Chapter 1

Random Variables (recall)

1.1 Basic concepts

A *random variable* $\boldsymbol{x}$ is a function $\boldsymbol{x} : \Omega \to \mathbb{R}$, where Ω denotes the space of elementary events. We associate to the random variable the probability $\mathbb{P}\{\boldsymbol{x} \leqslant x\}$, which denotes the probability that $\boldsymbol{x}$ takes values less or equal to x.

The *Cumulative Distribution Function* (CDF) is defined as:

$$F_{\boldsymbol{x}}(x) = P\{\boldsymbol{x} \leq x\}$$

Properties of the CDF:

- $F_{\boldsymbol{x}} : \mathbb{R} \to [0, 1]$
- $\lim_{x \to -\infty} F_{\boldsymbol{x}}(x) = 0 \quad , \quad \lim_{x \to +\infty} F_{\boldsymbol{x}}(x) = 1$
- $F_{\boldsymbol{x}}(x)$ is a non-decreasing function, i.e., if $x_1 < x_2$ then $F_{\boldsymbol{x}}(x_1) \leq F_{\boldsymbol{x}}(x_2)$
- $\mathbb{P}\{a < \boldsymbol{x} \leqslant b\} = F_{\boldsymbol{x}}(b) - F_{\boldsymbol{x}}(a)$

The *Probability Density Function* (PDF) (or *Probability Mass Function* when discrete random variables) is defined as

$$f_{\boldsymbol{x}}(x) = \frac{\partial}{\partial x} F_{\boldsymbol{x}}(x)$$

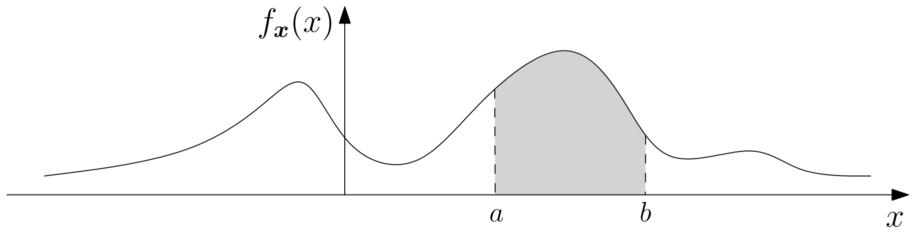
which implies

$$F_{\boldsymbol{x}}(x) = \int_{-\infty}^{x} f_{\boldsymbol{x}}(y)\, dy$$

The PDF completely characterizes a random variable.
Properties of the PDF:

- $f_{\boldsymbol{x}}(x) \geq 0 \ , \quad \forall x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f_{\boldsymbol{x}}(y)\, dy = 1$

- $\mathbb{P}\{a < \boldsymbol{x} \leqslant b\} = \int_a^b f_{\boldsymbol{x}}(y)\, dy$

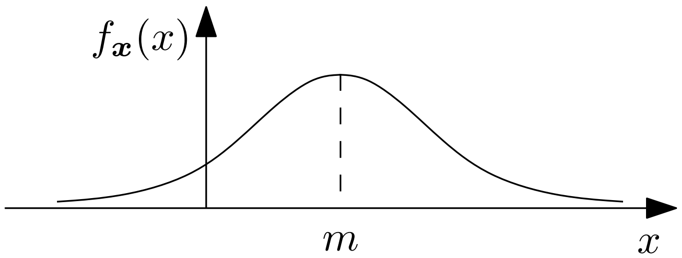


Examples:

- Gaussian (Normal) random variable

$$f_{\boldsymbol{x}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}, \quad m, \sigma \in \mathbb{R},\ \sigma > 0$$

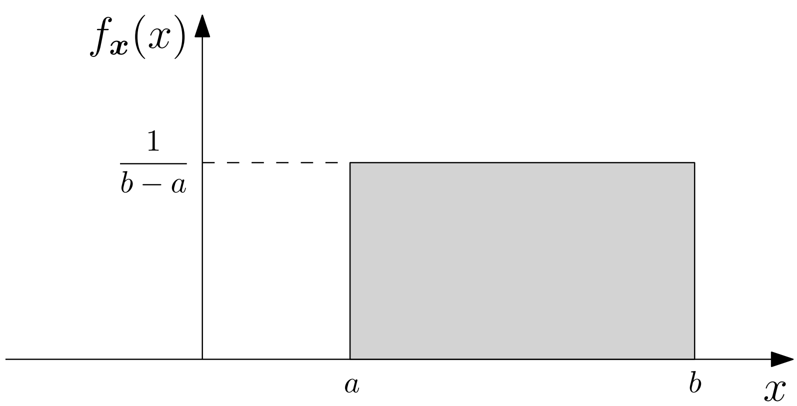
Notation $\boldsymbol{x} \sim \mathbb{N}(m, \sigma^2)$



- Uniform random variable

$$f_{\boldsymbol{x}}(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{else} \end{cases}$$

Notation $\boldsymbol{x} \sim \mathbb{U}(a, b)$



1.1.1 Multivariate Distributions

When dealing with more random variables $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, we can embed them in a vector $\boldsymbol{x}$, that is

$$\boldsymbol{x} : \Omega \to \mathbb{R}^n \ , \quad \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix}$$

The *Joint Cumulative Distribution Function (Joint CDF)* is defined as

$$F_{\boldsymbol{x}}(x) = P\left\{\boldsymbol{x}_1 \leqslant x_1, \boldsymbol{x}_2 \leqslant x_2, \dots, \boldsymbol{x}_n \leqslant x_n\right\}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- $F_{\boldsymbol{x}} : \mathbb{R}^n \to [0,1]$

The *Joint Probability Density Function (Joint PDF)* is defined as

$$f_{\boldsymbol{x}}(x) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F_{\boldsymbol{x}}(x)$$

- $f_{\boldsymbol{x}} : \mathbb{R}^n \to \mathbb{R}$
- $f_{\boldsymbol{x}}(x) \geqslant 0 \ , \quad \forall x \in \mathbb{R}^n$

The Joint PDF represents a complete probabilistic model of the n phenomena described by $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$.

