

Aerial Robotic Autonomy: Methods and Systems

Primer on Probability Theory

Kostas Alexis

Autonomous Robots Lab
Norwegian University of Science and Technology

Table of Contents

- 1 Introduction
- 2 Probability Density Functions
- 3 Gaussian Probability Density Functions
- 4 Gaussian Processes

Table of Contents

- 1 Introduction
- 2 Probability Density Functions
- 3 Gaussian Probability Density Functions
- 4 Gaussian Processes

Table of Contents

- 1 Introduction
- 2 Probability Density Functions**
- 3 Gaussian Probability Density Functions
- 4 Gaussian Processes

Definitions

Let x be a random variable distributed over a Probability Density Function (PDF) $p(x)$ over the interval $[a, b]$. Then we write:

$$\int_a^b p(x) dx = 1 \quad (1)$$

Thus satisfying the axiom of total probability.

For any two c, d within $[a, b]$ the probability that x lies within c and d , $Pr(c \leq x \leq d)$ takes the form:

$$Pr(c \leq x \leq d) = \int_c^d p(x) dx \quad (2)$$

Definitions

To introduce a conditional variable, let $p(x|y)$ be a PDF over $x \in [a, b]$ conditioned over $y \in [r, s]$ such that:

$$(\forall y) \int_a^b p(x|y) dx = 1 \quad (3)$$

For the case of N -dimensional continuous variables let $\mathbf{x} = (x_1, x_2, \dots, x_N)$ with $x_i \in [a_i, b_i]$. We then denote $p(\mathbf{x})$ or $p(x_1, x_2, \dots, x_N)$. The axiom of total probability requires:

$$\int_{\mathbf{a}}^{\mathbf{b}} p(\mathbf{x}) d\mathbf{x} = \int_{a_N}^{b_N} \dots \int_{a_2}^{b_2} \int_{a_1}^{b_1} p(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N = 1 \quad (4)$$

where $\mathbf{a} = (a_1, a_2, \dots, a_N)$ and $\mathbf{b} = (b_1, b_2, \dots, b_N)$.

Bayes' Rule and Inference

The Bayes' rule

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \quad (5)$$

Allows us to infer the posterior or likelihood of the state given the measurements, $p(\mathbf{x}|\mathbf{y})$, if we have a prior PDF over the state $p(\mathbf{x})$ and the sensor model $p(\mathbf{y}|\mathbf{x})$.

The denominator is computed by marginalization

$$p(\mathbf{y}) = p(\mathbf{y}) \int p(\mathbf{x}|\mathbf{y})d\mathbf{x} = \int p(\mathbf{x}|\mathbf{y})p(\mathbf{y})d\mathbf{x} = \int p(\mathbf{x}, \mathbf{y})d\mathbf{x} = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x} \quad (6)$$

We thus write:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}} \quad (7)$$

Moments

Moments of PDFs represent most notable characteristics.

The *zeroth* probability moment is always 1 (axiom of total probability).

The *first* probability moment is known as the *mean* μ

$$\mu = E[\mathbf{x}] = \int \mathbf{x}p(\mathbf{x})d\mathbf{x} \quad (8)$$

where $E[\cdot]$ denotes the expectation operator. For a general matrix function $\mathbf{F}(\mathbf{x})$ we write

$$E[\mathbf{F}(\mathbf{x})] = \int \mathbf{F}(\mathbf{x})p(\mathbf{x})d\mathbf{x} \quad (9)$$

The *second* probability moment is called the *covariance matrix* Σ

$$\Sigma = E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T] \quad (10)$$

The next two moments are called *skewness* and *kurtosis*.

- Skewness is a measure of symmetry, or more precisely, the lack of symmetry.
- Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution.

Sample Mean and Covariance

Let \mathbf{x} be a random variable with PDF $p(\mathbf{x})$. We can draw samples from this density

$$\mathbf{x}_{meas} \leftarrow p(\mathbf{x}) \quad (11)$$

Taking N such samples allow us to derive the *sample mean* and *sample covariance*

$$\boldsymbol{\mu}_{meas} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_{i,meas}$$

$$\boldsymbol{\Sigma}_{meas} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_{i,meas} - \boldsymbol{\mu}_{meas})(\mathbf{x}_{i,meas} - \boldsymbol{\mu}_{meas})^T \quad (12)$$

- Normalization term $N - 1$, rather than N , is called the *Bessel's correction* and reflects the fact that the sample covariance uses the difference of the measurements against the sample mean.

Statistically Independent, Uncorrelated

Let \mathbf{x}, \mathbf{y} be two random variables.

We say that \mathbf{x}, \mathbf{y} are statistically independent if

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y}) \quad (13)$$

We say that \mathbf{x}, \mathbf{y} are uncorrelated if

$$E[\mathbf{x}\mathbf{y}^T] = E[\mathbf{x}]E[\mathbf{y}]^T \quad (14)$$

- If the variables are statistically independent *this implies* that they are also uncorrelated.
- The reverse *is not always* true.

