# Statistical Signal Processing

i.i.d: independently identically distributed

## 1. Math

$\pi \approx 3.141\,59$ $\quad e \approx 2.718\,28$ $\quad \sqrt{2} \approx 1.414$ $\quad \sqrt{3} \approx 1.732$

**Binome, Trinome**
$(a \pm b)^2 = a^2 \pm 2ab + b^2$ $\qquad a^2 - b^2 = (a-b)(a+b)$
$(a \pm b)^3 = a^3 \pm 3a^2b + 3ab^2 \pm b^3$
$(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$

**Folgen und Reihen**
$\sum\limits_{k=1}^{n} k = \frac{n(n+1)}{2}$ Aritmetrische Summenformel $\qquad \sum\limits_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q}$ Geometrische Summenformel $\qquad \sum\limits_{n=0}^{\infty} \frac{z^n}{n!} = e^z$ Exponentialreihe

**Mittelwerte** $\quad (\sum$ von $i$ bis $N)$ $\qquad$ (Median: Mitte einer geordneten Liste)
$\overline{x}_{\text{ar}} = \frac{1}{N}\sum x_i$ (Arithmetisches) $\geq \overline{x}_{\text{geo}} = \sqrt[N]{\prod x_i}$ (Geometrisches Mittel) $\geq \overline{x}_{\text{hm}} = \frac{N}{\sum \frac{1}{x_i}}$ (Harmonisches)

**Ungleichungen:** $\qquad$ Bernoulli-Ungleichung: $(1+x)^n \geq 1 + nx$
$\left| |x| - |y| \right| \leq |x \pm y| \leq |x| + |y|$ Dreiecksungleichung $\qquad \left|\boldsymbol{x}^\top \cdot \boldsymbol{y}\right| \leq \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|$ Cauchy-Schwarz-Ungleichung

**Mengen:** De Morgan: $\overline{A \cap B} = \overline{A} \uplus \overline{B}$ $\qquad \overline{A \uplus B} = \overline{A} \cap \overline{B}$

### 1.1. Exp. und Log.

$e^x := \lim\limits_{n\to\infty}\left(1 + \frac{x}{n}\right)^n$ $\qquad e \approx 2,71828$
$a^x = e^{x \ln a}$ $\qquad \log_a x = \frac{\ln x}{\ln a}$ $\qquad \ln x \leq x - 1$
$\ln(x^a) = a\ln(x)$ $\qquad \ln(\frac{x}{a}) = \ln x - \ln a$ $\qquad \log(1) = 0$

### 1.2. Matrizen $\boldsymbol{A} \in \mathbb{K}^{m\times n}$

$\boldsymbol{A} = (a_{ij}) \in \mathbb{K}^{m\times n}$ hat $m$ Zeilen (Index $i$) und $n$ Spalten (Index $j$)
$(\boldsymbol{A}+\boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$ $\qquad (\boldsymbol{A}\cdot\boldsymbol{B})^\top = \boldsymbol{B}^\top \cdot \boldsymbol{A}^\top$
$(\boldsymbol{A}^\top)^{-1} = (\boldsymbol{A}^{-1})^\top$ $\qquad (\boldsymbol{A}\cdot\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
$\dim \mathbb{K} = n = \operatorname{rang} \boldsymbol{A} + \dim\ker \boldsymbol{A}$ $\qquad \operatorname{rang}\boldsymbol{A} = \operatorname{rang}\boldsymbol{A}^\top$

**1.2.1 Quadratische Matrizen** $A \in \mathbb{K}^{n\times n}$
regulär/invertierbar/nicht-singulär $\Leftrightarrow \det(\boldsymbol{A}) \neq 0 \Leftrightarrow \operatorname{rang}\boldsymbol{A} = n$
singulär/nicht-invertierbar $\Leftrightarrow \det(\boldsymbol{A}) = 0 \Leftrightarrow \operatorname{rang}\boldsymbol{A} \neq n$
orthogonal $\Leftrightarrow \boldsymbol{A}^\top = \boldsymbol{A}^{-1} \Rightarrow \det(\boldsymbol{A}) = \pm 1$
symmetrisch: $\boldsymbol{A} = \boldsymbol{A}^\top$ $\qquad$ schiefsymmetrisch: $\boldsymbol{A} = -\boldsymbol{A}^\top$

**1.2.2 Determinante von** $\boldsymbol{A} \in \mathbb{K}^{n\times n}$: $\det(\boldsymbol{A}) = |\boldsymbol{A}|$

$$\det\begin{bmatrix}\boldsymbol{A} & \boldsymbol{0}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix} = \det\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{0} & \boldsymbol{D}\end{bmatrix} = \det(\boldsymbol{A})\det(\boldsymbol{D})$$

$\det(\boldsymbol{A}) = \det(\boldsymbol{A}^\top)$ $\qquad \det(\boldsymbol{A}^{-1}) = \det(\boldsymbol{A})^{-1}$
$\det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A})\det(\boldsymbol{B}) = \det(\boldsymbol{B})\det(\boldsymbol{A}) = \det(\boldsymbol{B}\boldsymbol{A})$
Hat $\boldsymbol{A}$ 2 linear abhäng. Zeilen/Spalten $\Rightarrow |\boldsymbol{A}| = 0$

**1.2.3 Eigenwerte (EW) $\lambda$ und Eigenvektoren (EV) $\underline{v}$**

$$\boxed{\boldsymbol{A}\underline{v} = \lambda\underline{v} \qquad \det\boldsymbol{A} = \prod \lambda_i \qquad \operatorname{Sp}\boldsymbol{A} = \sum a_{ii} = \sum \lambda_i}$$

Eigenwerte: $\det(\boldsymbol{A} - \lambda\boldsymbol{1}) = 0$ Eigenvektoren: $\ker(\boldsymbol{A} - \lambda_i\boldsymbol{1}) = \underline{v}_i$
EW von Dreieck/Diagonal Matrizen sind die Elem. der Hauptdiagonale.

**1.2.4 Spezialfall $2\times 2$ Matrix $A$**
$\det(\boldsymbol{A}) = ad - bc$
$\operatorname{Sp}(\boldsymbol{A}) = a + d$
$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}^{-1} = \frac{1}{\det\boldsymbol{A}}\begin{bmatrix}d & -b\\ -c & a\end{bmatrix}$$
$$\lambda_{1/2} = \frac{\operatorname{Sp}\boldsymbol{A}}{2} \pm \sqrt{\left(\frac{\operatorname{sp}\boldsymbol{A}}{2}\right)^2 - \det\boldsymbol{A}}$$

**1.2.5 Differentiation**
$$\frac{\partial \boldsymbol{x}^\top\boldsymbol{y}}{\partial\boldsymbol{x}} = \frac{\partial \boldsymbol{y}^\top\boldsymbol{x}}{\partial\boldsymbol{x}} = \boldsymbol{y} \qquad \frac{\partial \boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{x}}{\partial\boldsymbol{x}} = (\boldsymbol{A}+\boldsymbol{A}^\top)\boldsymbol{x}$$
$$\frac{\partial \boldsymbol{x}^\top\boldsymbol{A}\boldsymbol{y}}{\partial\boldsymbol{A}} = \boldsymbol{x}\boldsymbol{y}^\top \qquad \frac{\partial\det(\boldsymbol{B}\boldsymbol{A}\boldsymbol{C})}{\partial\boldsymbol{A}} = \det(\boldsymbol{B}\boldsymbol{A}\boldsymbol{C})\left(\boldsymbol{A}^{-1}\right)^\top$$

**1.2.6 Ableitungsregeln** $(\forall \lambda, \mu \in \mathbb{R})$
Linearität: $(\lambda f + \mu g)'(x) = \lambda f'(x) + \mu g'(x_0)$
Produkt: $(f\cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$
Quotient: $\left(\frac{f}{g}\right)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{g(x)^2}$ $\quad \left(\frac{\text{NAZ}-\text{ZAN}}{\text{N}^2}\right)$
Kettenregel $(f(g(x)))' = f'(g(x))g'(x)$

