Statistical Signal Processing
Formula Collection

Lukas Schüttler

July 26, 2023

i.i.d: independent and identically distributed

1 Math Basics

Binome, Trinome

$$(a \pm b)^3 = a^3 \pm 3a^2b + 3ab^2 \pm b^3$$
$$(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$$

Sequences and Series

(Aritmetric Series) $$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$
(Geometric Series) $$\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q}$$
(Exponential Series) $$\sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z$$

Mean

$$\mu_{ar} = \frac{1}{N}\sum x_i \leq \quad \mu_{geo} = \sqrt[N]{\prod x_i} \leq \quad \mu_{har} = \frac{N}{\sum \frac{1}{x_i}}$$

Inequalities

Cauchy-Schwarz: $| \underline{x}^T \underline{y} | \leq || \underline{x} || \cdot || \underline{y} ||$

Bernoulli: $(1+x)^n \geq 1 + nx$

Triangle: $| a + b | \leq | a | + | b |$

Sets

De Morgan's Laws: $\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$

1.1 Differentiation ($\forall \lambda, \mu \in \mathbb{R}$)

$$(\lambda f(x) + \mu g(x))' = \lambda f'(x) + \mu g'(x)$$
$$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$
$$(f(g(x)))' = f'(g(x)) \cdot g'(x)$$
$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{g(x)^2}$$

1.2 Integration

$$\int f'(x) \cdot g(x)\, dx = f(x) \cdot g(x) - \int f(x) \cdot g'(x)\, dx$$
$$\int f(g(x)) \cdot g'(x)\, dx = \int f(u)\, du \quad \text{mit } u = g(x)$$

f(x)	F(x) - C	f'(x)
x^n	$\frac{1}{n+1}x^{n+1}$	nx^{n-1}
$\log(ax)$	$x\log(ax) - x$	$\frac{1}{x}$
$x \cdot e^x$	$(x-1)e^x$	$(x+1)e^x$
a^x	$\frac{a^x}{\log(a)}$	$a^x \cdot \log(a)$
$\sin(x)$	$-\cos(x)$	$\cos(x)$

1.3 Matrices

$\mathbf{A} \in \mathbb{K}^{m \times n}$: Matrix with m rows and n columns

$$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \qquad (\mathbf{A}\cdot\mathbf{B})^T = \mathbf{B}^T \cdot \mathbf{A}^T$$
$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T \qquad (\mathbf{A}\cdot\mathbf{B})^{-1} = \mathbf{B}^{-1} \cdot \mathbf{A}^{-1}$$

$\dim(\mathbf{A}) = n = \text{rank}(\mathbf{A}) + \dim\ker(\mathbf{A})$

1.3.1 Quadratic Matrices

$\mathbf{A} \in \mathbb{K}^{n \times n}$: Square matrix of order n

regular/invertible/non-singular: $\det(\mathbf{A}) \neq 0$, $\text{rank}(\mathbf{A}) = n$
singular/non-invertible: $\det(\mathbf{A}) = 0$, $\text{rank}(\mathbf{A}) < n$
$\mathbf{A}^{-1}$ exists for regular matrices
orthogonal: $\mathbf{A}^T \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{A}^T = \mathbf{I} \implies \det(\mathbf{A}) = \pm 1$
symmetric: $\mathbf{A}^T = \mathbf{A}$

1.3.2 Determinant of $\mathbf{A} \in \mathbb{K}^{n \times n}$

$\det \mathbf{A} = | \mathbf{A} |$

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

$\det \mathbf{A} = \det \mathbf{A}^T$
$\det(\mathbf{A} \cdot \mathbf{B}) = \det \mathbf{A} \cdot \det \mathbf{B} = \det \mathbf{B} \cdot \det \mathbf{A} = \det(\mathbf{B} \cdot \mathbf{A})$
If $\text{rank}(\mathbf{A}) < n$, then $\det(\mathbf{A}) = 0$

1.3.3 Eigenvalues and Eigenvectors

$$\mathbf{A} \cdot \underline{x} = \lambda \cdot \underline{x} \qquad \det(\mathbf{A}) = \prod \lambda_i \qquad \text{tr}\{\mathbf{A}\} = \sum \lambda_i$$
$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^T$$

