e^{sx}\mathrm{d}x.$

Probability-generating function

For a discrete random variable we use $G(t) = E_X[t^X] = \sum_{x=-\infty}^{\infty} p(x)t^x.$

Generating functions are useful because

1. The generating function determines the distribution and vice versa, in a manner similar to Laplace- and Z-transforms.
2. The generating function of a sum of independent random variables is the product of the generating functions.
3. The moments can be found by differentiating the generating function.

Examples of use of generating functions

Sum of Gaussians

Let X and Y be two independent Gaussian RVs with expectations a and b, and covariances q and r. What is the distribution of $Z = X + Y$?

Sum of Exponentials

Let X_i, $i = 1, \ldots, N$ be N i.i.d. exponential RVs with parameter λ, and let

$$Y = \sum_{i=1}^{N} X_i.$$

What is the distribution of Y?

Generalizations

- For a vector-valued random variable the moment-generating function is

$$M_X(\mathbf{s}) = E_X[e^{\mathbf{s}^\mathsf{T}\mathbf{x}}] = \int_\infty p(\mathbf{x}) e^{\mathbf{s}^\mathsf{T}\mathbf{x}} \mathrm{d}\mathbf{x}.$$

- In the PhD level course TK8102 probability-generating functions are generalized to **probability-generating functionals**.

Transformations of random variables

The other important tool that we have to derive new distributions is the following theorem.

Nonlinear transformations of random variables

Suppose that $\mathbf{y} = \mathbf{f}(\mathbf{x})$ where $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Denote the pdf's of $\mathbf{x}$ and $\mathbf{y}$ by $g(\mathbf{x})$ and $h(\mathbf{y})$, respectively. Then we have that

$$h(\mathbf{y}) = \sum_i g(\mathbf{f}_i^{-1}(\mathbf{y})) \, |\det(\mathbf{F}_i^{-1}(\mathbf{y}))|$$

where $\mathbf{f}_i^{-1}(\mathbf{y})$ range over all solutions of $\mathbf{y} = \mathbf{f}(\mathbf{x})$ with respect to $\mathbf{x}$, and $\mathbf{F}_i^{-1}(\mathbf{y})$ is the corresponding Jacobian matrix of the inverse mapping $\mathbf{f}_i^{-1}(\mathbf{y})$.

Square of zero-mean univariate Gaussian

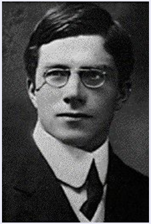
If $X \sim \mathcal{N}(0, 1)$, what is then the pdf of $Y = X^2$?

The Bayesian and frequentist paradigms

The frequentist approach

- All probability should be interpretable as a frequency.
- In the frequentist mindset, a quantity to be estimated is not random. Only the data are random.

Famous champion: Ronald Fisher.



The Bayesian approach

- Bayesians are inclined to represent all uncertainty probabilistically, even if they have to make subjective assignments of probabilities.
- Sensor fusion methods such as the Kalman filter are Bayesian because they rely on *a priori* uncertainties (e.g., in the process model).

Famous champion: Pierre-Simon Laplace.

Bayes' rule is valid for pdfs, not only for probabilities

To clarify what we mean by

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)} \tag{1}$$

let us first introduce the events A and B given by

$$\begin{aligned} A =& \{\ z \leq Z \leq z + \Delta z\ \} \\ B =& \{\ x \leq X \leq x + \Delta x\ \}. \end{aligned}$$

We can then define the conditional pdfs in (1) as follows:

$$\begin{aligned} p(z\,|\,x) &= \lim_{\Delta x \to 0} p(z\,|\,B) = \lim_{\Delta z \to 0} \lim_{\Delta x \to 0} \frac{P(z + \Delta z\,|\,B) - P(z\,|\,B)}{\Delta z} \\ p(x\,|\,z) &= \lim_{\Delta z \to 0} p(x\,|\,A) = \lim_{\Delta x \to 0} \lim_{\Delta z \to 0} \frac{P(x + \Delta x\,|\,A) - P(x\,|\,A)}{\Delta x}. \end{aligned}$$

Bayes' rule clearly holds for the events A and B. Its validity can be transferred to the pdf's via the cdf's by taking the appropriate limits. Details are in the book.

Bayes' rule is also valid for hybrid cases

Example: Poisson likelihood and Gamma prior

Suppose that $p(z|x) = \mathrm{Poisson}(z\,;\, x)$ and that $p(x) = \mathrm{Gamma}(x\,;\, k, \theta)$.

Then it can be shown that

$$p(z) = \int p(z|x)p(x)\,\mathrm{d}x = e^{-z}\theta^{z}\frac{\Gamma(z+k)}{\Gamma(z+1)\Gamma(k)}$$

and furthermore that

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)} = \frac{\dfrac{x^{z}e^{-z}}{z!}\cdot\dfrac{x^{k-1}e^{-x/\theta}}{\theta^{k}\Gamma(k)}}{e^{-z}\theta^{z}\dfrac{\Gamma(z+k)}{\Gamma(z+1)\Gamma(k)}} = \mathrm{Gamma}(x\,;\, z+k, \theta).$$

Observations:

- We see that it is straightforward to combine discrete and continuous uncertainty.
- In this particular example we see an example of a **conjugate prior**: The posterior ends up being of the same form as the prior.