Normalized Product

If $p_1(\mathbf{x}), p_2(\mathbf{x})$ are two PDFs of \mathbf{x} , the normalized product $p(\mathbf{x})$ is formulated as

$$p(\mathbf{x}) = \eta p_1(\mathbf{x}) p_2(\mathbf{x}), \quad \eta = \left(\int p_1(\mathbf{x}) p_2(\mathbf{x}) d\mathbf{x} \right)^{-1} \quad (15)$$

with η normalization constant ensuring that $p(\mathbf{x})$ satisfies the axiom of total probability.

The normalized product can be used to fuse independent estimates of a variable under the assumption of a uniform prior

$$p(\mathbf{x}|\mathbf{y}_1, \mathbf{y}_2) = \eta p(\mathbf{x}|\mathbf{y}_1) p(\mathbf{x}|\mathbf{y}_2), \quad \eta = \frac{p(\mathbf{y}_1) p(\mathbf{y}_2)}{p(\mathbf{y}_1, \mathbf{y}_2) p(\mathbf{x})} \quad (16)$$

Note: If we let the prior $p(\mathbf{x})$ be uniform for all values of \mathbf{x} then η is also a constant.

Shannon and Mutal Information

To assess our confidence for the estimate of the mean of a PDF we can use the *negative entropy* or *Shannon information* H

$$H(\mathbf{x}) = -E[\ln p(\mathbf{x})] = - \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} \quad (17)$$

To measure how much knowing one of the variables, reduces the uncertainty for the other we can use the *mutual information* $I(\mathbf{x}, \mathbf{y})$ between two random variables \mathbf{x}, \mathbf{y}

$$I(\mathbf{x}, \mathbf{y}) = E \left[\ln \left(\frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \right) \right] = \iint p(\mathbf{x}, \mathbf{y}) \ln \left(\frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \right) d\mathbf{x}d\mathbf{y} \quad (18)$$

- When \mathbf{x}, \mathbf{y} are statistically independent, $I(\mathbf{x}, \mathbf{y}) = 0$.
- When \mathbf{x}, \mathbf{y} are statistically dependent, $I(\mathbf{x}, \mathbf{y}) \geq 0$ and $I(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}) + H(\mathbf{y}) - H(\mathbf{x}, \mathbf{y})$

Cramér-Rao Lower Bound and Fisher Information

Considering a deterministic parameter θ that influences a random variable \mathbf{x} , $p(\mathbf{x}|\theta)$ and a sample $\mathbf{x}_{meas} \leftarrow p(\mathbf{x}|\theta)$, then the *Cramér-Rao lower bound (CRLB)* says that the covariance of any *unbiased estimate* $\hat{\theta}$ (based on \mathbf{x}_{meas}) of the deterministic parameter θ is bounded by the *Fisher Information Matrix* $\mathbf{I}(\mathbf{x}|\theta)$

$$\text{cov}(\hat{\theta}|\mathbf{x}_{meas}) = E \left[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T \right] \geq \mathbf{I}^{-1}(\mathbf{x}|\theta) \quad (19)$$

The Fisher Information Matrix takes the form

$$\mathbf{I}(\mathbf{x}|\theta) = E \left[\left(\frac{\partial \ln p(\mathbf{x}|\theta)}{\partial \theta} \right)^T \left(\frac{\partial \ln p(\mathbf{x}|\theta)}{\partial \theta} \right) \right] \quad (20)$$

Thus, CRLB sets a fundamental limit on how certain we can be about an estimate of a parameter, given our measurements.

Table of Contents

- 1 Introduction
- 2 Probability Density Functions
- 3 Gaussian Probability Density Functions**
- 4 Gaussian Processes

Definitions

Throughout most of our work in state estimation for robotics we will be working with Gaussian PDFs.

One-dimensional Gaussian PDF over random variable $x \in \mathbb{R}$

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) \quad (21)$$

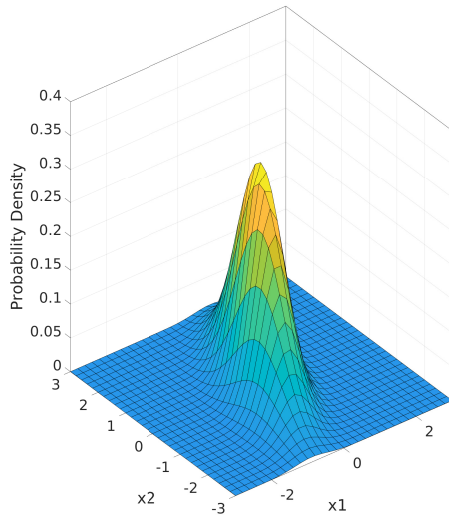
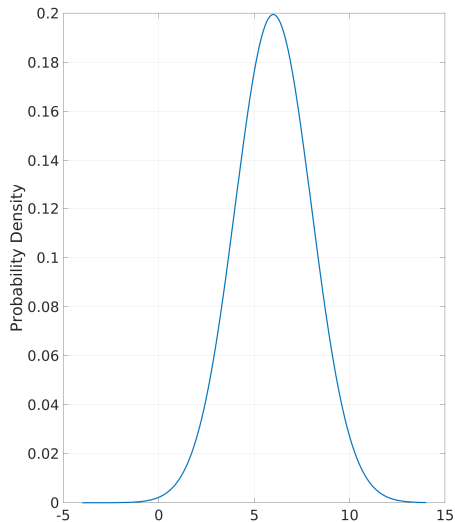
where μ, σ^2 are the mean and variance respectively (σ is called the standard deviation).