Eigenvectors of $\mathbf{A}$ span the range of $\mathbf{A}$
If only the trivial solution $\lambda = 0$ exists $\implies \underline{x} \in \ker(\mathbf{A})$
EW of Triangular/Diagonal Matrix: $\lambda_i = a_{ii}$ (diagonal elements)

1.3.4 Singularvalues and Singularvectors

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^T \quad \mathbf{A}^T \cdot \mathbf{A} = \mathbf{V} \cdot \mathbf{\Sigma}^2 \cdot \mathbf{V}^T \quad \mathbf{A} \cdot \mathbf{A}^T = \mathbf{U} \cdot \mathbf{\Sigma}^2 \cdot \mathbf{U}^T$$

Left singular vectors span the range of $\mathbf{A}$
Right singular vectors span the range of $\mathbf{A}^T$ (domain of $\mathbf{A}$)
If $\sigma_i = 0$ then $\underline{v}_i \in \ker(\mathbf{A})$ and $\underline{u}_i \in \ker(\mathbf{A}^T)$

1.4 Pseudo Inverse ($\mathbf{A} \in \mathbb{K}^{m \times n}$)

$(\mathbf{A}^T\mathbf{A})^{-1}$ exists $\implies \mathbf{A}^+_{\text{left}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \; m \geq n$
Is orthogonal projector onto range[$\mathbf{A}$]
$(\mathbf{A}\mathbf{A}^T)^{-1}$ exists $\implies \mathbf{A}^+_{\text{right}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \; m \leq n$
Is orthogonal projector onto range[$\mathbf{A}^T$]

1.4.1 Helpful Tricks

$$tr(\mathbf{A}\mathbf{B}^T + \mathbf{B}\mathbf{A}^T) = 2tr(\mathbf{A}\mathbf{B}^T)$$
$$\underline{a}^T\underline{b} = \underline{b}^T\underline{a} \implies \underline{a}^T\mathbf{M}\underline{b} = \underline{b}^T\mathbf{M}^T\underline{a}$$
$$\frac{d}{d\underline{h}}(\underline{y} - \mathbf{S}\underline{h})^T\mathbf{C}^{-1}(\underline{y} - \mathbf{S}\underline{h}) = -2\mathbf{S}^T\mathbf{C}^{-1}(\underline{y} - \mathbf{S}\underline{h})$$

2 Probability Theory

2.1 Combinatorics

Possible combinations/variations of choosing k elements out of n elements (distribute k elements into n bins):

	with repetition	without repetition
order	n^k	$\frac{n!}{(n-k)!}$
no order	$\binom{n+k-1}{k}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Permutations of n elements: $n!$

Permutations of n elements with k same elements: $\frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_n!}$

Binomialcoefficient: $\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{k!(n-k)!}$

$$\binom{n}{0} = 1 \qquad \binom{n}{1} = n \qquad \binom{n}{n} = 1$$

2.2 Probability space

Sample space: Set of all possible outcomes
$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$

Event space: Set of all possible events
$\mathbb{F} = \{A_1, A_2, \ldots, A_n\}$ with $A_i \subseteq \Omega$

Probability measure: Assigns probabilities to events
$P : \mathbb{F} \to [0, 1]$

Random variable: Maps outcomes to events
$X : \Omega \to \Omega$ with $X(\omega) = x \in A$

Observations: Single outcome of a random variable
$\{x_1, x_2, \ldots, x_n\} \subseteq \Omega$

Unknown parameters: Parameters of a probability distribution
$\theta \in \Theta$

Estimator: Function of observations that estimates θ
$T : \mathbb{X} \to \Theta \implies \hat{\theta} = T(X)$

2.3 Probability measure

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

The probability of the event A is the number of outcomes in A divided by the total number of outcomes in Ω

2.3.1 Axioms of Kolmogorov

with $A_i \cap A_j = \emptyset$ for $i \neq j$

$$P(A) \geq 0 \qquad P(\Omega) = 1 \qquad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

2.4 Independence

$X_1, X_2, \ldots, X_n$ are independent if and only if:

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

$$p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$$

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

2.5 Distribution

Probabilitydensity function (PDF):
$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$
$f_{X,Y}(x, y) = \frac{\mathrm{d}^2 F_{X,Y}(x,y)}{\mathrm{d}x \ \mathrm{d}y}$ (Joint PDF)