### 1.3. Integrale $\int e^x \mathrm{d}x = e^x = (e^x)'$

Partielle Integration: $\int uw' = uw - \int u'w$
Substitution: $\int f(g(x))g'(x)\,\mathrm{d}x = \int f(t)\,\mathrm{d}t$

| $F(x) - C$ | $f(x)$ | $f'(x)$ |
|---|---|---|
| $\frac{1}{q+1}x^{q+1}$ | $x^q$ | $qx^{q-1}$ |
| $\frac{2\sqrt{ax^3}}{3}$ | $\sqrt{ax}$ | $\frac{a}{2\sqrt{ax}}$ |
| $x\ln(ax) - x$ | $\ln(ax)$ | $\frac{1}{x}$ |
| $\frac{1}{a^2}e^{ax}(ax-1)$ | $x\cdot e^{ax}$ | $e^{ax}(ax+1)$ |
| $\frac{a^x}{\ln(a)}$ | $a^x$ | $a^x\ln(a)$ |
| $-\cos(x)$ | $\sin(x)$ | $\cos(x)$ |
| $\cosh(x)$ | $\sinh(x)$ | $\cosh(x)$ |
| $-\ln\lvert\cos(x)\rvert$ | $\tan(x)$ | $\frac{1}{\cos^2(x)}$ |

$$\int e^{at}\sin(bt)\,\mathrm{d}t = e^{at}\frac{a\sin(bt) + b\cos(bt)}{a^2+b^2}$$
$$\int \frac{\mathrm{d}t}{\sqrt{at+b}} = \frac{2\sqrt{at+b}}{a} \qquad \int t^2e^{at}\,\mathrm{d}t = \frac{(ax-1)^2+1}{a^3}e^{at}$$
$$\int te^{at}\,\mathrm{d}t = \frac{at-1}{a^2}e^{at} \qquad \int xe^{ax^2}\,\mathrm{d}x = \frac{1}{2a}e^{ax^2}$$

**1.3.1 Volumen und Oberfläche von Rotationskörpern um $x$-Achse**
$V = \pi\int_a^b f(x)^2\mathrm{d}x$ $\qquad O = 2\pi\int_a^b f(x)\sqrt{1 + f'(x)^2}\mathrm{d}x$

## 2. Probability Theory Basics

### 2.1. Kombinatorik

Mögliche Variationen/Kombinationen um $k$ Elemente von maximal $n$ Elementen zu wählen bzw. $k$ Elemente auf $n$ Felder zu verteilen:

| | Mit Reihenfolge | Reihenfolge egal |
|---|---|---|
| Mit Wiederholung | $n^k$ | $\binom{n+k-1}{k}$ |
| Ohne Wiederholung | $\frac{n!}{(n-k)!}$ | $\binom{n}{k}$ |

Permutation von $n$ mit jeweils $k$ gleichen Elementen: $\frac{n!}{k_1!\cdot k_2!\cdot\ldots}$
Binomialkoeffizient $\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!\cdot(n-k)!}$
$\binom{n}{0} = 1$ $\quad \binom{n}{1} = n$ $\quad \binom{4}{2} = 6$ $\quad \binom{5}{2} = 10$ $\quad \binom{6}{2} = 15$

### 2.2. Der Wahrscheinlichkeitsraum $(\Omega, \mathbb{F}, \mathrm{P})$

| | | |
|---|---|---|
| **Ergebnismenge** | $\Omega = \{\omega_1, \omega_2, \ldots\}$ | Ergebnis $\omega_j \in \Omega$ |
| **Ereignisalgebra** | $\mathbb{F} = \{A_1, A_2, \ldots\}$ | Ereignis $A_i \subseteq \Omega$ |
| **Wahrscheinlichkeitsmaß** | $\mathrm{P}: \mathbb{F} \to [0,1]$ | $\mathrm{P}(A) = \frac{\lvert A\rvert}{\lvert\Omega\rvert}$ |

### 2.3. Wahrscheinlichkeitsmaß P

$\mathrm{P}(A) = \frac{|A|}{|\Omega|}$ $\qquad \mathrm{P}(A\cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A\cap B)$

**2.3.1 Axiome von Kolmogorow**
Nichtnegativität: $\mathrm{P}(A) \geq 0 \Rightarrow \mathrm{P}: \mathbb{F} \mapsto [0,1]$
Normiertheit: $\mathrm{P}(\Omega) = 1$
Additivität: $\mathrm{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty}\mathrm{P}(A_i)$,
wenn $A_i \cap A_j = \emptyset, \forall i \neq j$

### 2.4. Bedingte Wahrscheinlichkeit

Bedingte Wahrscheinlichkeit für $A$ falls $B$ bereits eingetreten ist:
$\mathrm{P}_B(A) = \mathrm{P}(A|B) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}$

**2.4.1 Totale Wahrscheinlichkeit und Satz von Bayes**
Es muss gelten: $\bigcup\limits_{i\in I} B_i = \Omega$ für $B_i \cap B_j = \emptyset, \forall i \neq j$

Totale Wahrscheinlichkeit: $\mathrm{P}(A) = \sum\limits_{i\in I}\mathrm{P}(A|B_i)\,\mathrm{P}(B_i)$

Satz von Bayes: $\mathrm{P}(B_k|A) = \frac{\mathrm{P}(A|B_k)\,\mathrm{P}(B_k)}{\sum\limits_{i\in I}\mathrm{P}(A|B_i)\,\mathrm{P}(B_i)}$

**Multiplikationssatz:** $\mathrm{P}(A\cap B) = \mathrm{P}(A|B)\,\mathrm{P}(B) = \mathrm{P}(B|A)\,\mathrm{P}(A)$

### 2.5. Zufallsvariable

$X: \Omega \mapsto \Omega'$ ist Zufallsvariable, wenn für jedes Ereignis $A' \in \mathbb{F}'$ im Bildraum ein Ereignis $A$ im Urbildraum $\mathbb{F}$ existiert, sodass $\{\omega \in \Omega |\, X(\omega) \in A'\} \in \mathbb{F}$

### 2.6. Distribution

| Bezeichnung | Abk. | Zusammenhang |
|---|---|---|
| Wahrscheinlichkeitsdichte | pdf | $f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$ |
| Kumulative Verteilungsfkt. | cdf | $F_X(x) = \int\limits_{-\infty}^{x} f_X(\xi)\,\mathrm{d}\xi$ |

Joint CDF: $F_{X,Y}(x,y) = \mathrm{P}(\{X \leq x, Y \leq y\})$

### 2.7. Relations between $f_X(x)$, $f_{X,Y}(x,y)$, $f_{X\,|\,Y}(x|y)$

$$\boxed{\begin{gathered}\underbrace{f_{X,Y}(x,y)}_{\text{Joint PDF}} = f_{X\,|\,Y}(x,y)f_Y(y) = f_{Y\,|\,X}(y,x)f_X(x)\\ \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,\xi)\,\mathrm{d}\xi}_{\text{Marginalization}} = \underbrace{\int_{-\infty}^{\infty} f_{X\,|\,Y}(x,\xi)f_Y(\xi)\,\mathrm{d}\xi}_{\text{Total Probability}} = f_X(x)\end{gathered}}$$

### 2.8. Bedingte Zufallsvariablen

Ereignis A gegeben: $F_{X\,|\,A}(x|A) = \mathrm{P}(\{X \leq x\}|A)$
ZV $Y$ gegeben: $F_{X\,|\,Y}(x|y) = \mathrm{P}(\{X \leq x\}|\{Y = y\})$
$p_{X\,|\,Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$
$f_{X\,|\,Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\mathrm{d}F_{X\,|\,Y}(x|y)}{\mathrm{d}x}$

### 2.9. Unabhängigkeit von Zufallsvariablen

$X_1, \cdots, X_n$ sind stochastisch unabhängig, wenn für jedes $\underline{x} \in \mathbb{R}^n$ gilt:
$$F_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$
$$p_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$$
$$f_{X_1,\cdots,X_n}(x_1,\cdots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

## 3. Common Distributions

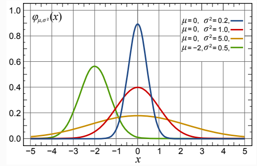
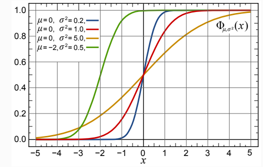
### 3.1. Binomialverteilung $\mathcal{B}(n,p)$ mit $p \in [0,1], n \in \mathbb{N}$

Folge von $n$ Bernoulli-Experimenten
$p$: Wahrscheinlichkeit für Erfolg $\qquad k$: Anzahl der Erfolge

$$p_X(k) = B_{n,p}(k) = \begin{cases}\binom{n}{k}p^k(1-p)^{n-k} & k \in \{0,\ldots,n\}\\ 0 & \text{sonst}\end{cases}$$

$\mathrm{E}[X] = np$ (Erwartungswert) $\qquad \mathrm{Var}[X] = np(1-p)$ (Varianz) $\qquad G_X(z) = (pz + 1 - p)^n$ (Wahrscheinlichkeitserz. Funktion)