Multi-variate Gaussian PDF over random variable $\mathbf{x} \in \mathbb{R}^N$

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (22)$$

where $\boldsymbol{\mu} \in \mathbb{R}^N, \boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ are the mean and covariance matrix respectively (symmetric, positive-definite).

Definitions



Definitions

Accordingly,

$$\boldsymbol{\mu} = E[\mathbf{x}] = \int_{-\infty}^{\infty} \mathbf{x} \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Sigma}}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \quad (23)$$

$$\boldsymbol{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] = \int_{-\infty}^{\infty} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{\Sigma}}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \quad (24)$$

We also write that \mathbf{x} is *normally* distributed

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (25)$$

We also say that a random variable is *standard normally distributed* if

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{1}) \quad (26)$$

where $\mathbf{1}$ is an $N \times N$ identity matrix.

Isserli's Theorem

Isserli's theorem allows to compute certain **higher-order moments of a Gaussian random variable** $\mathbf{x} = (x_1, x_2, \dots, x_{2M}) \in \mathbb{R}^{2M}$. In general, it says that

$$E[x_1 x_2 x_3 \dots x_{2M}] = \sum \prod E[x_i x_j], \quad (27)$$

where this implies summing over all distinct ways of partitioning into a product of M pairs. There are $(2M)!/(2^M M!)$ terms in the sum.

With four variables we write

$$E[x_i x_j x_k x_\ell] = E[x_i x_j] E[x_k x_\ell] + E[x_i x_k] E[x_j x_\ell] + E[x_i x_\ell] E[x_j x_k] \quad (28)$$

Isserli's Theorem

Assuming $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \in \mathbb{R}^N$ we will have occasion to compute expressions of the form

$$E[\mathbf{x}(\mathbf{x}^T \mathbf{x})^p \mathbf{x}^T] \quad (29)$$

where p non-negative integer.

- For $p = 0$, it holds that $E[\mathbf{x}\mathbf{x}^T] = \mathbf{\Sigma}$
- For $p = 1$, it holds that $E[\mathbf{x}\mathbf{x}^T \mathbf{x}\mathbf{x}^T] = \mathbf{\Sigma}(\text{tr}(\mathbf{\Sigma})\mathbf{1} + 2\mathbf{\Sigma})$

Note that for the scalar case, we have $x \sim \mathcal{N}(0, \sigma^2)$ and thus $E[x^4] = \sigma^2(\sigma^2 + 2\sigma^2) = 3\sigma^4$.

Isserli's Theorem

Let us also consider the case where

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0} \quad \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right), \quad \dim(\mathbf{x}_1) = N_1, \quad \dim(\mathbf{x}_2) = N_2 \quad (30)$$

Throughout our work in state estimation, we will need to compute expressions of the form

$$E[\mathbf{x}(\mathbf{x}_1^T \mathbf{x}_1)^p \mathbf{x}^T], \quad p \in \mathbb{N}_0 \quad (31)$$

For $p = 0$, $E[\mathbf{x}\mathbf{x}^T] = \boldsymbol{\Sigma}$

For $p = 1$ it holds that

$$E[\mathbf{x}\mathbf{x}_1^T \mathbf{x}_1 \mathbf{x}^T] = \boldsymbol{\Sigma} \left(\text{tr}(\boldsymbol{\Sigma}_{11})\mathbf{1} + 2 \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right) \quad (32)$$

Similarly,

$$E[\mathbf{x}\mathbf{x}_2^T \mathbf{x}_2 \mathbf{x}^T] = \boldsymbol{\Sigma} \left(\text{tr}(\boldsymbol{\Sigma}_{22})\mathbf{1} + 2 \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \quad (33)$$

Accordingly, we have the final check

$$E[\mathbf{x}\mathbf{x}^T\mathbf{x}\mathbf{x}^T] = E[\mathbf{x}(\mathbf{x}_1^T\mathbf{x}_1 + \mathbf{x}_2^T\mathbf{x}_2)\mathbf{x}^T] = E[\mathbf{x}\mathbf{x}_1^T\mathbf{x}_1\mathbf{x}^T] + E[\mathbf{x}\mathbf{x}_2^T\mathbf{x}_2\mathbf{x}^T] \quad (34)$$

Furthermore, it holds that

$$E[\mathbf{x}\mathbf{x}^T\mathbf{A}\mathbf{x}\mathbf{x}^T] = \mathbf{\Sigma}(\text{tr}(\mathbf{A}\mathbf{\Sigma})\mathbf{1} + \mathbf{A}\mathbf{\Sigma} + \mathbf{A}^T\mathbf{\Sigma}) \quad (35)$$

where \mathbf{A} is a compatible square matrix.