Cumulative distribution function (CDF):
$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t$
$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t)\mathrm{d}s \ \mathrm{d}t$

2.6 Conditional Probability

Probability of event A given that event B has occurred:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad P(A \cap B) = P(A \mid B) \cdot P(B)$$

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

$$f_Y(y) = \underbrace{\int f_{X,Y}(x, y)\mathrm{d}x}_{\text{marginalization}} = \int f_{Y|X}(y \mid x) \cdot f_X(x)\mathrm{d}x$$

Bayes' Theorem:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Conditional Stochastic Independence:

X and Z are conditionally independent given Y: XYZ

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

3 Common Distributions

3.1 Normal Distribution $\sim \mathcal{N}(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$f_{\vec{X}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^k \det(\mathbf{C})}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \mathbf{C}^{-1} (\vec{x} - \vec{\mu})\right)$$

with $\det(a\mathbf{A}) = a^n \det(\mathbf{A})$ if $\mathbf{A} \in \mathbb{R}^{n \times n}$

$$E[X] = \mu \qquad \mathrm{Var}[X] = \sigma^2$$

3.2 Uniform Distribution $\sim \mathcal{U}(a, b)$

$$f_X(x) = \frac{1}{b-a} \qquad \mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

3.3 Exponential Distribution $\sim$ Exp(λ)

$$f_X(x) = \lambda \exp(-\lambda x) \qquad \mu = \frac{1}{\lambda} \qquad \sigma^2 = \frac{1}{\lambda^2}$$

$$f_X(x; \theta) = \frac{h(x) \exp(a(\theta)t(x))}{\exp(b(\theta))}$$

3.4 Gamma Distribution $\sim \Gamma(\alpha, \beta)$

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) \qquad \mu = \frac{\alpha}{\beta} \qquad \sigma^2 = \frac{\alpha}{\beta^2}$$

3.5 Binomial Distribution $\sim \mathcal{B}(K;\theta)$

$$f_X(x) = \binom{K}{x}\theta^x(1-\theta)^{K-x} \qquad \mu = K\theta \qquad \sigma^2 = K\theta(1-\theta)$$

4 Important Properties

4.1 Expectation (first order moment)

The expectation of a random variable X is the average value of X over many trials.

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x \quad \hat{=} \sum_{x\in\mathcal{X}} x \cdot P_X(x)$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\mathrm{d}x \quad \hat{=} \sum_{x\in\mathcal{X}} g(x) \cdot P_X(x)$$

$$E[aX+b] = aE[X]+b$$
$$E[X+Y] = E[X]+E[Y]$$
$$E[XY] = E[X]E[Y] \qquad \text{if } X \text{ and } Y \text{ are independent}$$
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\mathrm{d}x$$

4.2 Variance (second order moment)

The variance of a random variable X is the average squared deviation from the mean.

$$\mathrm{Var}[X] = E[(X-E[X])^2] = E[X^2]-E[X]^2$$
$$\mathrm{Var}[\underline{X}] = \mathrm{E}[(\underline{X}-\vec{\mu})(\underline{X}-\vec{\mu})^T] = \mathrm{E}[\underline{X}\underline{X}^T]-\vec{\mu}\vec{\mu}^T$$

$$\mathrm{Var}[X] = \mathrm{Cov}[X,X]$$
$$\mathrm{Var}[aX+b] = a^2\mathrm{Var}[X]$$
$$\mathrm{Var}[X+Y] = \mathrm{Var}[X]+\mathrm{Var}[Y]+2\mathrm{Cov}[X,Y]$$
$$\mathrm{Var}[XY] = E[X^2]E[Y^2]-E[X]^2E[Y]^2+\mathrm{Cov}[X,Y]^2$$
$$\mathrm{Var}[X] = E[X^2]-E[X]^2$$
$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n}\mathrm{Var}[X_i] + \sum_{i\neq j}\mathrm{Cov}[X_i,X_j]$$

4.3 Covariance

The covariance of two random variables X and Y is a measure of how much one can be expressed by the other.