### 3.2. Normalverteilung

WDF/PDF:

KVF/CDF:


**WDF:** $\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R}}$ $\qquad \mu \in \mathbb{R}$, $\sigma > 0$

$\mathrm{E}(X) = \mu$ (Erwartungswert) $\qquad \mathrm{Var}(X) = \sigma^2$ (Varianz) $\qquad \varphi_X(\omega) = e^{j\omega\mu - \frac{\omega^2\sigma^2}{2}}$ (Charakt. Funktion)

### 3.3. Sonstiges

**Gammadistribution** $\Gamma(\alpha,\beta)$: $\mathrm{E}[X] = \frac{\alpha}{\beta}$
**Exponential:** $f(x,\lambda) = \lambda e^{-\lambda x}$ $\qquad \mathrm{E}[X] = \lambda^{-1}$ $\qquad \mathrm{Var}[X] = \lambda^{-2}$

## 4. Wichtige Parameter

### 4.1. Erwartungswert (1. zentrales Moment)

gibt den mittleren Wert einer Zufallsvariablen an

$$\boxed{\mu_X = \mathrm{E}[X] = \underbrace{\sum_{x\in\Omega'} x\cdot\mathrm{P}_X(x)}_{\text{diskrete } X:\Omega\to\Omega'} \;\hat{=}\; \underbrace{\int_{\mathbb{R}} x\cdot f_X(x)\,\mathrm{d}x}_{\text{stetige } X:\Omega\to\mathbb{R}}}$$

$\mathrm{E}[\alpha X + \beta Y] = \alpha\,\mathrm{E}[X] + \beta\,\mathrm{E}[Y]$ $\qquad X \leq Y \Rightarrow \mathrm{E}[X] \leq \mathrm{E}[Y]$
$\mathrm{E}[X^2] = \mathrm{Var}[X] + \mathrm{E}[X]^2$
$\mathrm{E}[X\,Y] = \mathrm{E}[X]\,\mathrm{E}[Y]$, falls $X$ und $Y$ stochastisch unabhängig
Umkehrung nicht möglichich: Unkorreliertheit $\not\Rightarrow$ Stoch. Unabhängig!

**4.1.1 Für Funktionen von Zufallsvariablen** $g(x)$
$\mathrm{E}[g(X)] = \sum\limits_{x\in\Omega'} g(x)\,\mathrm{P}_X(x) \;\hat{=}\; \int\limits_{\mathbb{R}} g(x)f_X(x)\,\mathrm{d}x$

### 4.2. Varianz (2. zentrales Moment)

ist ein Maß für die Stärke der Abweichung vom Erwartungswert

$$\boxed{\sigma_X^2 = \mathrm{Var}[X] = \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{E}[X^2] - \mathrm{E}[X]^2}$$

$\mathrm{Var}[\alpha X + \beta] = \alpha^2\,\mathrm{Var}[X]$ $\qquad \mathrm{Var}[X] = \mathrm{Cov}[X,X]$
$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n}\mathrm{Var}[X_i] + \sum_{j\neq i}\mathrm{Cov}[X_i, X_j]$$
**Standard Abweichung:** $\sigma = \sqrt{\mathrm{Var}[X]}$

### 4.3. Kovarianz

Maß für den linearen Zusammenhang zweier Variablen

$$\boxed{\begin{aligned}\mathrm{Cov}[X,Y] &= \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])^\top] =\\ &= \mathrm{E}[X\,Y^\top] - \mathrm{E}[X]\,\mathrm{E}[Y]^\top = \mathrm{Cov}[Y,X]\end{aligned}}$$

$\mathrm{Cov}[\alpha X + \beta, \gamma Y + \delta] = \alpha\gamma\,\mathrm{Cov}[X,Y]$
$\mathrm{Cov}[X+U, Y+V] = \mathrm{Cov}[X,Y] + \mathrm{Cov}[X,V] + \mathrm{Cov}[U,Y] + \mathrm{Cov}[U,V]$

**4.3.1 Korrelation = standardisierte Kovarianz**
$\rho(X,Y) = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\cdot\mathrm{Var}[Y]}} = \frac{C_{x,y}}{\sigma_x\cdot\sigma_y}$ $\qquad \rho(X,Y) \in [-1;1]$

**4.3.2 Kovarianzmatrix für** $\underline{z} = (\underline{x},\underline{y})^\top$
$$\mathrm{Cov}[\underline{z}] = \boldsymbol{C}_{\underline{z}} = \begin{bmatrix}C_X & C_{XY}\\ C_{XY} & C_Y\end{bmatrix} = \begin{bmatrix}\mathrm{Cov}[X,X] & \mathrm{Cov}[X,Y]\\ \mathrm{Cov}[Y,X] & \mathrm{Cov}[Y,Y]\end{bmatrix}$$
Immer symmetrisch: $C_{xy} = C_{yx}$! Für Matrizen: $\boldsymbol{C}_{\underline{xy}} = \boldsymbol{C}_{\underline{yx}}^\top$

# 5. Estimation

## 5.1. Estimation

Statistic Estimation treats the problem of inferring underlying characteristics of unknown random variables on the basis of observations of outputs of those random variables.

| | |
|---|---|
| Sample Space $\Omega$ | nonempty set of outputs of experiment |
| Sigma Algebra $\mathbb{F} \subseteq 2^{\Omega}$ | set of subsets of outputs (events) |
| Probability $\mathsf{P} : \mathbb{F} \mapsto [0,1]$ | |
| Random Variable $X : \Omega \mapsto \mathbb{X}$ | mapped subsets of $\Omega$ |
| Observations: $x_1, \dots, x_N$ | single values of $X$ |
| Observation Space $\mathbb{X}$ | possible observations of $X$ |
| Unknown parameter $\theta \in \Theta$ | parameter of propability function |
| Estimator $T : \mathbb{X} \mapsto \Theta$ | $T(X) = \hat{\theta}$, finds $\hat{\theta}$ from $X$ |

| | |
|---|---|
| unknown parm. $\theta$ | estimation of param. $\hat{\theta}$ |
| R.V. of param. $\Theta$ | estim. of R.V. of parm $T(X) = \hat{\Theta}$ |

## 5.2. Quality Properties of Estimators

Consistent: If $\lim\limits_{N\to\infty} T(x_1, \dots, x_N) = \theta$

Bias $\mathrm{Bias}(T) := \mathsf{E}[T(X_1, \dots, X_N)] - \theta$

unbiased if $\mathrm{Bias}(T) = 0$ (biased estimators can provide better estimates than unbiased estimators.)

Variance $\mathsf{Var}[T] := \mathsf{E}\left[(T - \mathsf{E}[T])^2\right]$

## 5.3. Mean Square Error (MSE)

The MSE is an extension of the Variance $\mathsf{Var}[T] := \mathsf{E}\left[(T - \mathsf{E}[T])^2\right]$:

$$\text{MSE: } \varepsilon[T] = \underbrace{\mathsf{E}\left[(T-\theta)^2\right]}_{=\mathsf{E}[(\hat{\theta}-\theta)^2]} = \mathsf{Var}(T) + (\mathrm{Bias}[T])^2$$

If $\Theta$ is also r.v. $\Rightarrow$ mean over both (e.g. Bayes est.):

**Mean MSE:** $\mathsf{E}[(T(X) - \Theta)^2] = \mathsf{E}\left[\mathsf{E}\left[(T(X) - \Theta)^2 | \Theta = \theta\right]\right]$

### 5.3.1 Minimum Mean Square Error (MMSE)

Minimizes mean square error: $\arg\min\limits_{\hat{\theta}} \mathsf{E}\left[(\hat{\theta} - \theta)^2\right]$

$\mathsf{E}\left[(\hat{\theta} - \theta)^2\right] = \mathsf{E}[\theta^2] - 2\hat{\theta}\,\mathsf{E}[\theta] + \hat{\theta}^2$

Solution: $\frac{\mathrm{d}}{\mathrm{d}\hat{\theta}} \mathsf{E}\left[(\hat{\theta} - \theta)^2\right] \stackrel{!}{=} 0 = -2\,\mathsf{E}[\theta] + 2\hat{\theta} \quad \Rightarrow \hat{\theta}_{\mathsf{MMSE}} = \mathsf{E}[\theta]$

## 5.4. Maximum Likelihood

Given model $\{\mathbb{X}, \mathbb{F}, \mathsf{P}_\theta; \theta \in \Theta\}$, assume $\mathsf{P}_\theta(\underline{x})$ or $f_{\underline{X}}(\underline{x}, \theta)$ for observed data $\underline{x}$. Estimate parameter $\theta$ so that the likelihood $L(\underline{x}, \theta)$ or $L(\theta | X = \underline{x})$ to obtain $\underline{x}$ is maximized.