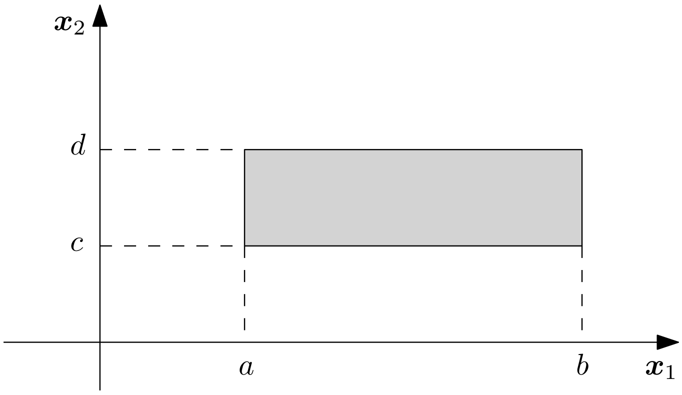
The *Marginal PDF* is defined as

$$f_{\boldsymbol{x}_1}(x_1) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1 \text{ times}} f_{\boldsymbol{x}}\left(x_1, y_2, \dots, y_n\right) dy_2 \dots dy_n$$

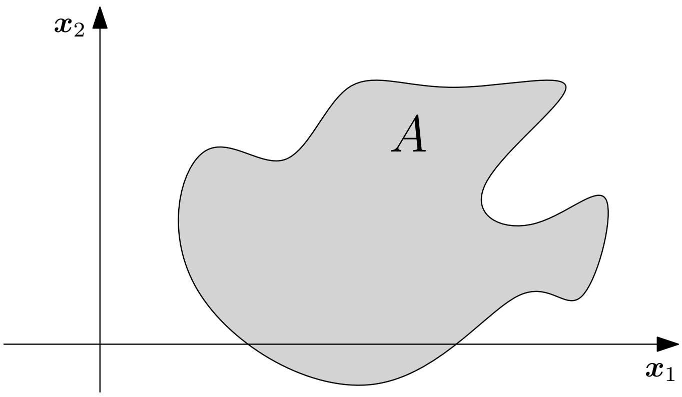
$$f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(x_1, x_2) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-2 \text{ times}} f_{\boldsymbol{x}}\left(x_1, x_2, y_3, \dots, y_n\right) dy_3 \dots dy_n$$

Example:

Given two random variables $\boldsymbol{x}_1, \boldsymbol{x}_2, \quad \boldsymbol{x} = [\boldsymbol{x}_1, \ \boldsymbol{x}_2]^T$.
Notation: $f_{\boldsymbol{x}}(x) = f_{\boldsymbol{x}_1,\boldsymbol{x}_2}\left(x_1, x_2\right)$ (the two notations are equivalent)
$\mathbb{P}\left\{a < \boldsymbol{x}_1 \leqslant b \text{ and } c < \boldsymbol{x}_2 \leqslant d\right\} = \int_c^d \int_a^b f_{\boldsymbol{x}_1,\boldsymbol{x}_2}\left(y_1, y_2\right) dy_1 dy_2$



$$\mathbb{P}\left\{\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}\in A\right\}=\mathbb{P}\{\boldsymbol{x}\in A\}=\iint_A f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(y_1,y_2)\,dy_1dy_2$$



Example (*n*-variate Gaussian distribution)

Let $\boldsymbol{x}\in\mathbb{R}^n$ be a vector of random variables. Define:

$$f_{\boldsymbol{x}}(x)=\frac{1}{\sqrt{(2\pi)^n\cdot\det P}}e^{-\frac{1}{2}(x-m)^TP^{-1}(x-m)}$$

$\boldsymbol{x}\sim\mathbb{N}(m,P)$ with $m\in\mathbb{R}^n,\ P=P^T\in\mathbb{R}^{n\times n},\ P>0$ (positive definite)

1.1.2 Mean and Variance

Let $\boldsymbol{x}\in\mathbb{R}$ be a scalar random variable, and $f_{\boldsymbol{x}}(x)$ the corresponding PDF.
The *mean* (or *expected value*) of $\boldsymbol{x}$ is

$$E[\boldsymbol{x}]=\int_{-\infty}^{\infty}xf_{\boldsymbol{x}}(x)\,dx=m_{\boldsymbol{x}}$$

where $E[\cdot]$ denotes the *expectation* operator.
Given a function $g:\mathbb{R}\to\mathbb{R}$, the expected value of $g(\boldsymbol{x})$ is

$$E[g(\boldsymbol{x})]=\int_{-\infty}^{\infty}g(x)f_{\boldsymbol{x}}(x)\,dx$$

The *variance* of $\boldsymbol{x}$ is defined as

$$\mathrm{Var}[\boldsymbol{x}]=E\left[(\boldsymbol{x}-m_{\boldsymbol{x}})^2\right]=\int_{-\infty}^{\infty}(x-m_{\boldsymbol{x}})^2f_{\boldsymbol{x}}(x)\,dx=\sigma_{\boldsymbol{x}}^2$$