Joint Gaussian PDFs, their Factors, and Inference

We can have a joint Gaussian over a pair of variables (\mathbf{x}, \mathbf{y})

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} \right) \quad (36)$$

with $\boldsymbol{\Sigma}_{yx} = \boldsymbol{\Sigma}_{xy}^T$.

Schur complement

Suppose $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are respectively $p \times p$, $p \times q$, $q \times p$, $q \times q$, $p, q \in \mathbb{N}_0$ matrices of complex numbers. Let matrix \mathbf{M} with dimensions $(p + q) \times (p + q)$

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad (37)$$

the Schur complements of \mathbf{A} in \mathbf{M} are the matrices of the form $\mathbf{S} = \mathbf{D} - \mathbf{CaB}$, where \mathbf{a} is the generalized inverse of \mathbf{A} . If \mathbf{A} is invertible, this is $\mathbf{S} = \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}$

Joint Gaussian PDFs, their Factors, and Inference

Breaking a joint density into the product of two factors $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y})$ we can work out the details for the joint Gaussian using the Schur complement. It turns out that:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) \quad (38)$$

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}) \quad (39)$$

$$p(\mathbf{y}) = (\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}) \quad (40)$$

with both $p(\mathbf{x}|\mathbf{y}), p(\mathbf{y})$ are Gaussian PDFs.

If we know the value of \mathbf{y} (i.e., it is measured), then we can work out the likelihood of \mathbf{x} by computing $p(\mathbf{x}|\mathbf{y})$ using Eq. 39. This is in fact the *cornerstone of Gaussian Inference*: We start with a prior about our state, $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$ and then narrow down based on some measurements \mathbf{y}_{meas} . In Eq. 39 we see that both the mean $\boldsymbol{\mu}_x$ and the covariance $\boldsymbol{\Sigma}_{xx}$ are adjusted.

Statistically Independent, Uncorrelated

For Gaussian PDFs, statistically independent variables are also uncorrelated **and** uncorrelated variables are also statistically independent.

Assuming statistical independence, $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$ and so $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$. This implies

$$\boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) = \mathbf{0} \quad (41)$$

$$\boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx} = \mathbf{0} \quad (42)$$

thus, also $\boldsymbol{\Sigma}_{xy} = \mathbf{0}$. Furthermore, since

$$\boldsymbol{\Sigma}_{xy} = E[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{y} - \boldsymbol{\mu}_y)^T] = E[\mathbf{x}\mathbf{y}^T] - E[\mathbf{x}]E[\mathbf{y}]^T \quad (43)$$

we conclude the uncorrelated condition

$$E[\mathbf{x}\mathbf{y}^T] = E[\mathbf{x}]E[\mathbf{y}]^T \quad (44)$$

Linear Change of Variables

Suppose we have a Gaussian random variable $\mathbf{x} \in \mathbb{R}^N \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$ and that we have a second random variable $\mathbf{y} \in \mathbb{R}^M$ related to \mathbf{x} through the linear map

$$\mathbf{y} = \mathbf{G}\mathbf{x}, \quad \mathbf{G} \in \mathbb{R}^{M \times N} \text{ (const)} \quad (45)$$

With respect to the statistical properties of \mathbf{y} it holds that

$$\boldsymbol{\mu}_y = E[\mathbf{y}] = E[\mathbf{G}\mathbf{x}] = \mathbf{G}E[\mathbf{x}] = \mathbf{G}\boldsymbol{\mu}_x \quad (46)$$

$$\boldsymbol{\Sigma}_{yy} = E[(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)^T] = \mathbf{G}E[(\mathbf{x} - \boldsymbol{\mu}_x)(\mathbf{x} - \boldsymbol{\mu}_x)^T]\mathbf{G}^T = \mathbf{G}\boldsymbol{\Sigma}_{xx}\mathbf{G}^T \quad (47)$$

and thus we write that $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}) = \mathcal{N}(\mathbf{G}\boldsymbol{\mu}_x, \mathbf{G}\boldsymbol{\Sigma}_{xx}\mathbf{G}^T)$.

Linear Change of Variables

An alternative way to conclude the same is through a change of variables. We assume that the linear map is *injective* (that is that two values of \mathbf{x} cannot map to a single \mathbf{y} value) and in fact \mathbf{G} is *invertible* (thus also $M = N$). By the axiom of total probability

$$\int_{-\infty}^{\infty} p(\mathbf{x}) d\mathbf{x} = 1 \quad (48)$$

A small volume of \mathbf{x} is related to a small volume \mathbf{y} by $d\mathbf{y} = |\det \mathbf{G}| d\mathbf{x}$. We can then make a substitution of variables to have

$$1 = \int_{-\infty}^{\infty} p(\mathbf{x}) d\mathbf{x} = \dots = \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{G}\mathbf{\Sigma}_{xx}\mathbf{G}^T)}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{G}\boldsymbol{\mu}_x)^T (\mathbf{G}\mathbf{\Sigma}_{xx}\mathbf{G}^T)^{-1} (\mathbf{y} - \mathbf{G}\boldsymbol{\mu}_x)\right) d\mathbf{y} \quad (49)$$

where we have $\boldsymbol{\mu}_y = \mathbf{G}\boldsymbol{\mu}_x$, $\mathbf{\Sigma}_{yy} = \mathbf{G}\mathbf{\Sigma}_{xx}\mathbf{G}^T$ as derived before.