**Likelihood Function:** (Prob. for $\theta$ given $\underline{x}$)

Discrete: $\quad L(x_1, \dots, x_N; \theta) = \mathsf{P}_\theta(x_1, \dots, x_N)$

Continuous: $\quad L(x_1, \dots, x_N; \theta) = f_{X_1, \dots, X_N}(x_1, \dots, x_N, \theta)$

If $N$ observations are Identically Independently Distributed (i.i.d.):

$L(\underline{x}, \theta) = \prod\limits_{i=1}^{N} \mathsf{P}_\theta(x_i) = \prod\limits_{i=1}^{N} f_{X_i}(x_i)$

**ML Estimator** (Picks $\theta$): $T_{\mathsf{ML}} : X \mapsto \underset{\theta\in\Theta}{\operatorname{argmax}}\{L(X, \theta)\} =$

$= \underset{\theta\in\Theta}{\operatorname{argmax}}\{\log L(\underline{X}, \theta)\} \stackrel{\text{i.i.d.}}{=} \underset{\theta\in\Theta}{\operatorname{argmax}}\{\sum \log L(x_i, \theta)\}$

Find Maximum: $\frac{\partial L(\underline{x}, \theta)}{\partial \theta} = \frac{\mathrm{d}}{\mathrm{d}\theta} \log L(x; \theta)\Big|_{\theta = \hat{\theta}} \stackrel{!}{=} 0$

Solve for $\theta$ to obtain ML estimator function $\hat{\theta}_{\mathsf{ML}}$

Check quality of estimator with MSE

Maximum-Likelihood Estimator is Asymptotically Efficient. However, there might be not enough samples and the likelihood function is often not known.

## 5.5. Uniformly Minimum Variance Unbiased (UMVU) Estimators (Best unbiased estimators)

Best unbiased estimator: Lowest Variance of all estimators.

Fisher's Information Inequality: Estimate lower bound of variance if

- $L(x, \theta) > 0, \forall x, \theta$
- $L(x, \theta)$ is diffable for $\theta$
- $\int_{\mathbb{X}} \frac{\partial}{\partial\theta} L(x, \theta)\,\mathrm{d}x = \frac{\partial}{\partial\theta} \int_{\mathbb{X}} L(x, \theta)\,\mathrm{d}x$

**Score Function:**

$g(x, \theta) = \frac{\partial}{\partial\theta} \log L(x, \theta) = \frac{\frac{\partial}{\partial\theta} L(x, \theta)}{L(x, \theta)} \qquad \mathsf{E}[g(x, \theta)] = 0$

**Fischer Information:**

$I_{\mathsf{F}}(\theta) := \mathsf{Var}[g(X, \theta)] = \mathsf{E}[g(x, \theta)^2] = -\,\mathsf{E}\left[\frac{\partial^2}{\partial\theta^2} \log L(X, \theta)\right]$

**Cramér-Rao Lower Bound (CRB):** (if $T$ is unbiased)

$$\mathsf{Var}[T(X)] \geq \left(\frac{\partial\,\mathsf{E}[T(X)]}{\partial\theta}\right)^2 \frac{1}{I_{\mathsf{F}}(\theta)} \qquad \mathsf{Var}[T(X)] \geq \frac{1}{I_{\mathsf{F}}(\theta)}$$

For $N$ i.i.d. observations: $I_{\mathsf{F}}^{(N)}(x, \theta) = N \cdot I_{\mathsf{F}}^{(1)}(x, \theta)$

### 5.5.1 Exponential Models

If $f_X(x) = \frac{h(x)\exp\left(a(\theta)t(x)\right)}{\exp(b(\theta))}$ then $I_F(\theta) = \frac{\partial a(\theta)}{\partial\theta} \frac{\partial E[t(X)]}{\partial\theta}$

**Some Derivations:** (check in exam)

Uniformly: Not diffable $\Rightarrow$ no $I_F(\theta)$

Normal $\mathcal{N}(\theta, \sigma^2)$: $g(x, \theta) = \frac{(x-\theta)}{\sigma^2} \qquad I_{\mathsf{F}}(\theta) = \frac{1}{\sigma^2}$

Binomial $\mathcal{B}(\theta, K)$: $g(x, \theta) = \frac{x}{\theta} - \frac{K-x}{1-\theta} \qquad I_{\mathsf{F}}(\theta) = \frac{K}{\theta(1-\theta)}$

## 5.6. Bayes Estimation (Conditional Mean)

A Priori information about $\theta$ is known as probability $f_\Theta(\theta; \sigma)$ with random variable $\Theta$ and parameter $\sigma$. Now the conditional pdf $f_{X|\Theta}(x, \theta)$ is used to find $\theta$ by minimizing the mean MSE instead of uniformly MSE.

Mean MSE for $\Theta$: $\mathsf{E}\left[\mathsf{E}[(T(X) - \Theta)^2 | \Theta = \theta]\right]$

**Conditional Mean Estimator:**

$T_{\mathsf{CM}} : x \mapsto \mathsf{E}[\Theta | X = x] = \int_\Theta \theta \cdot f_{\Theta|X}(\theta|x)\,\mathrm{d}\theta$

Posterior $f_{\Theta|\underline{X}}(\theta|\underline{x}) = \frac{f_{\underline{X}|\Theta}(\underline{x})f_\Theta(\theta)}{\int_\Theta f_{\underline{X},\xi}(\underline{x}, \xi)\,\mathrm{d}\xi} = \frac{f_{\underline{X}|\theta}(\underline{x})f_\Theta(\theta)}{f_X(x)}$

**Hint:** to calculate $f_{\Theta|\underline{X}}(\theta|\underline{x})$: Replace every factor not containing $\theta$, such as $\frac{1}{f_X(x)}$ with a factor $\gamma$ and determine $\gamma$ at the end such that $\int_\Theta f_{\Theta|\underline{X}}(\theta|\underline{x})\,\mathrm{d}\theta = 1$

MMSE: $\mathsf{E}[\mathsf{Var}[X | \Theta = \theta]]$

**Multivariate Gaussian:** $X, \Theta \sim \mathcal{N} \quad \Rightarrow \sigma_X^2 = \sigma_{X|\Theta=\theta}^2 + \sigma_\Theta$

$T_{\mathsf{CM}} : x \mapsto \mathsf{E}[\Theta | X = x] = \underline{\boldsymbol{\mu}}_\Theta + \underset{\sim}{\boldsymbol{C}}_{\Theta,X} \underset{\sim}{\boldsymbol{C}}_X^{-1} (\underline{x} - \underline{\boldsymbol{\mu}}_X)$

MMSE:

$\mathsf{E}\left[\| T_{\mathsf{CM}} - \Theta \|_2^2\right] = \operatorname{tr}(\underset{\sim}{\boldsymbol{C}}_{\theta|X}) = \operatorname{tr}(\underset{\sim}{\boldsymbol{C}}_\Theta - \underset{\sim}{\boldsymbol{C}}_{\Theta,X} \underset{\sim}{\boldsymbol{C}}_X^{-1} \underset{\sim}{\boldsymbol{C}}_{X,\Theta})$

**Orthogonality Principle:**

$T_{\mathsf{CM}}(\underline{X}) - \Theta \perp h(\underline{X}) \quad \Rightarrow \quad \mathsf{E}[(T_{\mathsf{CM}}(\underline{X}) - \Theta)h(\underline{X})] = 0$

**MMSE Estimator:** $\hat{\theta}_{\mathsf{MMSE}} = \underset{\theta\in\Theta}{\arg\min}\ \mathsf{MSE}$

minimizes the MSE for all estimators

## 5.7. Example:

Estimate mean $\theta$ of $X$ with prior knowledge $\theta \in \Theta \sim \mathcal{N}$:

$X \sim \mathcal{N}(\theta, \sigma_{X|\Theta=\theta}^2)$ and $\Theta \sim \mathcal{N}(m, \sigma_\Theta^2)$

$\hat{\theta}_{\mathsf{CM}} = \mathsf{E}[\Theta | \underline{X} = \underline{x}] = \frac{N\sigma_\Theta^2}{\sigma_{X|\Theta=\theta}^2 + N\sigma_\Theta^2}\hat{\theta}_{\mathsf{ML}} + \frac{\sigma_{X|\Theta=\theta}^2}{\sigma_{X|\Theta=\theta}^2 + N\sigma_\Theta^2} m$

For $N$ independent observations $x_i$: $\hat{\theta}_{\mathsf{ML}} = \frac{1}{N}\sum x_i$

Large $N \Rightarrow$ ML better, small $N \Rightarrow$ CM better

# 6. Linear Estimation

$t$ is now the unknown parameter $\theta$, we want to estimate $y$ and $\underline{x}$ is the input vector... review regression problem $\underline{y} = \underset{\sim}{\boldsymbol{A}}\underline{x}$ (we solve for $\underline{x}$), here we solve for $\underline{t}$, because $\underline{x}$ is known (measured)! Confusing...