$$\mathrm{Cov}[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] = E[XY]-\mu_X\mu_Y$$
$$\mathrm{Cov}[\underline{X},\underline{Y}] = \mathrm{E}[(\underline{X}-\vec{\mu}_X)(\underline{Y}-\vec{\mu}_Y)^T]$$

$$\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X]$$
$$\mathrm{Cov}[aX+b,cY+d] = ac\mathrm{Cov}[X,Y]$$
$$\mathrm{Cov}[X+U,Y] = \mathrm{Cov}[X,Y]+\mathrm{Cov}[U,Y]$$
$$\mathrm{Cov}[\underline{z}] = \mathbf{C}_{\underline{z}} = \begin{bmatrix}\mathbf{C}_{\underline{x}} & \mathbf{C}_{\underline{x},\underline{y}}\\ \mathbf{C}_{\underline{y},\underline{x}} & \mathbf{C}_{\underline{y}}\end{bmatrix} \text{ with } \underline{z} = [\underline{x}^T,\underline{y}^T]^T$$

4.3.1 Correlation

The correlation is the normalized covariance.

$$\rho(X,Y) = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} = \frac{\mathbf{C}_{X,Y}}{\sigma_X\sigma_Y}$$

4.4 Sample Correlation and Covariance

The sample covariance is the covariance of two random variables X and Y estimated from a sample of N observations.

$$\hat{\mathbf{C}}_{X,Y} = \frac{1}{N}\sum_{i=1}^{N}(\underline{x}_i-\underline{\mu}_X)(\underline{y}_i-\underline{\mu}_Y)^T$$
$$\hat{\mathbf{R}}_{X,Y} = \frac{\hat{\mathbf{C}}_{X,Y}}{\sqrt{\hat{\mathbf{C}}_{X,X}\hat{\mathbf{C}}_{Y,Y}}} = \frac{1}{N}\sum_{i=1}^{N}\underline{x}_i\underline{y}_i^T$$

5 Estimation

5.1 Estimator Quality

Consistency: $\lim_{N\to\infty}\mathrm{Var}[T] = 0$ or $\lim_{N\to\infty} T = \theta$
Unbiasedness: $E[T] = \theta$ with $\mathrm{Bias}[T] = E[T]-\theta$
Efficiency: $\sqrt{N}E[T(x_1\ldots x_N)-\theta] \overset{N\to\infty}{\sim} \mathcal{N}(0, I_F^{(1)})$
Variance: $\mathrm{Var}[T] = E[(T-E[T])^2]$

5.1.1 Mean Squared Error (MSE)

The MSE is the expected value of the squared error.

$$\epsilon[T] = E[(T-\theta)^2] = \mathrm{Var}[T]+\mathrm{Bias}^2[T]$$
$$\epsilon[\underline{T}] = \mathrm{E}[\|\underline{T}-\underline{\theta}\|^2] = \mathrm{tr}\{E[(\underline{T}-\underline{\theta})(\underline{T}-\underline{\theta})^T]\}$$

5.1.2 Minimum Mean Squared Error (MMSE)

The MMSE is the minimum MSE over all possible estimators.

$$\arg\min_T E[(T-\theta)^2]$$

An estimator is MMSE if it minimizes the MSE for all θ.

5.2 Comparison of Estimators

Estimate mean θ of X with prior knowledge $\Theta \sim \mathcal{N}(\mu,\sigma_\Theta^2)$ and $X \sim \mathcal{N}(\theta,\sigma_X^2)$. For N i.i.d. observations x_i

$$\hat{\theta}_{\mathrm{CM}} = \frac{N\sigma_\Theta^2}{N\sigma_\Theta^2+\sigma_{X|\theta}^2}\hat{\theta}_{\mathrm{ML}} + \frac{\sigma_X^2}{N\sigma_\Theta^2+\sigma_{X|\theta}^2}\mu$$
$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

For large N or $\sigma_\Theta^2 \gg \sigma_{X|\theta}^2$ the ML estimator used.
For small N large $\sigma_{X|\theta}^2$ or small σ_Θ^2 the knowledge about Θ is used to improve estimation.

5.3 Maximum Likelihood Estimation (ML)

The ML estimator is the value of θ that maximizes the likelihood function $L(x;\theta)$ given $f_X(x;\theta)$.