Linear Change of Variables

Likewise, we can derive the statistics of \mathbf{x} from \mathbf{y} given that this linear mapping is invertible. This however is a bit trickier as the resulting covariance of \mathbf{x} will blow up since we are dilating to a larger space. To overcome this problem, we switch to *information form*. Let

$$\mathbf{u} = \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{y} \quad (50)$$

we have

$$\mathbf{u} \sim \mathcal{N}(\boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}^{-1}) \quad (51)$$

Likewise, let

$$\mathbf{v} = \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{x} \quad (52)$$

we have

$$\mathbf{v} \sim \mathcal{N}(\boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}^{-1}) \quad (53)$$

Linear Change of Variables

Since the mapping from \mathbf{y} to \mathbf{x} is not unique, we need to specify what we want to do. One choice is

$$\mathbf{v} = \mathbf{G}^T \mathbf{u} \Leftrightarrow \boldsymbol{\Sigma}_{xx}^{-1} \mathbf{x} = \mathbf{G}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{y} \quad (54)$$

then we take the expectations

$$\boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\mu}_x = E[\mathbf{v}] = \mathbf{G}^T E[\mathbf{u}] = \mathbf{G}^T \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\mu}_y \quad (55)$$

$$\boldsymbol{\Sigma}_{xx}^{-1} = E[(\mathbf{v} - \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\mu}_x)(-\boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\mu}_x)^T] = \mathbf{G}^T \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{G} \quad (56)$$

Note: if $\boldsymbol{\Sigma}_{xx}^{-1}$ is not full rank then we cannot recover $\boldsymbol{\Sigma}_{xx}, \boldsymbol{\mu}_x$ and must keep them in information form. However, multiple such estimates can be fused together.

Normalized Product of Gaussians

The normalized product of K Gaussian PDFs is *also a Gaussian PDF*

$$\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \equiv \eta \prod_{k=1}^K \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right) \quad (57)$$

where $\boldsymbol{\Sigma}^{-1} = \sum_{k=1}^K \boldsymbol{\Sigma}_k^{-1}$, $\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \sum_{k=1}^K \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k$, and η is a normalization constant to enforce the axiom of total probability.

- The normalized product of Gaussians comes up when fusing multiple estimates together.

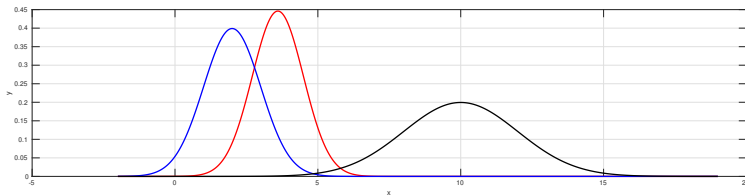


Figure: Illustrational example

Normalized Product of Gaussians

We also have that

$$\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \equiv \eta \prod_{k=1}^K \exp\left(-\frac{1}{2}(\mathbf{G}_k \mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{G}_k \mathbf{x} - \boldsymbol{\mu}_k)\right) \quad (58)$$

where

$$\boldsymbol{\Sigma}^{-1} = \sum_{k=1}^K \mathbf{G}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{G}_k \quad (59)$$

$$\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \sum_{k=1}^K \mathbf{G}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k \quad (60)$$

in the case that the matrices, $\mathbf{G}_k \in \mathbb{R}^{M_k \times N}$, are present, with $M_k \leq N$ and η is again a normalization constant.

Sherman-Morrison-Woodburry (SMW) Identity

In state estimation, the Sherman-Morrison-Woodburry (SMW) *matrix identity(ies)* (also called the *matrix inversion lemma*) is commonly used. SMW is in fact four identities coming from the same derivation.

A matrix can be factored into either a *lower-diagonal-upper* (LDU) or *upper-diagonal-lower* (UDL) form:

$$\begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{CA} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} + \mathbf{CAB} \end{bmatrix} \begin{bmatrix} \mathbf{1} & -\mathbf{AB} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \quad (\text{LDU}) \quad (61)$$

$$\begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & -\mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{BD}^{-1}\mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{1} \end{bmatrix} \quad (\text{UDL}) \quad (62)$$

Sherman-Morrison-Woodburry (SMW) Identity

We then invert each of these forms. For the LDU

$$\begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{1} & \mathbf{AB} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & (\mathbf{D} + \mathbf{CAB})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ -\mathbf{CA} & \mathbf{1} \end{bmatrix} \Rightarrow \quad (63)$$

$$\begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A} - \mathbf{AB}(\mathbf{D} + \mathbf{CAB})^{-1}\mathbf{CA} & \mathbf{AB}(\mathbf{D} + \mathbf{CAB})^{-1} \\ -(\mathbf{D} + \mathbf{CAB})^{-1}\mathbf{CA} & (\mathbf{D} + \mathbf{CAB})^{-1} \end{bmatrix} \quad (64)$$