1. Training $\to$ 2. Estimation

Training: We observe $y$ and $\underline{x}$ (knowing both) and then based on that we try to estimate $y$ given $\underline{x}$ (only observe $\underline{x}$) with a linear model $\hat{y} = \underline{x}^\top \underline{t}$

$$\text{Estimation: } \hat{y} = \underline{x}^\top \underline{t} + m \quad \text{or} \quad \hat{y} = \underline{x}^\top \underline{t}$$

Given: $N$ observations $(y_i, \underline{x}_i)$, unknown parameters $\underline{t}$, noise $m$

$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad \underset{\sim}{\boldsymbol{X}} = \begin{bmatrix} \underline{x}_1^\top \\ \vdots \\ \underline{x}_n^\top \end{bmatrix} \qquad$ Note: $\hat{y} \neq y!!$

Problem: Estimate $y$ based on given (known) observations $\underline{x}$ and unknown parameter $\underline{t}$ with assumed linear Model: $\hat{y} = \underline{x}^\top \underline{t}$

Note $y = \underline{x}^\top \underline{t} + m \to y = \underline{x}'^\top \underline{t}'$ with $\underline{x}' = \begin{pmatrix} \underline{x} \\ 1 \end{pmatrix}, \quad t' = \begin{pmatrix} \underline{t} \\ m \end{pmatrix}$

Sometimes in Exams: $\hat{y} = \underline{x}^\top \underline{t} \Leftrightarrow \hat{\underline{x}} = \underset{\sim}{\boldsymbol{T}}^\top \underline{y}$

estimate $\underline{x}$ given $\underline{y}$ and unknown $\underset{\sim}{\boldsymbol{T}}$

## 6.1. Least Square Estimation (LSE)

Tries to minimize the square error for linear Model: $\hat{y}_{\mathsf{LS}} = \underline{x}^\top \underline{t}_{\mathsf{LS}}$

Least Square Error: $\min \left[\sum\limits_{i=1}^{N} (y_i - \underline{x}_i^\top \underline{t})^2\right] = \min\limits_{\underline{t}} \left\|\underline{y} - \underset{\sim}{\boldsymbol{X}}\underline{t}\right\|$

$$\underline{t}_{\mathsf{LS}} = (\underset{\sim}{\boldsymbol{X}}^\top \underset{\sim}{\boldsymbol{X}})^{-1} \underset{\sim}{\boldsymbol{X}}^\top \underline{y}$$

$\hat{\underline{y}}_{\mathsf{LS}} = \underset{\sim}{\boldsymbol{X}} \underline{t}_{\mathsf{LS}} \in span(X)$

**Orthogonality Principle:** $N$ observations $\underline{x}_i \in \mathbb{R}^d$

$\underset{\sim}{\boldsymbol{Y}} - \underset{\sim}{\boldsymbol{X}}\underset{\sim}{\boldsymbol{T}}_{\mathsf{LS}} \perp \operatorname{span}[\underset{\sim}{\boldsymbol{X}}] \Leftrightarrow \underset{\sim}{\boldsymbol{Y}} - \underset{\sim}{\boldsymbol{X}}\underset{\sim}{\boldsymbol{T}}_{\mathsf{LS}} \in \operatorname{null}[\underset{\sim}{\boldsymbol{X}}^\top]$, thus

$\underset{\sim}{\boldsymbol{X}}^\top(\underset{\sim}{\boldsymbol{Y}} - \underset{\sim}{\boldsymbol{X}}\underset{\sim}{\boldsymbol{T}}_{\mathsf{LS}}) = 0$ and if $N > d \wedge \operatorname{rang}[\underset{\sim}{\boldsymbol{X}}] = d$:

$\underset{\sim}{\boldsymbol{T}}_{\mathsf{LS}} = (\underset{\sim}{\boldsymbol{X}}^\top \underset{\sim}{\boldsymbol{X}})^{-1} \underset{\sim}{\boldsymbol{X}}^\top \underset{\sim}{\boldsymbol{Y}}$

## 6.2. Linear Minimum Mean Square Estimator (LMMSE)

Estimate $y$ with linear estimator $\underline{t}$, such that $\hat{y} = \underline{t}^\top \underline{x} + m$

Note: the Model does not need to be linear! The estimator is linear!

$$\hat{y}_{\mathsf{LMMSE}} = \underset{t,m}{\arg\min}\,\mathsf{E}\left[\left\|\underline{y} - (\underline{t}^\top \underline{x} + m)\right\|_2^2\right]$$

If Random joint variable $\underline{z} = \begin{pmatrix} \underline{x} \\ y \end{pmatrix}$ with

$\underline{\boldsymbol{\mu}}_{\underline{z}} = \begin{pmatrix} \underline{\boldsymbol{\mu}}_{\underline{x}} \\ \mu_y \end{pmatrix}$ and $\underset{\sim}{\boldsymbol{C}}_{\underline{z}} = \begin{bmatrix} \underset{\sim}{\boldsymbol{C}}_{\underline{x}} & \underline{\boldsymbol{c}}_{\underline{x}y} \\ c_{y\underline{x}} & c_y \end{bmatrix}$ then

LMMSE Estimation of $y$ given $x$ is

$\hat{y} = \mu_y + \underline{\boldsymbol{c}}_{y\underline{x}} \underset{\sim}{\boldsymbol{C}}_{\underline{x}}^{-1} (\underline{x} - \underline{\boldsymbol{\mu}}_{\underline{x}}) = \underbrace{\underline{\boldsymbol{c}}_{y\underline{x}} \underset{\sim}{\boldsymbol{C}}_{\underline{x}}^{-1}}_{=\underline{t}^\top} \underline{x} \underbrace{- \mu_y + \underline{\boldsymbol{c}}_{y\underline{x}} \underset{\sim}{\boldsymbol{C}}_{\underline{x}}^{-1} \underline{\boldsymbol{\mu}}_{\underline{x}}}_{=m}$

Minimum MSE: $\mathsf{E}\left[\left\|\underline{y} - (\underline{x}^\top \underline{t} + m)\right\|_2^2\right] = c_y - c_{y\underline{x}} C_{\underline{x}}^{-1} \underline{\boldsymbol{c}}_{\underline{x}y}$

**Hint:** First calculate $\hat{y}$ in general and then set variables according to system equation.

**Multivariate:** $\hat{\underline{y}} = \underset{\sim}{\boldsymbol{T}}_{\mathsf{LMMSE}}^\top \underline{x} \qquad \underset{\sim}{\boldsymbol{T}}_{\mathsf{LMMSE}}^\top = \underset{\sim}{\boldsymbol{C}}_{\underline{\mathbf{y}}\underline{\mathbf{x}}} \underset{\sim}{\boldsymbol{C}}_{\underline{\mathbf{x}}}^{-1}$

---

If $\underline{\boldsymbol{\mu}}_{\underline{z}} = \underline{\boldsymbol{0}}$ then

Estimator $\hat{y} = \underline{\boldsymbol{c}}_{y,\underline{x}} \underset{\sim}{\boldsymbol{C}}_{\underline{x}}^{-1} \underline{x}$

Minimum MSE: $\mathsf{E}[c_{y,\underline{x}}] = c_y - \underline{t}^\top \underline{\boldsymbol{c}}_{\underline{x},y}$

## 6.3. Matched Filter Estimator (MF)

For channel $\underline{y} = \underline{h}x + \underline{v}$, Filtered: $\underline{t}^\top \underline{y} = \underline{t}^\top \underline{h}x + \underline{t}^\top \underline{v}$

Find Filter $\underline{t}^\top$ that maximizes $\mathsf{SNR} = \frac{\|\underline{h}x\|}{\|\underline{v}\|}$