Estimators

Definition: Estimator.

We have some data z that were generated at random according to $p(z|x)$. We would like to infer knowledge about x from z. Let $x \in \mathcal{X}$ and let $z \in \mathcal{Z}$. An estimator is a function $\theta : \mathcal{Z} \rightarrow \mathcal{X}$ so that $f(z)$ gives an estimate of x.

Estimators that maximize PDFs/probabilities.

- The maximum-likelihood (ML) estimator is given by

$$\theta = \arg\max_x p(z|x).$$

- The maximum *a posteriori* (MAP) estimator is given by

$$\theta = \arg\max_x p(x|z) = \arg\max_x p(z|x)p(x).$$

To implement such estimators, various optimization techniques may be required:

- Steepest descent
- Newton
- Gauss-Newton
- Levenberg-Marquadt
- MILP
- Viterbi algorithm
- Genetic algorithms
- Lagrangian Relaxation
- RANSAC

Estimators are random variables

Example 1: ML estimator of Rayleigh distribution parameter.

Let $\mathbf{z} = [z_1, \ldots, z_M]^\mathsf{T}$ consist of IID samples from a Rayleigh distribution:

$$p(\mathbf{z} \,|\, \eta) = \prod_{i=1}^{M} \frac{z_i}{\eta} \exp\left(-\frac{z_i^2}{2\eta}\right).$$

By differentiating the logarithm of $p(\mathbf{z} \,|\, \eta)$ and equating the derivative to zero, we get the ML estimator

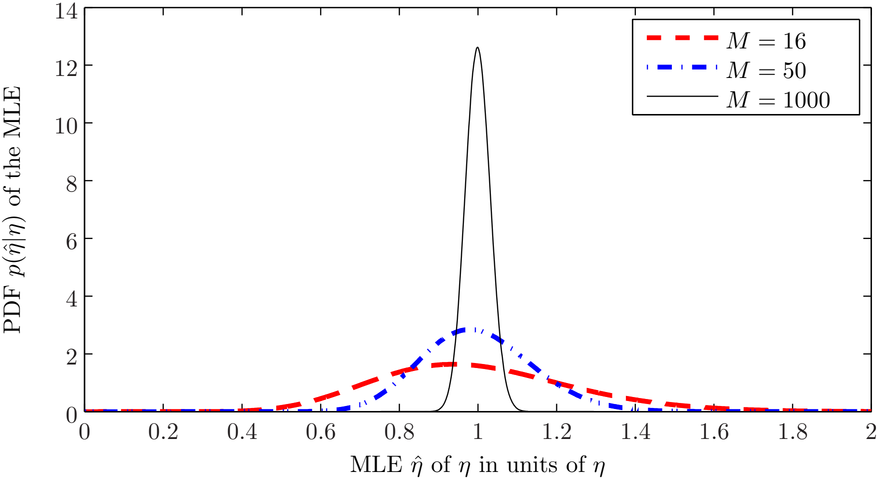
$$\hat{\eta} = \frac{1}{2M} \sum_{i=1}^{M} z_i^2.$$

This entity (a sum of IID exponential random variables) has a Gamma distribution:[a]

$$p(\hat{\eta} \,|\, \eta) = \mathrm{Gamma}\left(\hat{\eta}\,;\, M, \frac{2\eta}{M}\right) = \frac{M^M}{\Gamma(M)} \left(\frac{\hat{\eta}}{\eta}\right)^M \exp\left(-M\frac{\hat{\eta}}{\eta}\right). \tag{2}$$

[a] See Papoulis & Pillai (2002): “Probability, Random Variables and Stochastic Processes” for more on Gamma, Rayleigh, Exponential and all kinds of distributions.

Estimators are random variables



LS and MMSE estimators

The least squares (LS) estimator.

For any estimation problem on the form $\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{w}$ the least squares estimator is

$$\hat{\mathbf{x}}_{\text{LS}} = \arg\min_{\mathbf{x}} \|\mathbf{z} - \mathbf{h}(\mathbf{x})\|_2^2$$

- The LS estimator does not make any assumptions about the measurement “noise” $\mathbf{w}$. It is therefore of a non-probabilistic nature.
- If $\mathbf{w}$ is IID multivariate Gaussian, then the LS estimator is identical to the MLE.

The minimum mean square error (MMSE) estimator.

This is the probabilistic counterpart of the LS estimator. It is given by

$$\hat{\mathbf{x}}_{\text{MMSE}} = \arg\min_{\hat{\mathbf{x}}} E\left[(\hat{\mathbf{x}} - \mathbf{x})^{\mathsf{T}}(\hat{\mathbf{x}} - \mathbf{x}) \mid \mathbf{z}\right] = E[\mathbf{x}|\mathbf{z}] = \int \mathbf{x}p(\mathbf{x}|\mathbf{z})\mathrm{d}\mathbf{x}$$

MMSE estimator versus MAP estimator.