The square root of the variance is called *standard deviation*

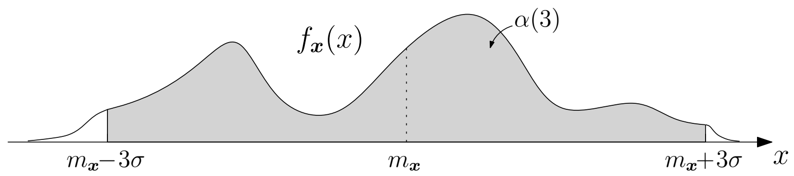
$$\sigma_{\boldsymbol{x}} = \sqrt{\mathrm{Var}[\boldsymbol{x}]}$$

Notation: $\boldsymbol{x} \sim (m_{\boldsymbol{x}}, \sigma_{\boldsymbol{x}}^2)$ means that $\boldsymbol{x}$ is a random variable with mean $m_{\boldsymbol{x}}$ and variance $\sigma_{\boldsymbol{x}}^2$.

1.1.3 Confidence Intervals

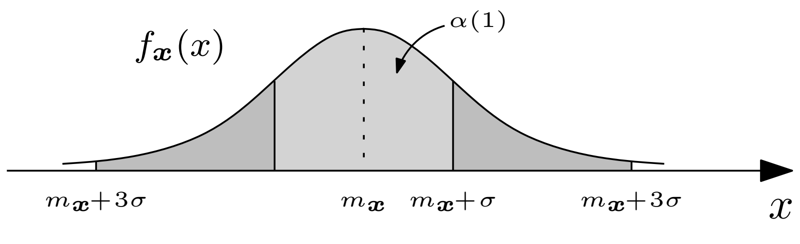
Confidence intervals are intervals within which the random variable falls with a certain probability:

$$\alpha(k) = \mathbb{P}\left\{m_{\boldsymbol{x}} - k\sigma_{\boldsymbol{x}} < \boldsymbol{x} \leq m_{\boldsymbol{x}} + k\sigma_{\boldsymbol{x}}\right\} \qquad k = 1, 2, 3, \dots$$



For a Gaussian random variable:

k	$\alpha(k)$
1	0.683
2	0.954
3	0.997
$\vdots$	
6	0.999999998



1.1.4 Covariance e Correlation

Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be two random variables. The *cross-covariance* of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is

$$\begin{aligned}\sigma_{\boldsymbol{x}_1\boldsymbol{x}_2} &= E\left[(\boldsymbol{x}_1 - m_{\boldsymbol{x}_1})(\boldsymbol{x}_2 - m_{\boldsymbol{x}_2})\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x_1 - m_{\boldsymbol{x}_1})(x_2 - m_{\boldsymbol{x}_2})\, f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(x_1, x_2)\, dx_1 dx_2\end{aligned}$$

The random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are *independent* if

$$f_{\boldsymbol{x}_1\boldsymbol{x}_2}(x_1, x_2) = f_{\boldsymbol{x}_1}(x_1) \cdot f_{\boldsymbol{x}_2}(x_2)$$

The random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are *uncorrelated* if

$$E\left[\boldsymbol{x}_1\boldsymbol{x}_2\right] = E\left[\boldsymbol{x}_1\right] \cdot E\left[\boldsymbol{x}_2\right] = m_{\boldsymbol{x}_1} \cdot m_{\boldsymbol{x}_2}$$

which is equivalent to $\sigma_{\boldsymbol{x}_1\boldsymbol{x}_2} = 0$.

Exercise: Prove that if $\boldsymbol{x}_1, \boldsymbol{x}_2$ are independent then they are uncorrelated.

Let $\boldsymbol{x}$ be a vector of random variables, $\boldsymbol{x} \in \mathbb{R}^n$. The mean of $\boldsymbol{x}$ is

$$\begin{aligned} E[\boldsymbol{x}] &= \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ times}} x f_{\boldsymbol{x}}(x)\, dx_1 \ldots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} f_{\boldsymbol{x}}\left(x_1, x_2, \ldots, x_n\right)\, dx_1 dx_2 \ldots dx_n \end{aligned}$$

Notice that $E[\boldsymbol{x}] \in \mathbb{R}^n$.

Let us consider the first entry, i.e. that related to $\boldsymbol{x}_1$:

$$\begin{aligned} E[\boldsymbol{x}_1] &= \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n \text{ times}} x_1 f_{\boldsymbol{x}}(x_1, \ldots, x_n)\, dx_n \ldots dx_1 \\ &= \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1 \text{ times}} x_1 f_{\boldsymbol{x}}(x_1, \ldots, x_{n-1})\, dx_{n-1} \ldots dx_1 \\ &\vdots \\ &= \int_{-\infty}^{\infty} x_1 f_{\boldsymbol{x}_1}\left(x_1\right)\, dx_1 \end{aligned}$$

Thus

$$E[\boldsymbol{x}] = \int_{-\infty}^{\infty} x f_{\boldsymbol{x}}(x)\, dx = \begin{bmatrix} E[\boldsymbol{x}_1] \\ E[\boldsymbol{x}_2] \\ \vdots \\ E[\boldsymbol{x}_n] \end{bmatrix} = m_{\boldsymbol{x}} \in \mathbb{R}^n$$