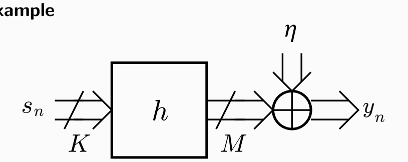
$\underset{\sim}{\boldsymbol{t}}_{\mathsf{MF}} = \max\limits_{t} \left\{ \frac{\mathsf{E}\left[(\underline{t}^\top \underline{h}x)^2\right]}{\mathsf{E}\left[(\underline{t}^\top \underline{v})^2\right]} \right\}$

In the lecture (estimate $\underline{h}$):

$\underset{\sim}{\boldsymbol{T}}_{\mathsf{MF}} = \max\limits_{T} \left\{ \frac{\left|\mathsf{E}\left[\hat{\underline{h}}^H \underline{h}\right]\right|^2}{\operatorname{tr}\left[\mathsf{Var}[\underset{\sim}{\boldsymbol{T}}\underline{n}]\right]} \right\}$

$\hat{\underline{h}}_{\mathsf{MF}} = \underset{\sim}{\boldsymbol{T}}_{\mathsf{MF}} \underline{y} \qquad \underset{\sim}{\boldsymbol{T}}_{\mathsf{MF}} \propto \underset{\sim}{\boldsymbol{C}}_{\underline{h}} \underset{\sim}{\boldsymbol{S}}^H \underset{\sim}{\boldsymbol{C}}_{\underline{n}}^{-1}$

## 6.4. Example




System Model: $\underline{y}_n = \underset{\sim}{\boldsymbol{H}} \underline{s}_n + \eta_n$

with $\underset{\sim}{\boldsymbol{H}} = (h_{m,k}) \in \mathbb{C}^{M\times K} \qquad (m \in [1, M], k \in [1, K])$

**Linear Channel Model** $\underline{y} = \underset{\sim}{\boldsymbol{S}} \underline{h} + \underline{n}$ with $\underline{h} \sim \mathcal{N}(0, \underset{\sim}{\boldsymbol{C}}_{\underline{h}})$ and $\underline{n} \sim \mathcal{N}(0, \underset{\sim}{\boldsymbol{C}}_{\underline{n}})$

Linear Estimator $\underset{\sim}{\boldsymbol{T}}$ estimates $\hat{\underline{h}} = \underset{\sim}{\boldsymbol{T}} \underline{y} \in \mathbb{C}^{MK}$

$\underset{\sim}{\boldsymbol{T}}_{\mathsf{MMSE}} = \underset{\sim}{\boldsymbol{C}}_{\underline{h}\underline{y}} \underset{\sim}{\boldsymbol{C}}_{\underline{y}}^{-1} = \underset{\sim}{\boldsymbol{C}}_{\underline{h}} \underset{\sim}{\boldsymbol{S}}^{\mathsf{H}} (\underset{\sim}{\boldsymbol{S}} \underset{\sim}{\boldsymbol{C}}_{\underline{h}} \underset{\sim}{\boldsymbol{S}}^{\mathsf{H}} + \underset{\sim}{\boldsymbol{C}}_{\underline{n}})^{-1}$

$\underset{\sim}{\boldsymbol{T}}_{\mathsf{ML}} = \underset{\sim}{\boldsymbol{T}}_{\mathsf{Cor}} = (\underset{\sim}{\boldsymbol{S}}^{\mathsf{H}} \underset{\sim}{\boldsymbol{C}}_{\underline{n}}^{-1} \underset{\sim}{\boldsymbol{S}})^{-1} \underset{\sim}{\boldsymbol{S}}^{\mathsf{H}} \underset{\sim}{\boldsymbol{C}}_{\underline{n}}^{-1}$

$\underset{\sim}{\boldsymbol{T}}_{\mathsf{MF}} \propto \underset{\sim}{\boldsymbol{C}}_{\underline{h}} \underset{\sim}{\boldsymbol{S}}^{\mathsf{H}} \underset{\sim}{\boldsymbol{C}}_{\underline{n}}^{-1}$

For Assumption $\underset{\sim}{\boldsymbol{S}}^{\mathsf{H}} \underset{\sim}{\boldsymbol{S}} = N\sigma_s^2 \underset{\sim}{\boldsymbol{1}}_{K\times M}$ and $\underset{\sim}{\boldsymbol{C}}_{\underline{n}} = \sigma_\eta^2 \underset{\sim}{\boldsymbol{1}}_{N\times M}$

| Estimator | Averaged Squared Bias | Variance |
|---|---|---|
| ML/Correlator | 0 | $KM\,\frac{\sigma_\eta^2}{N\sigma_s^2}$ |
| Matched Filter | $\sum\limits_{i=1}^{KM} \lambda_i \left(\frac{\lambda_i}{\lambda_1} - 1\right)^2$ | $\sum\limits_{i=1}^{KM} \left(\frac{\lambda_i}{\lambda_1}\right)^2 \frac{\sigma_\eta^2}{N\sigma_s^2}$ |
| MMSE | $\sum\limits_{i=1}^{KM} \lambda_i \left(\frac{1}{1 + \frac{\sigma_\eta^2}{\lambda_i N\sigma_s^2}} - 1\right)^2$ | $\sum\limits_{i=1}^{KM} \frac{1}{\left(1 + \frac{\sigma_\eta^2}{\lambda_i N\sigma_s^2}\right)^2} \frac{\sigma_\eta^2}{N\sigma_s^2}$ |

## 6.5. Estimators

Upper Bound: Uniform in $[0; \theta]$ : $\hat{\theta}_{\mathsf{ML}} = \frac{2}{N}\sum x_i$

Probability $p$ for $\mathcal{B}(p, N)$: $\hat{p}_{\mathsf{ML}} = \frac{x}{N} \qquad \hat{p}_{\mathsf{CM}} = \frac{x+1}{N+2}$

Mean $\mu$ for $\mathcal{N}(\mu, \sigma^2)$ : $\hat{\mu}_{\mathsf{ML}}^2 = \frac{1}{N}\sum\limits_{i=1}^{N} x_i$

Variance $\sigma^2$ for $\mathcal{N}(\mu, \sigma^2)$ : $\hat{\sigma}_{\mathsf{ML}}^2 = \frac{1}{N}\sum\limits_{i=1}^{N} (x_i - \mu)^2$

# 7. Gaussian Stuff

### 7.1. Gaussian Channel

Channel: $Y = hs_i + N$ with $h \sim \mathcal{N}, N \sim \mathcal{N}$

$L(y_1, ..., y_N) = \prod_{i=1}^{n} f_{Y_i}(y_i, h)$

$f_{Y_i}(y_i, h) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - hs_i)^2\right)$

$\hat{h}_{ML} = \underset{h}{\operatorname{argmin}}\{\left\|\underline{y} - h\underline{s}\right\|^2\} = \frac{\underline{s}^\top \underline{y}}{\underline{s}^\top \underline{s}}$

If multidimensional channel: $\underline{y} = \underline{S}\underline{h} + \underline{n}$:

$L(\underline{y}, \underline{h}) = \frac{1}{\sqrt{\det(2\pi\underline{C})}} \exp\left(-\frac{1}{2}(\underline{y} - \underline{S}\underline{h})^\top \underline{C}^{-1}(\underline{y} - \underline{S}\underline{h})\right)$

$l(\underline{y}, \underline{h}) = \frac{1}{2}\left(\log(\det(2\pi\underline{C}) - (\underline{y} - \underline{S}\underline{h})^\top \underline{C}^{-1}(\underline{y} - \underline{S}\underline{h})\right)$

$\frac{\mathrm{d}}{\mathrm{d}h}(\underline{y} - \underline{S}\underline{h})^\top \underline{C}^{-1}(\underline{y} - \underline{S}\underline{h}) = -2\underline{S}^\top \underline{C}^{-1}(\underline{x} - \underline{S}\underline{h})$

**Gaussian Covariance:** if $Y \sim \mathcal{N}(0, \sigma^2)$, $N \sim \mathcal{N}(0, \sigma^2)$:

$\underline{C}_Y = \mathrm{Cov}[Y, Y] = \mathrm{E}[(Y - \mu)(Y - \mu)^\top] = \mathrm{E}[Y Y^\top]$

For Channel $Y = Sh + N$: $\mathrm{E}[Y Y^\top] = S\,\mathrm{E}[hh^\top]S^\top + \mathrm{E}[NN^\top]$