Likelihood function:
$L(x_1,\ldots,x_n;\theta) = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n;\theta)$
$L(x_1,\ldots,x_n;\theta) = P_\theta(X_1 = x_1,\ldots,X_n = x_n)$

If N observations are i.i.d.:

$$L(x;\theta) = \prod_{i=1}^{N} f_{X_i}(x_i;\theta) \qquad l(x;\theta) = \sum_{i=1}^{N}\log f_{X_i}(x_i;\theta)$$

Maximum Likelihood Estimator:

$$T_{ML} = \arg\max_\theta\{L(x;\theta)\} = \arg\max_\theta\{l(x;\theta)\}$$
$$\frac{\delta L(x;\theta)}{\delta\theta} = \frac{\delta l(x;\theta)}{\delta\theta} \overset{!}{=} 0$$

Properties: The ML Estimator is consistent, asymptotically unbiased and asymptotically efficient.

5.4 Uniformly Minimum Variance Unbiased Estimator (UMVU)

The UMVU estimator is the unbiased estimator with the smallest variance. (Best unbiased estimator)

Fisher's Information Inequality: Estimate lower bound for the variance if

$L(x, \theta) > 0 \quad \forall x, \theta$

$L(x, \theta)$ is differentiable in respect to θ

$\frac{\delta}{\delta\theta} \int L(x,\theta)\mathrm{d}x = \int \frac{\delta}{\delta\theta} L(x,\theta)\mathrm{d}x$

Score function:

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

Fisher information:

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

Cramer-Rao Lower Bound (CRB):

$$\boxed{\mathrm{Var}[T] \geq \left(\frac{\delta E[T(X)]}{\delta\theta}\right)^2 \frac{1}{I_F(\theta)} \qquad \mathrm{Var}[T] \geq \frac{1}{I_F(\theta)}}$$

with T being unbiased $\implies E[T(X)] = \theta$

For N i.i.d. observations: $I_F^{(N)}(x,\theta) = N \cdot I_F(x,\theta)$

5.4.1 Exponential Models

$$f_X(x;\theta) = \frac{h(x)\exp(a(\theta)t(x))}{\exp(b(\theta))} \qquad I_F(\theta) = \frac{\delta a}{\delta\theta}\frac{\delta E[t(X)]}{\delta\theta}$$

5.4.2 Useful derivations

Uniform $\mathcal{U}(a,b)$: Not differentiable $\implies$ no $I_F(\theta)$

Normal $\mathcal{N}(\mu,\sigma^2)$: $g(x,\theta) = \frac{x-\theta}{\sigma^2}$ $I_F(\theta) = \frac{1}{\sigma^2}$

Binomial $\mathcal{B}(K,\theta)$: $g(x,\theta) = \frac{x}{\theta} - \frac{K-x}{1-\theta}$ $I_F(\theta) = \frac{K}{\theta(1-\theta)}$

5.5 Bayes Estimation (Conditional Mean)

A priori information about θ is given by the pdf $f_\Theta(\theta;\sigma)$. The conditional pdf (posterior pdf) $f_{X|\Theta}(x \mid \theta)$ is used to find θ by minimizing the mean MSE instead of the uniform MSE.

Mean MSE for Θ: $E[E[(T(X)-\Theta)^2 \mid \Theta = \theta]]$

Conditional Mean Estimator:

$$T_{CM} : x \mapsto E[\Theta \mid X = x] = \int \theta f_{\Theta|X}(\theta \mid x)\mathrm{d}\theta$$

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

with $f_X(x) = \text{const.}$ $\implies$ can be replaced by a factor $\frac{1}{\gamma}$. γ can be determined such that $\int_\Theta f_{\Theta|X}(\theta \mid x)\mathrm{d}\theta = 1$

MSE if $\Theta \sim \mathcal{N}(\mu,\sigma_\Theta^2)$, $X \sim \mathcal{N}(\mu,\sigma_X^2)$: $E[\mathrm{Var}[\Theta \mid X]]$

Jointly Gaussian: $(\Theta, X \sim \mathcal{N})$

$$T_{CM} = E[\Theta \mid X = x] = \underline{\mu_\Theta} + \mathbf{C}_{\Theta X}\mathbf{C}_X^{-1}\underbrace{(\underline{x} - \underline{\mu_X})}_{\Delta\underline{x}}$$

$$E[\| T_{CM} - \Theta \|^2] = \mathrm{tr}\{\mathbf{C}_{\Theta|X=x}\}$$

$$\mathbf{C}_{\Theta|X} = \mathbf{C}_\Theta - \mathbf{C}_{\Theta X}\mathbf{C}_X^{-1}\mathbf{C}_{X\Theta}$$