If $\boldsymbol{x} \in \mathbb{R}^n$ the concept of variance is expressed by means of the *covariance matrix*

$$\underbrace{\mathrm{Cov}(\boldsymbol{x})}_{\mathbb{R}^{n\times n}} = P_{\boldsymbol{x}} = E[\underbrace{(\boldsymbol{x} - m_{\boldsymbol{x}})}_{\mathbb{R}^{n\times 1}} \underbrace{(\boldsymbol{x} - m_{\boldsymbol{x}})^T}_{\mathbb{R}^{1\times n}}]$$

$$= E\left[\begin{pmatrix} \boldsymbol{x}_1 - m_{\boldsymbol{x}_1} \\ \boldsymbol{x}_2 - m_{\boldsymbol{x}_2} \\ \vdots \\ \boldsymbol{x}_n - m_{\boldsymbol{x}_n} \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{x}_1 - m_{\boldsymbol{x}_1} & \boldsymbol{x}_2 - m_{\boldsymbol{x}_2} & \boldsymbol{x}_n - m_{\boldsymbol{x}_n} \end{pmatrix}\right]$$

$$= \begin{bmatrix} E\left[(\boldsymbol{x}_1 - m_{\boldsymbol{x}_1})^2\right] & E\left[(\boldsymbol{x}_1 - m_{\boldsymbol{x}_1})(\boldsymbol{x}_2 - m_{\boldsymbol{x}_2})\right] & \ldots & E\left[(\boldsymbol{x}_1 - m_{\boldsymbol{x}_1})(\boldsymbol{x}_n - m_{\boldsymbol{x}_n})\right] \\ E\left[(\boldsymbol{x}_2 - m_{\boldsymbol{x}_2})(\boldsymbol{x}_1 - m_{\boldsymbol{x}_1})\right] & E\left[(\boldsymbol{x}_2 - m_{\boldsymbol{x}_2})^2\right] & \ldots & E\left[(\boldsymbol{x}_2 - m_{\boldsymbol{x}_2})(\boldsymbol{x}_n - m_{\boldsymbol{x}_n})\right] \\ & \vdots & & \\ E\left[(\boldsymbol{x}_n - m_{\boldsymbol{x}_n})(\boldsymbol{x}_1 - m_{\boldsymbol{x}_1})\right] & E\left[(\boldsymbol{x}_n - m_{\boldsymbol{x}_n})(\boldsymbol{x}_2 - m_{\boldsymbol{x}_2})\right] & \ldots & E\left[(\boldsymbol{x}_n - m_{\boldsymbol{x}_n})^2\right] \end{bmatrix}$$

Properties:

- $P_{\boldsymbol{x}} = P_{\boldsymbol{x}}^T$ symmetric
- $P_{\boldsymbol{x}} \geq 0$ non-negative definite
- The diagonal entries of $P_{\boldsymbol{x}}$ are the variances of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

1.1.5 Gaussian Random Variables

Let $\boldsymbol{x} \in \mathbb{R}^n$, the PDF is defined as

$$f_{\boldsymbol{x}}(x) = \frac{1}{\sqrt{(2\pi)^n \det P}} e^{-\frac{1}{2}(x-m)^T P^{-1}(x-m)}$$

with $m \in \mathbb{R}^n, P = P^T > 0, P \in \mathbb{R}^{n\times n}$. Notation: $\boldsymbol{x} \sim \mathbb{N}(m, P)$.
Properties:

- $E[\boldsymbol{x}] = m$
- $E\left[(\boldsymbol{x} - m)(\boldsymbol{x} - m)^T\right] = P$

  In the scalar case ($n = 1$):
  $\boldsymbol{x} \sim \mathbb{N}(m, \sigma^2),\ \ E[\boldsymbol{x}] = m,\ E\left[(\boldsymbol{x} - m)^2\right] = \sigma^2,\ f_{\boldsymbol{x}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}$
- Let $\boldsymbol{x} \sim \mathbb{N}\left(m_{\boldsymbol{x}}, P_{\boldsymbol{x}}\right), \boldsymbol{x} \in \mathbb{R}^n$. Define the new random variable $\boldsymbol{y}$ as

  $$\boldsymbol{y} = \underbrace{A}_{m\times n}\boldsymbol{x} + \underbrace{b}_{m\times 1}, \quad \boldsymbol{y} \in \mathbb{R}^m$$

  Then $\boldsymbol{y} \sim \mathbb{N}\left(Am_{\boldsymbol{x}} + b, AP_{\boldsymbol{x}}A^T\right)$, that is, Gaussianity is preserved by affine transformations.
  Proof:

  - $E[\boldsymbol{y}] = E[A\boldsymbol{x} + b] = AE[\boldsymbol{x}] + b = Am_{\boldsymbol{x}} + b \qquad$ (by the linearity of $E[\cdot]$)
  - $E\left[(\boldsymbol{y} - m_{\boldsymbol{y}})(\boldsymbol{y} - m_{\boldsymbol{y}})^T\right] = E\left[(A\boldsymbol{x} + b - (Am_{\boldsymbol{x}} + b))(A\boldsymbol{x} + b - (Am_{\boldsymbol{x}} + b))^T\right]$
    $= E\left[(A\boldsymbol{x} + \not{b} - Am_{\boldsymbol{x}} - \not{b})(A\boldsymbol{x} - Am_{\boldsymbol{x}})^T\right] = E\left[(A(\boldsymbol{x} - m_{\boldsymbol{x}}))(A(\boldsymbol{x} - m_{\boldsymbol{x}}))^T\right]$
    $= E\left[A(\boldsymbol{x} - m_{\boldsymbol{x}})(\boldsymbol{x} - m_{\boldsymbol{x}})^T A^T\right] = A\underbrace{E\left[(\boldsymbol{x} - m_{\boldsymbol{x}})(\boldsymbol{x} - m_{\boldsymbol{x}})^T\right]}_{P_{\boldsymbol{x}}}A^T = AP_{\boldsymbol{x}}A^T$
- If $\boldsymbol{x}, \boldsymbol{y}$ are Gaussian and $E\left[(\boldsymbol{x} - m_{\boldsymbol{x}})(\boldsymbol{y} - m_{\boldsymbol{y}})^T\right] = 0$, then $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent.