### 7.2. Multivariate Gaussian Distributions

A vector $\underline{x}$ of $n$ independent Gaussian random variables $x_i$ is jointly Gaussian. If $\underline{x} \sim \mathcal{N}(\underline{\mu}_{\underline{x}}, \underline{C}_{\underline{x}})$:

$$f_{\underline{x}}(\underline{x}) = f_{x_1,...,x_n}(x_1, ..., x_n) =$$
$$= \frac{1}{\sqrt{\det(2\pi\underline{C}_{\underline{x}})}} \exp\left(-\frac{1}{2}\left(\underline{x} - \underline{\mu}_{\underline{x}}\right)^\top \underline{C}_{\underline{x}}^{-1}\left(\underline{x} - \underline{\mu}_{\underline{x}}\right)\right)$$

Affine transformations $\underline{y} = \underline{A}\underline{x} + \underline{b}$ are jointly Gaussian with

$\underline{y} \sim \mathcal{N}(\underline{A}\underline{\mu}_{\underline{x}} + \underline{b}, \underline{A}\underline{C}_{\underline{x}}\underline{A}^\top)$

All marginal PDFs are Gaussian as well

**Contour Lines**

Ellipsoid with central point $\mathrm{E}[\underline{y}]$ and main axis are the eigenvectors of $\underline{C}_{\underline{y}}^{-1}$

### 7.3. Conditional Gaussian

$\underline{A} \sim \mathcal{N}(\underline{\mu}_A, \underline{C}_A)$, $\underline{B} \sim \mathcal{N}(\underline{\mu}_B, \underline{C}_B)$

$\Rightarrow (\underline{A}|\underline{B} = b) \sim \mathcal{N}(\underline{\mu}_{A|B}, \underline{C}_{A|B})$

**Conditional Mean:**

$\mathrm{E}[\underline{A}|\underline{B} = \underline{b}] = \underline{\mu}_{A|B=b} = \underline{\mu}_A + \underline{C}_{AB}\,\underline{C}_{BB}^{-1}\left(\underline{b} - \underline{\mu}_B\right)$

**Conditional Variance:**

$\underline{C}_{A|B} = \underline{C}_{AA} + \underline{C}_{AB}\,\underline{C}_{BB}^{-1}\,\underline{C}_{BA}$

### 7.4. Misc

If CDF of gaussian distribution given $\Phi(z) \sim \mathcal{N}(0, 1)$ then for $X \sim \mathcal{N}(1, 1)$ the CDF is given as $\Phi(x - \mu_x)$

# 8. Sequences

### 8.1. Random Sequences

Sequence of a random variable. Example: result of a dice is RV, roll a dice several times is a random sequence.

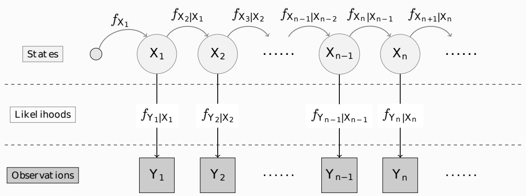
### 8.2. Markov Sequence $X_n : \Omega \to X_n$

Sequence of memoryless state transitions with certain probabilities.

1. state: $f_{X_1}(x_1)$
2. state: $f_{X_2 \mid X_1}(x_2|x_1)$
n. state: $f_{X_n \mid X_{n-1}}(x_n|x_{n-1})$

### 8.3. Hidden Markov Chains

Problem: states $X_i$ are not visible and can only be guessed indirectly as a random variable $Y_i$.




Conditional pdf $f_{\underline{X}_n|\underline{Y}_n}$ Likelihood pdf $f_{Y_n \mid X_n}$

State-transision pdf $f_{X_n \mid X_{n-1}}$

**Estimation:**

$$f_{\underline{X}_n|\underline{Y}_n} \propto f_{\underline{Y}_n|\underline{X}_n} \cdot \int_{\mathbb{X}} f_{\underline{X}_n|\underline{X}_{n-1}} \cdot f_{\underline{X}_{n-1}|\underline{Y}_{n-1}}\,\mathrm{d}\underline{x}_{n-1}$$

# 9. Recursive Estimation

### 9.1. Kalman-Filter

recursively calculates the most likely state from previous state estimates and current observation. Shows optimum performance for Gauss-Markov Sequences.

**State space:**

$\underline{x}_n = \underline{G}_n\underline{x}_{n-1} + \underline{B}\underline{u}_n + \underline{v}_n$
$\underline{y}_n = \underline{H}_n\underline{x}_n + \underline{w}_n$

With gaussian process/measurement noise $\underline{v}_n / \underline{w}_n$

Short notation: $\mathrm{E}[\underline{x}_n|\underline{y}_{n-1}] = \hat{\underline{x}}_{n|n-1}$ $\quad \mathrm{E}[\underline{x}_n|\underline{y}_n] = \hat{\underline{x}}_{n|n}$

$\mathrm{E}[\underline{y}_n|\underline{y}_{n-1}] = \hat{\underline{y}}_{n|n-1}$ $\quad \mathrm{E}[\underline{y}_n|\underline{y}_n] = \hat{\underline{y}}_{n|n}$

**1. step: Prediction**

Mean: $\hat{\underline{x}}_{n|n-1} = \underline{G}_n\hat{\underline{x}}_{n-1|n-1}$

Covariance: $\underline{C}_{\underline{x}_{n|n-1}} = \underline{G}_n\underline{C}_{\underline{x}_{n-1|n-1}}\underline{G}_n^\top + \underline{C}_{\underline{v}}$

**2. step: Update**

Mean: $\hat{\underline{x}}_{n|n} = \hat{\underline{x}}_{n|n-1} + \underline{K}_n\left(\underline{y}_n - \underline{H}_n\hat{\underline{x}}_{n|n-1}\right)$

Covariance: $\underline{C}_{\underline{x}_{n|n}} = \underline{C}_{\underline{x}_{n|n-1}} + \underline{K}_n\underline{H}_n\underline{C}_{\underline{x}_{n|n-1}}$

$$\hat{\underline{x}}_{n|n} = \underbrace{\hat{\underline{x}}_{n|n-1}}_{\text{estimation } \mathrm{E}[X_n \mid Y_{n-1} = y_{n-1}]} + \overbrace{\underline{K}_n\underbrace{\left(\underline{y}_n - \underline{H}_n\hat{\underline{x}}_{n|n-1}\right)}_{\text{innovation: } \Delta y_n}}^{\text{correction: } \mathrm{E}[X_n \mid \Delta Y_n = y_n]}$$

With optimal **Kalman-gain** (prediction for $\underline{x}_n$ based on $\Delta y_n$):

$$\underline{K}_n = \underline{C}_{\underline{x}_{n|n-1}}\underline{H}_n^\top\underbrace{(\underline{H}_n\underline{C}_{\underline{x}_{n|n-1}}\underline{H}_n^\top + \underline{C}_{\underline{w}_n})}_{\underline{C}_{\delta y_n}}{}^{-1}$$

**Innovation**: closeness of the estimated mean value to the real value

$\Delta\underline{y}_n = \underline{y}_n - \hat{\underline{y}}_{n|n-1} = \underline{y}_n - \underline{H}_n\hat{\underline{x}}_{n|n-1}$

**Init:** $\hat{\underline{x}}_{0|-1} = \mathrm{E}[X_0]$ $\quad \sigma^2_{0|-1} = \mathrm{Var}[X_0]$

**MMSE Estimator:** $\hat{\underline{x}} = \int \underline{x}_n f_{X_n \mid Y_{(n)}}(\underline{x}_n|\underline{y}_{(n)})\,\mathrm{d}\underline{x}_n$

For non linear problems: Suboptimum nonlinear Filters: Extended KF, Unscented KF, ParticleFilter

### 9.2. Extended Kalman (EKF)

Linear approximation of non-linear $g, h$

$\underline{x}_n = g_n(\underline{x}_{n-1}, \underline{v}_n)$ $\quad \underline{v}_n \sim \mathcal{N}$
$\underline{y}_n = h_n(\underline{x}_{n-1}, \underline{w}_n)$ $\quad \underline{w}_n \sim \mathcal{N}$

### 9.3. Unscented Kalman (UKF)

Approximation of desired PDF $f_{X_n|Y_n}(x_n|y_n)$ by Gaussian PDF.