Orthogonality Principle:

$$T_{CM}(X) - \Theta \perp h(X) \implies E[(T_{CM}(X)-\Theta)h(X)] = 0$$

$\implies$ Error has no correlation with the estimator or random variable

Properties: The CM ist LMMSE if it is linear in X e.q. the estimator is a linear function of X

6 Linear Estimation

$$\boxed{\hat{y} = \mathbf{x}^T\underline{t} + m \quad \hat{y} = \mathbf{X}'\underline{t}' \quad \underline{t}' = \begin{bmatrix}\underline{t} \\ m\end{bmatrix} \quad \mathbf{X}' = \begin{bmatrix}\mathbf{X} & \underline{1}\end{bmatrix}}$$

Given N observations y_i based on a input $\underline{x}_i$.

$$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}$$

Estimation of $\underline{x}$:

$$\underline{\hat{y}} = \underline{x}^T\mathbf{T} \implies \underline{\hat{x}} = \mathbf{T}^T\underline{y}$$

6.1 Comparison

With MIMO channel $\mathbf{H} \in \mathbb{K}^{M\times N}$ (N inputs, M outputs, K Observations)

Estimator	Average squared bias	Variance
ML	0	$N \cdot M \cdot \gamma$
Matched Filter	$\sum_{i=1}^{KM}\lambda_i\left(\frac{\lambda_i}{\lambda_1}-1\right)$	$\sum_{i=1}^{KM}\frac{\lambda_i^2}{\lambda_1^2}\gamma$
MMSE	$\sum_{i=1}^{KM}\frac{\lambda_i}{(1+\gamma^{-1}\lambda_i)^2}$	$\sum_{i=1}^{KM}\frac{\gamma}{(1+\gamma\lambda_i^{-1})^2}$

Note: $\gamma = \frac{\sigma_n^2}{K\sigma_s^2}$

6.2 Linear Minimum Mean Square Error Estimation (LMMSE)

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

Note: The underlying model is not necessarily linear.

$$\boxed{\underline{\hat{y}}_{\mathrm{LMMSE}} = \arg\min_{\underline{t},m} E\left[\| \underline{y} - \underline{x}^T\underline{t} - m \|^2\right]}$$

Joint Variable:

$$\underline{z} = \begin{bmatrix}\underline{x} \\ y\end{bmatrix} \qquad \underline{\mu_z} = \begin{bmatrix}\underline{\mu_x} \\ \mu_y\end{bmatrix} \qquad \mathbf{C}_z = \begin{bmatrix}\mathbf{C}_x & \underline{c}_{xy} \\ \underline{c}_{xy}^T & c_y^2\end{bmatrix}$$

LMMSE Estimation of y given $\underline{x}$:

$$\hat{y} = \mu_y + \underline{c}_{xy}^T\mathbf{C}_x^{-1}(\underline{x} - \underline{\mu}_x) = \underbrace{\underline{c}_{xy}^T\mathbf{C}_x^{-1}}_{\underline{t}^T}\underline{x} + \underbrace{\mu_y - \underline{c}_{xy}^T\mathbf{C}_x^{-1}\underline{\mu}_x}_{m}$$

$$E\left[\| \underline{y} - \underline{x}^T\underline{t} - m \|^2\right] = c_y^2 - \underline{c}_{xy}^T\mathbf{C}_x^{-1}\underline{c}_{xy}$$

Hint: Use general form of $\hat{y}$ then insert variables according to the given problem.

Properties: The LMMSE estimator depends on first and second order moments of the distribution. It does not consider the distribution of the random variables. The LMMSE ist not a random variable.

The LMMSE estimator is MMSE (CM Estimator) if the underlying model is linear (jointly Gaussian).

6.3 Least Squares Estimation (LSE)

Minimize the squared error between the observations $\underline{y}$ and the model $\mathbf{X}\underline{t}$.