Central Limit Theorem

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ independent random variables, $E[\boldsymbol{x}_i] = m_i,\ E\left[(\boldsymbol{x}_i - m_i)^2\right] = \sigma_i^2,\ i = 1, \ldots, n$. Define

$$\boldsymbol{z}_n = \frac{\sum_{i=1}^{n} \boldsymbol{x}_i - \sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

Then,

$$\lim_{n \to +\infty} f_{\boldsymbol{z}_n}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \sim \mathbb{N}(0,1) \qquad \text{(Standard Gaussian distribution)}$$

Special case:

$\boldsymbol{x}_i$ are independent and identically distributed (i.i.d), that is:

$m_1 = m_2 = \ldots = m_n = m$ and $\sigma_1 = \sigma_2 = \ldots = \sigma_n = \sigma$.

$$\boldsymbol{z}_n = \frac{\left(\sum_{i=1}^{n} \boldsymbol{x}_i\right) - nm}{\sqrt{n \cdot \sigma^2}} = \frac{\left(\frac{1}{n} \cdot \sum_{i=1}^{n} \boldsymbol{x}_i\right) - m}{\sqrt{\frac{\sigma^2}{n}}} \sim \mathbb{N}(0,1)$$

Thus

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \underset{n \to \infty}{\sim} \mathbb{N}\left(m, \frac{\sigma^2}{n}\right)$$

1.1.6 Functions of random variables

Scalar case:

Let $\boldsymbol{x} \in \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$. $\boldsymbol{x} \sim f_{\boldsymbol{x}}(x)$, $\boldsymbol{y} = g(\boldsymbol{x})$. We want to find $f_{\boldsymbol{y}}(y)$.

Theorem:

$$f_{\boldsymbol{y}}(y) = \sum_{i=1}^{m} \frac{f_{\boldsymbol{x}}(x_i)}{|g'(x_i)|}$$

where

- $x_1, x_2, \ldots, x_m$ satisfy $g(x_1) = g(x_2) = \cdots = g(x_m) = y$
- $g'(x_i) = \left.\frac{d}{dx} g(x)\right|_{x = x_i}$

Multivariable case:

Let $\boldsymbol{x} \in \mathbb{R}^n,\ g : \mathbb{R}^n \to \mathbb{R}^n,\ \boldsymbol{y} = g(\boldsymbol{x})$. One has

$$f_{\boldsymbol{y}}(y) = \sum_{i=1}^{m} \frac{f_{\boldsymbol{x}}(x_i)}{|\det J_g(x_i)|}$$

where:

- $x_1, x_2, \ldots, x_m$ satisfy $g(x_1) = g(x_2) = \ldots = g(x_m) = y\ , \quad x_i \in \mathbb{R}^n$

- $J_g(x_i) = \frac{\partial g}{\partial x}\Big|_{x=x_i} = \left.\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}\right|_{x=x_i} \in \mathbb{R}^{n\times n}$ Jacobian of g (computed in x_i)

Linear case:

$\boldsymbol{y} = A\boldsymbol{x}$ with $\det(A) \neq 0$, $J(x) = A$, $m = 1$, $x_1 = A^{-1}y$

$$f_{\boldsymbol{y}}(y) = \frac{f_{\boldsymbol{x}}\left(A^{-1}y\right)}{|\det(A)|}$$

Example:

$\boldsymbol{x} \sim \mathbb{U}[-2,2]$, $\boldsymbol{y} = \boldsymbol{x}^2 = g(\boldsymbol{x})$. We want to find $f_{\boldsymbol{y}}(y)$.

For all y we must find $x_1, \dots x_m$ such that $x_i^2 = y$.

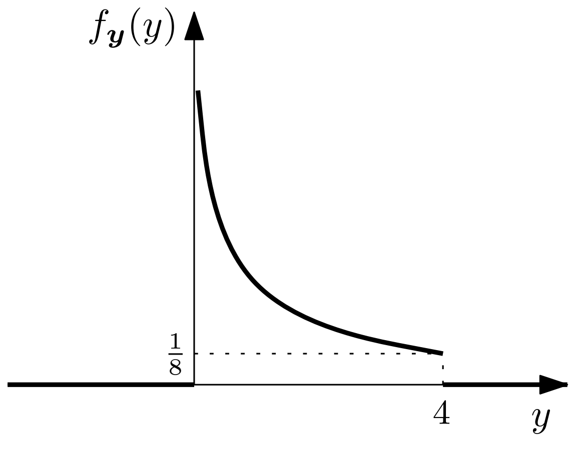
$$\begin{cases} \text{if } y > 0 \rightarrow x_1 = \sqrt{y},\ x_2 = -\sqrt{y} \ \rightarrow\ m = 2 \\ \text{if } y < 0 \rightarrow \ \text{no solutions} \end{cases}$$

For $y > 0$, $\quad f_{\boldsymbol{y}}(y) = \dfrac{f_{\boldsymbol{x}}\left(x_1\right)}{|g'\left(x_1\right)|} + \dfrac{f_{\boldsymbol{x}}\left(x_2\right)}{|g'\left(x_2\right)|}$

$g'(x) = 2x$

$$f_{\boldsymbol{x}} = \begin{cases} 1/4 & \text{if } -2 < x \leq 2 \\ 0 & \text{else} \end{cases}$$

$$f_{\boldsymbol{y}}(y) = \begin{cases} \dfrac{1/4}{|2\sqrt{y}|} + \dfrac{1/4}{|-2\sqrt{y}|} & \text{if } 0 < y \leq 4 \\ 0 & \text{else} \end{cases} = \begin{cases} \dfrac{1}{4\sqrt{y}} & \text{if } 0 < y \leq 4 \\ 0 & \text{else} \end{cases}$$



Exercise:

Verify that $\displaystyle\int_{-\infty}^{\infty} f_{\boldsymbol{y}}(y)\, dy = 1$.