- MMSE and MAP are equal for all symmetric posterior PDFs.
- The MMSE estimator is a Bayes estimator which minimizes expected Bayes risk.
- Care should be exercised in choosing estimator if $p(\mathbf{x}|\mathbf{z})$ is multimodal or skewed.

Unbiasedness, MSE

Definition of unbiased estimator.

Let $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ be the estimation error. We then require that $E[\tilde{\mathbf{x}}] = 0$.

- The MMSE estimator is always unbiased.
- The ML and MAP estimators are in general biased.

Unbiasedness is desirable. However, there may exist biased estimators with lower MSE, and there exist estimation problems where the requirement of unbiasedness will lead to unacceptable degradation in MSE.

Variance and MSE: Vector case.

The variance is given by

$$\mathrm{Cov}(\hat{\mathbf{x}}) = E\left[(\hat{\mathbf{x}} - E[\hat{\mathbf{x}}])(\hat{\mathbf{x}} - E[\hat{\mathbf{x}}])^{\mathsf{T}}\right]$$

The MSE is given by

$$\mathrm{MSE}(\hat{\mathbf{x}}) = E\left[(\hat{\mathbf{x}} - \mathbf{x})^{\mathsf{T}}(\hat{\mathbf{x}} - \mathbf{x})\right] = \mathrm{tr}\left(E[(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^{\mathsf{T}}]\right)$$

LMMSE estimators

When the MMSE is too complicated we may settle for the best linear estimator. This entails finding the best estimator $\hat{\mathbf{x}}$ on the form

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{z} + \mathbf{b}$$

that mimimizes

$$\mathrm{MSE}(\hat{\mathbf{x}}) = E\left(\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2\right).$$

This minimization problem has the solution

$$\hat{\mathbf{x}} = E[\mathbf{x}] + \mathrm{Cov}(\mathbf{x}, \mathbf{z})\mathrm{Cov}(\mathbf{z})^{-1}(\mathbf{z} - E[\mathbf{z}]). \tag{3}$$

To use the LMMSE estimator we need to know the first two moments (expectation and covariance) of the joint PDF $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{z}, \mathbf{x})p(\mathbf{x})$.

The MSE of the LMMSE estimator.

Let $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ be the estimation error. The MSE is then given by the matrix

$$E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^{\mathsf{T}}] = \mathrm{Cov}(\mathbf{x}) - \mathrm{Cov}(\mathbf{x}, \mathbf{z})\mathrm{Cov}(\mathbf{z})^{-1}\mathrm{Cov}(\mathbf{x}, \mathbf{z})^{\mathsf{T}} \tag{4}$$

- Thus, the LMMSE estimator provides a simple measure of its own performance.
- You recognize (3) and (4) as the update step of the Kalman filter.
- (4) may be misleading if we have wrong values of $\mathrm{Cov}(\mathbf{x})$, $\mathrm{Cov}(\mathbf{x}, \mathbf{z})$ or $\mathrm{Cov}(\mathbf{z})$.

Example of LMMSE estimation[1]

Sensitivity to incorrect prior variance

The random variable x with prior mean $\bar{x}$ and variance σ_0^2 is measured via $z = x + w$ where w is zero mean, with variance σ^2 and independent of x.

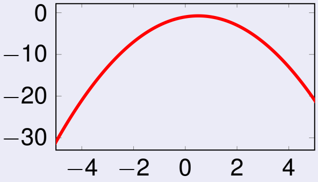
1) Write the LMMSE estimator $\hat{x}$ in terms of z and the MSE σ_1^2 associated with this estimator.
2) Write the estimate x^* of x as above but under the **incorrect** assumption that the prior variance is σ_p^2.
3) Find the actual MSE, σ_a^2 associated with 2).
4) Verify the expression for σ_a^2 by inserting $\sigma_p^2 = \sigma_0^2$ and compare with σ_1^2.
5) Let $\sigma_p^2 = s\sigma_0^2$. Investigate how x^* behaves in the limits $s \to 0$ and $s \to \infty$.

[1] Adapted from Exercise 3.2 in Bar-Shalom, Kirubarajan & Li (2001)

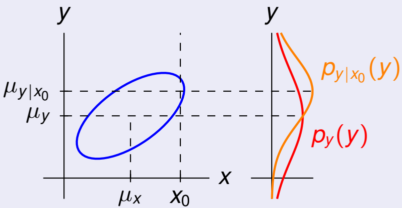
Next week

Next week's topic is the multivariate Gaussian distribution.

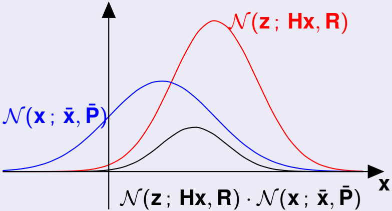
The importance of quadratic forms



Convenient manipulation rules



The product identity



Product identity $\longrightarrow$ Kalman filter