### 9.4. Particle-Filter

For non linear state space *and* non-gaussian noise

**Non-linear State space:**

$\underline{x}_n = g_n(\underline{x}_{n-1}, \underline{v}_n)$
$\underline{y}_n = h_n(\underline{x}_{n-1}, \underline{w}_n)$

**Posterior Conditional PDF:** $f_{X_n|Y_n}(x_n|y_n) \propto \overbrace{f_{Y_n|X_n}(y_n|x_n)}^{\text{likelihood}} \cdot$
$\cdot \int_{\mathbb{X}} \underbrace{f_{X_n|X_{n-1}}(x_n|x_{n-1})}_{\text{state transition}}\,\underbrace{f_{X_{n-1}|Y_{n-1}}(x_{n-1}|y_{n-1})}_{\text{last conditional PDF}}\,\mathrm{d}x_{n-1}$

$N$ random Particles with particle weight $w_n^i$ at time $n$

**Monte-Carlo-Integration:** $I = \mathrm{E}[g(X)] \approx I_N = \frac{1}{N}\sum_{i=1}^{N}\tilde{g}(x^i)$

**Importance Sampling:** Instead of $f_X(x)$ use **Importance Density** $q_X(x)$

$I_N = \frac{1}{N}\sum_{i=1}^{N}\tilde{w}^i g(x^i)$ with weights $\tilde{w}^i = \frac{f_X(x^i)}{q_X(x^i)}$

If $\int f_{X_n}(x)\,\mathrm{d}x \neq 1$ then $I_N = \sum_{i=1}^{N}\tilde{w}^i g(x^i)$

### 9.5. Conditional Stochastical Independence

$\mathrm{P}(A \cap B|E) = \mathrm{P}(A|E) \cdot \mathrm{P}(B|E)$

Given $Y$, $X$ and $Z$ are independent if

$f_{Z \mid Y,X}(z|y, x) = f_{Z \mid Y}(z|y)$ or

$f_{X,Z \mid Y}(x, z|y) = f_{Z \mid Y}(z|y) \cdot f_{X \mid Y}(x|y)$

$f_{Z \mid X,Y}(z|x, y) = f_{Z \mid Y}(z|y)$ or $f_{X \mid Z,Y}(x|z, y) = f_{X \mid Y}(x|y)$

# 10. Hypothesis Testing

making a decision based on the observations

### 10.1. Definition

Null hypothesis $H_0 : \theta \in \Theta_0$ (Assumed first to be true)

Alternate hypothesis $H_1 : \theta \in \Theta_1$ (The one to proof)

Decision rule $\varphi : \mathbb{X} \to [0, 1]$ with

$\varphi(x) = 1$: decide for $H_1$, $\varphi(x) = 0$: decide for $H_0$ Error level $\alpha$ with $\mathrm{E}[d(X)|\theta] \le \alpha, \forall\theta \in \Theta_0$

| Error Type | Decision \ Reality | $H_1$ false ($H_0$ true) | $H_1$ true ($H_0$ false) |
|---|---|---|---|
| 1 (FA) False Alarm | $H_1$ rejected ($H_0$ accepted) | True Negative $\mathrm{P} = 1 - \alpha$ | False Negative (Type 2) $\mathrm{P} = \beta$ |
| 2 (DE) Detection Error | $H_1$ accepted ($H_0$ rejected) | False Positive (Type 1) $\mathrm{P} = \alpha$ | True Positive $\mathrm{P} = 1 - \beta$ |

Power: Sensitivity/Recall/Hit Rate: $\frac{\mathrm{TP}}{\mathrm{TP+FN}} = 1 - \beta$

Specificity/True negative rate: $\frac{\mathrm{TN}}{\mathrm{FP+TN}} = 1 - \alpha$

Precision/Positive Prediciton rate: $\frac{\mathrm{TP}}{\mathrm{TP+FP}}$

Accuracy: $\frac{\mathrm{TP+TN}}{\mathrm{P+N}} = \frac{2-\alpha-\beta}{2}$

#### 10.1.1 Design of a test

Cost criterion $G_\varphi : \Theta \to [0, 1], \theta \mapsto \mathrm{E}[d(X)|\theta]$

False Positive lower than $\alpha$: $G_d(\theta)|_{\theta\in\Theta_0} \le \alpha, \forall\theta \in \Theta_0$

False Negative small as possible: $\max\{G_d(\theta)|_{\theta\in\Theta_1}\}, \forall\theta \in \Theta_1$

### 10.2. Sufficient Statistics

Sufficiency for a test $T(X)$ means that no other test statistic, i.e., function of the observations $\underline{x}$, contains additional information about the parameter $\theta$ to be estimated:

$f_{X \mid T}(x|T(x) = t, \theta) = f_{X \mid T}(x|T(x) = t)$

# 11. Tests

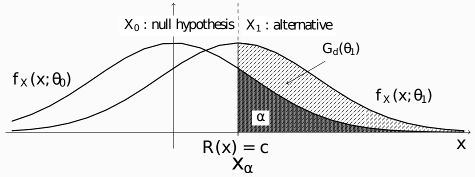
### 11.1. Neyman-Pearson-Test

The best test of $\mathrm{P}_0$ against $\mathrm{P}_1$ is

$d_{\mathrm{NP}}(x) = \begin{cases} 1 & R(x) > c \\ \gamma & R(x) = c \\ 0 & R(x) < c \end{cases}$ $\quad$ Likelihood-Ratio: $R(x) = \frac{f_X(x;\theta_1)}{f_X(x;\theta_0)}$

$\gamma = \frac{\alpha - \mathrm{P}_0(\{R > c\})}{\mathrm{P}_0(\{R = c\})}$ $\quad$ Errorlevel $\alpha$

Steps: For $\alpha$ calculate $x_\alpha$, then $c = R(x_\alpha)$




**Maximum Likelihood Detector:** $d_{\mathrm{ML}}(x) = \begin{cases} 1 & R(x) > 1 \\ 0 & \text{otherwise} \end{cases}$

**ROC Graphs:** plot $G_d(\theta_1)$ as a function of $G_d(\theta_0)$

### 11.2. Bayes Test (MAP Detector)

Prior knowledge on possible hypotheses: $\mathrm{P}(\{\theta \in \Theta_0\}) + \mathrm{P}(\{\theta \in \Theta_1\}) = 1$, minimizes the probability of a wrong decision.

$d_{\mathrm{Bayes}} = \begin{cases} 1 & \frac{f_X(x|\theta_1)}{f_X(x|\theta_0)} > \frac{c_0\,\mathrm{P}(\theta_0|x)}{c_1\,\mathrm{P}(\theta_1|x)} \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & \mathrm{P}(\theta_1|x) > \mathrm{P}(\theta_0|x) \\ 0 & \text{otherwise} \end{cases}$

Risk weights $c_0, c_1$ are 1 by default.

If $\mathrm{P}(\theta_0) = \mathrm{P}(\theta_1)$, the Bayes test is equivalent to the ML test

**Loss Function** $L(d(x), \theta) = \begin{cases} c_0 & \text{type 1 } d(x) = 1, \text{ but } \theta = \theta_0 \\ c_1 & \text{type 2 } d(x) = 0, \text{ but } \theta = \theta_1 \end{cases}$

$\mathrm{risk}(d) = \mathrm{E}[L(d(X), \theta)] = \mathrm{E}[\mathrm{E}[L(d(x), \theta)|x = X]]$

**Multiple Hypothesis** $d_{\mathrm{Bayes}} = \begin{cases} 0 & x \in \mathbb{X}_0 \\ 1 & x \in \mathbb{X}_1 \\ 2 & x \in \mathbb{X}_2 \end{cases}$

### 11.3. Linear Alternative Tests

$d : \mathbb{X} \to \mathbb{R}, \underline{x} \mapsto \begin{cases} 1 & \underline{w}^\top\underline{x} - w_0 > 0 \\ 0 & \text{otherwise} \end{cases}$

Estimate normal vector $\underline{w}^\top$, which separates $\mathbb{X}$ into $\mathbb{X}_0$ and $\mathbb{X}_1$

$\log R(\underline{x}) = \frac{\ln(\det(\underline{C}_0))}{\ln(\det(\underline{C}_1))} + \frac{1}{2}(\underline{x} - \underline{\mu}_0)^\top\underline{C}_0^{-1}(\underline{x} - \underline{\mu}_0) -$
$-\frac{1}{2}(\underline{x} - \underline{\mu}_1)^\top\underline{C}_1^{-1}(\underline{x} - \underline{\mu}_1) = 0$

**For 2 Gaussians**, with $\underline{C}_0 = \underline{C}_1 = \underline{C}$: $\underline{w}^\top = (\underline{\mu}_1 - \underline{\mu}_0)^\top\underline{C}$

and constant translation $w_0 = \frac{(\underline{\mu}_1 - \underline{\mu}_0)^\top\underline{C}(\underline{\mu}_1 - \underline{\mu}_0)}{2}$