Exercise:

Let $\boldsymbol{x}_i \sim \mathbb{U}\left[-\frac{1}{2}, \frac{1}{2}\right]$ independent. Compute the PDF of

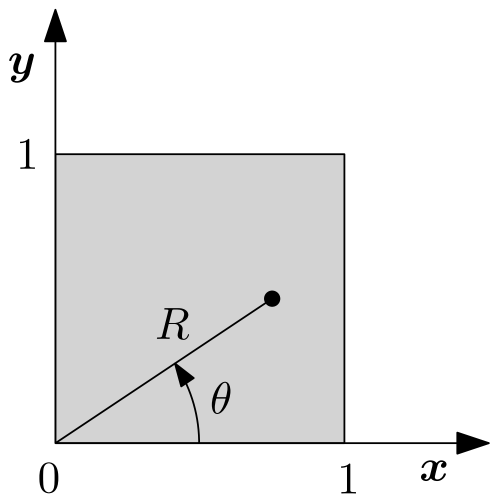
$$Y_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$$

for $n = 2, 3$ and compare the resulting PDF with the Gaussian $\mathbb{N}\left(0, \frac{1}{12n}\right)$.
Hint ($n = 2$):
$\boldsymbol{y}_1 = \frac{1}{2}(\boldsymbol{x}_1 + \boldsymbol{x}_2)$
$\boldsymbol{y}_2 = \boldsymbol{x}_2$
$$\boldsymbol{y} = \underbrace{\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}}_{A} \boldsymbol{x}$$
Compute the PDF of $\boldsymbol{y}$ and then the marginal PDF of $\boldsymbol{y}_1$.

Exercise:

Let $\boldsymbol{x}, \boldsymbol{y} \sim \mathbb{U}[0, 1]$ independent random variables.
Consider the new two random variables:

$$R = \sqrt{\boldsymbol{x}^2 + \boldsymbol{y}^2}, \quad \theta = \arctan\frac{\boldsymbol{y}}{\boldsymbol{x}}, \quad g : \mathbb{R}^2 \to \mathbb{R}^2 \qquad \text{(polar coordinates)}$$



- Compute the Joint PDF of the vector $w = \begin{bmatrix} R \\ \theta \end{bmatrix}$
- Repeat assuming that $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \mathbb{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$

1.1.7 Correlation Index

Let $\boldsymbol{x}, \boldsymbol{y}$ be two scalar random variables. Let us define the correlation index as

$$\rho_{\boldsymbol{xy}} = \frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}} \cdot \sigma_{\boldsymbol{y}}}$$

Property:

- $-1 \leq \rho_{\boldsymbol{xy}} \leq 1$

Example:

Let $\boldsymbol{x} \sim \mathbb{N}(0, 1)$, and set $\boldsymbol{y} = 3\boldsymbol{x}$. One has $\boldsymbol{y} \sim \mathbb{N}(Am_{\boldsymbol{x}} + b,\, AP_{\boldsymbol{x}}A^T) = \mathbb{N}(0, 9)$.

$$\sigma_{\boldsymbol{xy}} = E[(\boldsymbol{x} - 0)(\boldsymbol{y} - 0)] = E[\boldsymbol{x} \cdot 3\boldsymbol{x}] = 3E\left[\boldsymbol{x}^2\right] = 3$$

$$\rho_{\boldsymbol{xy}} = \frac{\sigma_{\boldsymbol{xy}}}{\sigma_{\boldsymbol{x}} \cdot \sigma_{\boldsymbol{y}}} = \frac{E\left[(\boldsymbol{x} - m_{\boldsymbol{x}})(\boldsymbol{y} - m_{\boldsymbol{y}})\right]}{\sqrt{E\left[(\boldsymbol{x} - m_{\boldsymbol{x}})^2\right]} \cdot \sqrt{E\left[(\boldsymbol{y} - m_{\boldsymbol{y}})^2\right]}} = \frac{3}{\sqrt{1}\sqrt{9}} = 1.$$

So, $\boldsymbol{x}$ and $\boldsymbol{y}$ show a perfect positive correlation.

1.1.8 Conditional Distributions

Given two random variables $\boldsymbol{x}, \boldsymbol{y}$ with Joint PDF $f_{\boldsymbol{x},\boldsymbol{y}}(x, y)$, we are interested in $\boldsymbol{x}$. Remember that the Marginal PDF

$$f_{\boldsymbol{x}}(x) = \int_{-\infty}^{\infty} f_{\boldsymbol{x},\boldsymbol{y}}(x, y)\, dy$$

denotes the knowledge of $\boldsymbol{x}$ before observing $\boldsymbol{y}$, and hence it represents the a-priori information on $\boldsymbol{x}$.
Assume that we measured $\boldsymbol{y}$, that is $\boldsymbol{y} = \bar{y}$. The question is: how the information on $\boldsymbol{y}$ (i.e., the data) affects the information on $\boldsymbol{x}$?
The a-posteriori PDF of $\boldsymbol{x}$ after observing $\boldsymbol{y} = \bar{y}$ is the *Conditional PDF of* $\boldsymbol{x}$ *given* $\boldsymbol{y}$