For the UDL case

$$\begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A}^{-1} + \mathbf{BD}^{-1}\mathbf{C})^{-1} & (\mathbf{A}^{-1} + \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A}^{-1} + \mathbf{BD}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{A}^{-1} + \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} \end{bmatrix} \quad (65)$$

Sherman-Morrison-Woodburry (SMW) Identity

Comparing from Eq. 63 and 64 we conclude the SMW identities:

SMW Identities

$$(\mathbf{A}^{-1} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} \equiv \mathbf{A} - \mathbf{A}\mathbf{B}(\mathbf{D} + \mathbf{C}\mathbf{A}\mathbf{B})^{-1}\mathbf{C}\mathbf{A} \quad (66)$$

$$(\mathbf{D} + \mathbf{C}\mathbf{A}\mathbf{B})^{-1} \equiv \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{C}(\mathbf{A}^{-1} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \quad (67)$$

$$\mathbf{A}\mathbf{B}(\mathbf{D} + \mathbf{C}\mathbf{A}\mathbf{B})^{-1} \equiv (\mathbf{A}^{-1} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \quad (68)$$

$$(\mathbf{D} + \mathbf{C}\mathbf{A}\mathbf{B})^{-1}\mathbf{C}\mathbf{A} \equiv \mathbf{D}^{-1}\mathbf{C}(\mathbf{A}^{-1} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} \quad (69)$$

The SMW identities are frequently used when manipulating expressions involving covariance matrices of Gaussian PDFs.

Passing a Gaussian through a Nonlinearity

We can examine the process of passing a Gaussian PDF through a *stochastic nonlinearity*, namely computing

$$p(\mathbf{y}) = \int_{-\infty}^{\infty} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}, \quad p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{g}(\mathbf{x}), \mathbf{R}), \quad p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx}) \quad (70)$$

and $\mathbf{g}(\cdot)$, $:\mathbf{x} \rightarrow \mathbf{y}$ is a nonlinear map that is then corrupted by zero-mean Gaussian noise with covariance \mathbf{R} .

- We shall require this type of stochastic nonlinearity when modeling sensors.
- Passing a Gaussian through this type of function is required when performing full Bayesian inference.

Passing a Gaussian through a Nonlinearity

Scalar Deterministic Case via Change of Variables

Let a Gaussian random variable $x \in \mathbb{R}$, $x \sim \mathcal{N}(0, \sigma^2)$, i.e., $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right)$.

Consider the nonlinear mapping

$$y = \exp(x) \rightarrow x = \ln(y) \quad (71)$$

The infinitesimal integration volumes for x, y are then related by

$$dy = \exp(x) dx \text{ or } dx = \frac{1}{y} dy \quad (72)$$

According to the axiom of total probability

$$1 = \int_{-\infty}^{\infty} p(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right) dx \quad (73)$$

Passing a Gaussian through a Nonlinearity

$$= \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(\ln(y))^2}{\sigma^2}\right) \frac{1}{y} dy = \int_0^{\infty} p(y) dy \quad (74)$$

giving the exact expression for $p(y)$ which is visualized in the Figure below.

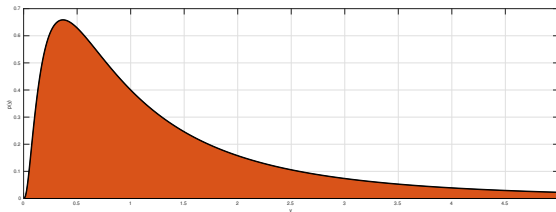


Figure: The PDF, $p(y)$, resulting from $p(x)$ being Gaussian PDF and passing through the nonlinearity $y = \exp(x)$.

Note that $p(y)$ is no longer Gaussian owing to the nonlinear change of variables.

Passing a Gaussian through a Nonlinearity

General Case via Linearization: Such analytical calculation -as before - is not generally possible. Moreover, when the nonlinearity is stochastic (i.e., $\mathbf{R} > 0$), our mapping will never be invertible due to the extra input coming from the noise. One way to solve for processing the nonlinear effect is through *linearization*. We linearize the nonlinear map such that

$$\begin{aligned}\mathbf{g}(\mathbf{x}) &\approx \boldsymbol{\mu}_y + \mathbf{G}(\mathbf{x} - \boldsymbol{\mu}_x) \\ \mathbf{G} &= \left. \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\boldsymbol{\mu}_x} \\ \boldsymbol{\mu}_y &= \mathbf{g}(\boldsymbol{\mu}_x)\end{aligned}\tag{75}$$

where \mathbf{G} is the *Jacobian* of $\mathbf{g}(\cdot)$ with respect to \mathbf{x} . This process allows us to pass the Gaussian through the linearized function in closed form and it is thus an approximation that works well for mildly nonlinear maps.