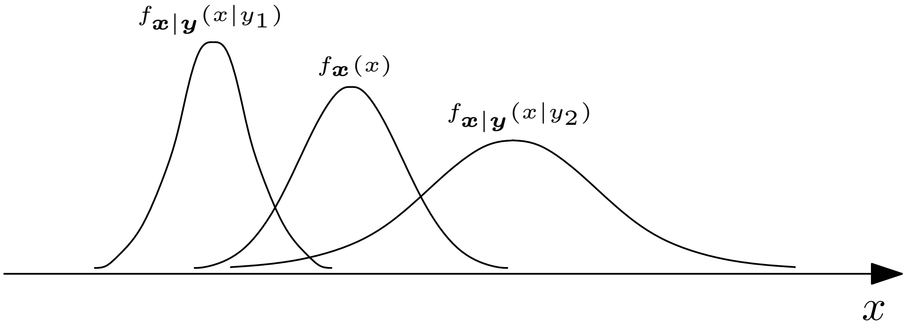
$$f_{\boldsymbol{x}|\boldsymbol{y}}(x|\bar{y}) = \frac{f_{\boldsymbol{xy}}(x, \bar{y})}{f_{\boldsymbol{y}}(\bar{y})}$$

We can define the *Conditional Mean* and *Conditional Variance* as:

$$m_{\boldsymbol{x}|\boldsymbol{y}} = E[\boldsymbol{x}|\boldsymbol{y}] = \int_{-\infty}^{\infty} x f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)\, dx \quad \text{(Conditional mean)}$$

$$\begin{aligned}\sigma^2_{\boldsymbol{x}|\boldsymbol{y}} = \mathrm{Var}(\boldsymbol{x}|\boldsymbol{y}) &= E\left[\left(\boldsymbol{x} - m_{\boldsymbol{x}|\boldsymbol{y}}\right)^2 \mid \boldsymbol{y}\right] \\ &= \int_{-\infty}^{\infty} \left(x - m_{\boldsymbol{x}|\boldsymbol{y}}\right)^2 f_{\boldsymbol{x}|\boldsymbol{y}}(x|y)\, dx\end{aligned} \quad \text{(Conditional variance)}$$

Notice that $m_{\boldsymbol{x}|\boldsymbol{y}}(y)$ and $\sigma^2_{\boldsymbol{x}|\boldsymbol{y}}(y)$ can be viewed as functions of the measurement y. They represent how the mean and the variance change due to the observation $\boldsymbol{y} = y$.



Gaussian conditional distributions

Let $\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \mathbb{N}(m, P) = \mathbb{N}\left(\begin{bmatrix} m_{\boldsymbol{x}} \\ m_{\boldsymbol{y}} \end{bmatrix}, \begin{bmatrix} P_{\boldsymbol{x}} & P_{\boldsymbol{xy}} \\ P_{\boldsymbol{xy}}^T & P_{\boldsymbol{y}} \end{bmatrix}\right) \quad , \quad \boldsymbol{x} \in \mathbb{R}^{n_{\boldsymbol{x}}}$

The a-priori PDF of $\boldsymbol{x}$ is

$$f_{\boldsymbol{x}}(x) = \frac{1}{\sqrt{(2\pi)^{n_{\boldsymbol{x}}} \det(P_{\boldsymbol{x}})}} e^{-\frac{1}{2}(x-m_{\boldsymbol{x}})^\top P_{\boldsymbol{x}}^{-1}(x-m_{\boldsymbol{x}})}$$

The Conditional PDF of $\boldsymbol{x}$ given $\boldsymbol{y} = y$ is

$$f_{\boldsymbol{x}|\boldsymbol{y}}(x|y) \sim \mathbb{N}\left(m_{\boldsymbol{x}|\boldsymbol{y}}, P_{\boldsymbol{x}|\boldsymbol{y}}\right)$$

where

$$m_{\boldsymbol{x}|\boldsymbol{y}} = m_{\boldsymbol{x}} + P_{\boldsymbol{xy}} \cdot P_{\boldsymbol{y}}^{-1}(y - m_{\boldsymbol{y}})$$
$$P_{\boldsymbol{x}|\boldsymbol{y}} = P_{\boldsymbol{x}} - P_{\boldsymbol{xy}} P_{\boldsymbol{y}}^{-1} P_{\boldsymbol{xy}}^T$$

Example (estimation problem)

Let $\boldsymbol{x}$ be an unknown scalar quantity.
Suppose to measure $\boldsymbol{x}$ by means of a sensor that provides the following measure

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}$$

where $\boldsymbol{v}$ denotes the measurement noise.
Assumptions:

- $\boldsymbol{x} \sim \mathbb{N}(m_{\boldsymbol{x}}, \sigma_{\boldsymbol{x}}^2)$ a-priori information on $\boldsymbol{x}$
- $\boldsymbol{v} \sim \mathbb{N}(0, \sigma_{\boldsymbol{v}}^2)$ information on the noise
- $\boldsymbol{x}$ and $\boldsymbol{v}$ independent

Problem: Construct an estimator of $\boldsymbol{x}$ based on the observation $\boldsymbol{y} = y$.
An estimator $T : \mathbb{R} \to \mathbb{R}$ is a function that associates to y an estimate $\hat{x}$ of $\boldsymbol{x}$

$$\hat{x} = T(y)$$

Idea: use the conditional mean.

$$
\begin{aligned}
\hat{x} &= E[\boldsymbol{x}|\boldsymbol{y}] = m_{\boldsymbol{x}} + P_{\boldsymbol{xy}} P_{\boldsymbol{y}}^{-1} \left(y - m_{\boldsymbol{y}}\right) \\
m_{\boldsymbol{y}} &= E[\boldsymbol{y}] = E[\boldsymbol{x} + \boldsymbol{v}] = E[\boldsymbol{x}] + E[\boldsymbol{v}] = m_{\boldsymbol{x}} + 0 = m_{\boldsymbol{x}} \\
P_{\boldsymbol{xy}} &= E\left[\left(\boldsymbol{x} - m_{\boldsymbol{x}}\right)\left(\boldsymbol{y} - m_{\boldsymbol{y}}\right)\right] = E\left[\left(\boldsymbol{x} - m_{\boldsymbol{x}}\right)\left(\boldsymbol{x} + \boldsymbol{v} - m_{\boldsymbol{x}}\right)\right] = \\
&= E\left[\left(\boldsymbol{x} - m_{\boldsymbol{x}}\right)^2\right] + \underbrace{E\left[\left(\boldsymbol{x} - m_{\boldsymbol{x}}\right)\boldsymbol{v}\right]}_{0} = \sigma_{\boldsymbol{x}}^2 \\
P_{\boldsymbol{y}} &= E\left[\left(\boldsymbol{y} - m_{\boldsymbol{y}}\right)^2\right] = E\left[\left(\boldsymbol{x} + \boldsymbol{v} - m_{\boldsymbol{x}}\right)^2\right] = \\
&= E\left[\left(\boldsymbol{x} - m_{\boldsymbol{x}}\right)^2 + \left(\boldsymbol{v} - 0\right)^2 + 2\left(\boldsymbol{x} - m_{\boldsymbol{x}}\right)\left(\boldsymbol{v} - 0\right)\right] = \sigma_{\boldsymbol{x}}^2 + \sigma_{\boldsymbol{v}}^2
\end{aligned}
$$

So,

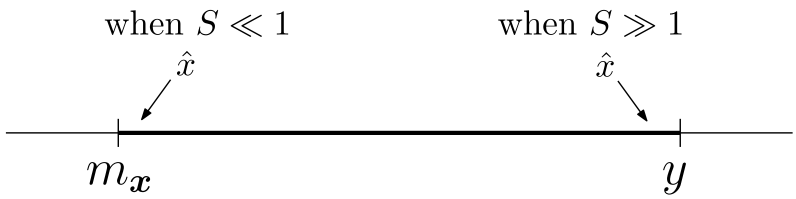
$$
\hat{x} = m_{\boldsymbol{x}} + \frac{\sigma_{\boldsymbol{x}}^2}{\sigma_{\boldsymbol{x}}^2 + \sigma_{\boldsymbol{v}}^2}(y - m_{\boldsymbol{x}}) = m_{\boldsymbol{x}} + \frac{\frac{\sigma_{\boldsymbol{x}}^2}{\sigma_{\boldsymbol{v}}^2}}{\frac{\sigma_{\boldsymbol{x}}^2}{\sigma_{\boldsymbol{v}}^2} + 1}(y - m_{\boldsymbol{x}})
$$

Define the signal-to-noise ratio as $S = \dfrac{\sigma_{\boldsymbol{x}}^2}{\sigma_{\boldsymbol{v}}^2}$, $S \in (0, +\infty)$.

Thus,

$$
\begin{aligned}
\hat{x} &= m_{\boldsymbol{x}} + \frac{S}{S+1}\left(y - m_{\boldsymbol{x}}\right) = m_{\boldsymbol{x}} + \frac{S}{S+1} y - \frac{S}{S+1} m_{\boldsymbol{x}} \\
&= \left(1 - \frac{S}{S+1}\right) m_{\boldsymbol{x}} + \frac{S}{S+1} y = \frac{1}{S+1} m_{\boldsymbol{x}} + \frac{S}{S+1} y \\
&= \alpha m_{\boldsymbol{x}} + (1-\alpha) y, \quad \alpha = \frac{1}{S+1}, \ \alpha \in (0,1)
\end{aligned}
$$

Then, $\hat{x}$ is a convex combination of $m_{\boldsymbol{x}}$ and y.



The variance of $\boldsymbol{x}$ after the observation is (since Gaussian):

$$
\begin{aligned}
\sigma_{\boldsymbol{x}|\boldsymbol{y}}^2 &= P_{\boldsymbol{x}} - P_{\boldsymbol{xy}} P_{\boldsymbol{y}}^{-1} P_{\boldsymbol{xy}}^T = \sigma_{\boldsymbol{x}}^2 - \sigma_{\boldsymbol{x}}^2 \frac{1}{\sigma_{\boldsymbol{x}}^2 + \sigma_{\boldsymbol{v}}^2} \sigma_{\boldsymbol{x}}^2 = \\
&= \sigma_{\boldsymbol{x}}^2 \left(1 - \frac{\sigma_{\boldsymbol{x}}^2}{\sigma_{\boldsymbol{x}}^2 + \sigma_{\boldsymbol{v}}^2}\right) = \sigma_{\boldsymbol{x}}^2 \cdot \frac{\sigma_{\boldsymbol{v}}^2}{\sigma_{\boldsymbol{x}}^2 + \sigma_{\boldsymbol{v}}^2} = \sigma_{\boldsymbol{x}}^2 \frac{1}{S+1} < \sigma_{\boldsymbol{x}}^2 \,, \quad \forall S \in (0, +\infty)
\end{aligned}
$$

So, the knowledge of $\boldsymbol{y} = y$ reduces the variance on $\boldsymbol{x}$.