Passing a Gaussian through a Nonlinearity

Returning to

$$p(\mathbf{y}) = \int_{-\infty}^{\infty} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x} \quad (76)$$

we have that

$$p(\mathbf{y}) = \eta \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(\mathbf{y} - (\boldsymbol{\mu}_y + \mathbf{G}(\mathbf{x} - \boldsymbol{\mu}_x)))^T \mathbf{R}^{-1}(\mathbf{y} - (\boldsymbol{\mu}_y + \mathbf{G}(\mathbf{x} - \boldsymbol{\mu}_x)))\right) \\ \times \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^T \boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x)\right) d\mathbf{x} \quad (77)$$

$$= \eta \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_y)^T \mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_x)^T (\boldsymbol{\Sigma}_{xx}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G})(\mathbf{x} - \boldsymbol{\mu}_x)\right) \\ \times \exp\left((\mathbf{y} - \boldsymbol{\mu}_y)^T \mathbf{R}^{-1} \mathbf{G}(\mathbf{x} - \boldsymbol{\mu}_x)\right) d\mathbf{x} \quad (78)$$

Passing a Gaussian through a Nonlinearity

After manipulation

$$p(\mathbf{y}) = \rho \exp \left(-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_y)^T (\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{G} (\mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} + \boldsymbol{\Sigma}_{xx}^{-1})^{-1} \mathbf{G}^T \mathbf{R}^{-1}) (\mathbf{y} - \boldsymbol{\mu}_y) \right) \quad (79)$$

By exploiting the SMW inequalities it turns out

$$p(\mathbf{y}) = \rho \left(-\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_y)^T (\mathbf{R} + \mathbf{G} \boldsymbol{\Sigma}_{xx} \mathbf{G}^T)^{-1} (\mathbf{y} - \boldsymbol{\mu}_y) \right) \quad (80)$$

where ρ is the new normalization constant. Accordingly we write

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy}) = \mathcal{N}(\mathbf{g}(\boldsymbol{\mu}_x), \mathbf{R} + \mathbf{G} \boldsymbol{\Sigma}_{xx} \mathbf{G}^T) \quad (81)$$

Shannon Information of a Gaussian

Shannon information of a Gaussian PDF

$$\begin{aligned} H(\mathbf{x}) &= - \int_{-\infty}^{\infty} p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{2} \ln \left((2\pi)^N \det \mathbf{\Sigma} \right) + \frac{1}{2} E \left[(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \end{aligned} \quad (82)$$

where the second term is written as an expectation and in fact corresponds to a squared *Mahalanobis distance*.

The Mahalanobis distance is a measure of the distance between a point P and a distribution D, introduced by P. C. Mahalanobis in 1936.

Shannon Information of a Gaussian

Mahalanobis distance

For an observation \mathbf{x} from a set of observations with mean μ and covariance matrix Σ the Mahalanobis distance is defined as

$$D_M(\mathbf{x}) = \sqrt{(\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)} \quad (83)$$

The Mahalanobis distance can also be defined as the dissimilarity between two vectors \mathbf{x}, \mathbf{y} , of the same distribution with covariance Σ , as

$$d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y})} \quad (84)$$

If the covariance matrix is the identity matrix, the Mahalanobis distance reduces to the Euclidean distance. If the covariance matrix is diagonal, then the resulting distance measure is called a standardized Euclidean distance.

Shannon Information of a Gaussian

The quadratic function inside the expectation in $H(\mathbf{x})$ can be written as

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = \text{tr}(\boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T) \quad (85)$$

Accordingly it can be shown that

$$E[(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})] = \text{tr}(E[\boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T]) = N \quad (86)$$

Substituting back to into the expression for Shannon information

$$H(\mathbf{x}) = \frac{1}{2} \ln \left((2\pi e)^N \det \boldsymbol{\Sigma} \right) \quad (87)$$

which is purely a function of the covariance matrix $\boldsymbol{\Sigma}$ of the Gaussian PDF.

Geometric interpretation: $\sqrt{\det \boldsymbol{\Sigma}}$ is proportional to the *volume of the uncertainty ellipsoid* formed by the Gaussian.

Mutual Information of a Joint Gaussian PDF

Let the joint Gaussian for variables $\mathbf{x} \in \mathbb{R}^N, \mathbf{y} \in \mathbb{R}^M$

$$p(\mathbf{x}, \mathbf{y}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix}\right) \quad (88)$$

It can be shown that the *mutual information for the joint Gaussian* is

$$I(\mathbf{x}, \mathbf{y}) = -\frac{1}{2} \ln \left(\frac{\det \boldsymbol{\Sigma}}{\det \boldsymbol{\Sigma}_{xx} \det \boldsymbol{\Sigma}_{yy}} \right) \quad (89)$$

And by further processing it turns out that

$$I(\mathbf{x}, \mathbf{y}) = -\frac{1}{2} \ln \det(\mathbf{1} - \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx}) = \quad (90)$$

$$= -\frac{1}{2} \ln \det(\mathbf{1} - \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xy}) \quad (91)$$

The two forms above are equivalent based on *Sylvester's determinant theorem* which states that $\det(\mathbf{1} - \mathbf{AB}) = \det(\mathbf{1} - \mathbf{BA})$ even when \mathbf{A}, \mathbf{B} are not square.

Cramér-Rao Lower Bound for Gaussian PDFs

Let K samples (measurements) $\mathbf{x}_{meas,k} \in \mathbb{R}^N$ drawn from a Gaussian PDF. The K *statistically independent random variables* associated with these measurements are

$$(\forall) \mathbf{x}_k \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (92)$$

Due to statistical independence, it holds that $E(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_\ell - \boldsymbol{\mu})^T] = 0$, $k \neq \ell$. It can be shown that the *Fisher Information Matrix* takes the form

$$I(\mathbf{x}|\boldsymbol{\mu}) = K\boldsymbol{\Sigma}^{-1} \quad (93)$$

Thus, the CRLB says

$$\text{COV}(\hat{\mathbf{x}}|\mathbf{x}_{meas,1}, \dots, \mathbf{x}_{meas,K}) \geq \frac{1}{K}\boldsymbol{\Sigma} \quad (94)$$

i.e., the more measurements, the smaller the lower limit of the uncertainty in the estimate.

Cramér-Rao Lower Bound for Gaussian PDFs

Note that in computing the CRLB there was no need to specify the form of the unbiased estimator. *The CRLB is the lower bound for **any** unbiased estimator.* An estimator that performs exactly at the CRLB can be found

$$\hat{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_{meas,k} \quad (95)$$

where

$$E[\hat{\mathbf{x}}] = E\left[\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k\right] = \frac{1}{K} \sum_{k=1}^K E[\mathbf{x}_k] = \frac{1}{K} \sum_{k=1}^K \boldsymbol{\mu} = \boldsymbol{\mu} \quad (96)$$

$$\begin{aligned} \text{COV}(\hat{\mathbf{x}} | \mathbf{x}_{meas,1}, \dots, \mathbf{x}_{meas,K}) &= E[(\hat{\mathbf{x}} - \boldsymbol{\mu})(\hat{\mathbf{x}} - \boldsymbol{\mu})^T] = \\ &= E\left[\left(\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k - \boldsymbol{\mu}\right) \left(\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k - \boldsymbol{\mu}\right)^T\right] = \frac{1}{K} \boldsymbol{\Sigma} \end{aligned} \quad (97)$$

which is exactly at the CRLB.

Table of Contents

- 1 Introduction
- 2 Probability Density Functions
- 3 Gaussian Probability Density Functions
- 4 Gaussian Processes**

We denote a Gaussian random variable $\mathbf{x} \in \mathbb{R}^N$

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (98)$$

we use this type of random variable extensively to represent discrete-time quantities.

To represent state quantities that are continuous functions of time, t , we introduce *Gaussian Processes (GPs)*. For GPs we say that there is a *mean function*, $\boldsymbol{\mu}(t)$, and a *covariance function*, $\boldsymbol{\Sigma}(t, t')$.

- The whole trajectory is viewed as a single variable belonging to a class of functions.
- The closer a function is to the mean function, the more likely it is.
- The covariance controls how smooth the function is by describing the correlation between two times t and t' .

$$\mathbf{x}(t) \sim \mathcal{GP}(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t, t'))$$

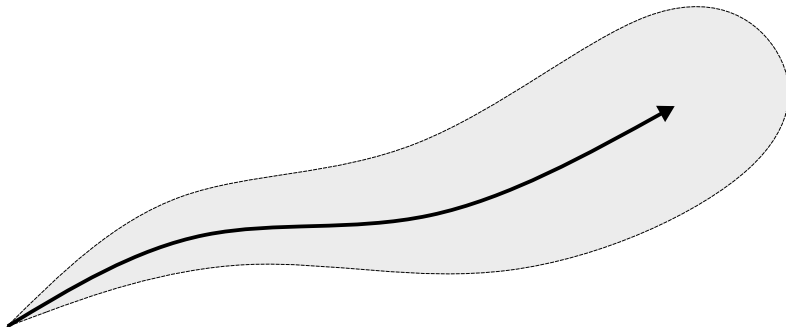


Figure: Continuous-time trajectories can be represented using Gaussian processes, which have a mean function (dark line) and a covariance function (shaded area).

Gaussian Processes

To indicate that a continuous-time trajectory is a *Gaussian Process (GP)*, we write

$$\mathbf{x}(t) \sim \mathcal{GP}(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t, t')) \quad (99)$$

If we consider a variable at a single particular time of interest, τ , we write

$$\mathbf{x}(\tau) \sim \mathcal{N}(\boldsymbol{\mu}(\tau), \boldsymbol{\Sigma}(\tau, \tau)) \quad (100)$$

where $\boldsymbol{\Sigma}(\tau, \tau)$ is a simple covariance matrix. We have marginalized out all of the other instants of time, leaving $\mathbf{x}(\tau)$ as a usual Gaussian variable.

Zero-mean, White noise process

$$\mathbf{w}(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{Q}\delta(t - t')) \quad (101)$$

where $\delta(t)$ is Dirac's delta and \mathbf{Q} is a *power spectral density*. This is a *stationary* noise process as it depends only on the difference $t - t'$.

Thank you

Q&A

Assignments

Other matters