LS Error: $\min\left[\sum(y_i - \underline{x}_i^T\underline{t})^2\right] = \min\left[\| \underline{y} - \mathbf{X}\underline{t} \|^2\right]$

$$\boxed{\underline{t}_{\mathrm{LS}} = \underbrace{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T}_{\text{Pseudo inverse } X^+}\underline{y} \qquad \underline{\hat{y}}_{\mathrm{LS}} = \mathbf{X}\underline{t}_{\mathrm{LS}}}$$

Orthogonality principle:

$$\underline{y} - \mathbf{X}\underline{t}_{\text{LS}} \perp \text{range}[\mathbf{X}] \implies \underline{y} - \mathbf{X}\underline{t}_{\text{LS}} \in \ker[\mathbf{X}^T]$$

$$\mathbf{X}^T(\underline{y} - \mathbf{X}\underline{t}_{\text{LS}}) = \mathbf{0} \implies \mathbf{X}^T\underline{y} = \mathbf{X}^T\mathbf{X}\underline{t}_{\text{LS}}$$

if $N \geq \text{rank}[\mathbf{X}]$ (All columns of $\mathbf{X}$ are independent, $(\mathbf{X}^T\mathbf{X})^{-1}$ exists)

6.4 Matched Filter

The optimal linear filter for maximizing the SNR of a signal in the presence of additive stochastic noise.

For a channel $\underline{y} = \underline{h}x + \underline{n}$ filtered with $\mathbf{T}$: $\mathbf{T}\underline{y} = \mathbf{T}\underline{h}x + \mathbf{T}\underline{n}$
Such that $\hat{\underline{h}} = \mathbf{T}\underline{y}$ is the optimal estimate of $\underline{h}$.

$$\boxed{\mathbf{T}_{\text{MF}} \implies \max_{\mathbf{T}} \left\{ \frac{E[(\mathbf{T}\underline{h}x)^2]}{E[(\mathbf{T}\underline{n})^2]} \right\} = \max_{\mathbf{T}} \left\{ \frac{|\, E[\hat{\underline{h}}^T \underline{h}]\,|^2}{\operatorname{tr}\{\text{Var}[\mathbf{T}\underline{n}]\}} \right\}}$$

With MIMO channel $\mathbf{H} \in \mathbb{K}^{M\times N}$ (N inputs, M outputs, K Observations)

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{N} \implies \underline{y} = (\mathbf{S}^T \otimes \mathbf{I}_M)\underline{h} + \underline{n}$$

All matrices $\mathbf{Y}, \mathbf{H}, \mathbf{N}$ are stacked column-wise into vectors $\underline{y}, \underline{h}, \underline{n}$

$$\hat{\underline{h}} = \mathbf{T}\underline{y} \qquad \mathbf{T}_{\text{MF}} = \mathbf{C}_h(\mathbf{S}^T \otimes \mathbf{I}_M)^T\mathbf{C}_n^{-1}$$

7 Sequences

7.1 Random Sequences

Sequence of random variables $X_1, X_2, \ldots, X_N$. For example multiple dice rolls after each other.

7.2 Markov Sequences

Random sequence where value depends only on the previous value.
1. state: $f_{X_1}(x_1)$
2. state: $f_{X_2}(x_2 \mid x_1)$
n. state: $f_{X_n}(x_n \mid x_{n-1}, \ldots, x_1) = f_{X_n}(x_n \mid x_{n-1})$

7.2.1 Hidden Markov Chain

Markov sequence where the state is not directly observable. X can only be guessed from Y.

$$X_n = G_n(X_{n-1}) \qquad Y_n = H_n(X_n)$$

State-transition PDF: $f_{X_n|X_{n-1}}(x_n \mid x_{n-1})$

Estimation:

$$f_{X_n|Y_n} \propto \underbrace{f_{Y_n|X_n}}_{\text{likelihood}} \int_{\mathbb{X}} \underbrace{f_{X_n|X_{n-1}}}_{\text{state transition}} \underbrace{f_{X_{n-1}|Y_{n-1}}}_{\text{last state}} \mathrm{d}x_{n-1}$$

$$\hat{x}_{n|n} = E[X_n \mid Y_1, \ldots, Y_n] = E[X_n \mid Y_{(n)}] \quad \hat{y}_{n|n} = E[Y_n \mid Y_{(n)}]$$

Hint: Estimators like CM and LMMSE can be used to estimate $\hat{\underline{x}}_{n|n}$

For non linear problems the Kalman-Filter can be modified with the Extended Kalman-Filter (EKF) or the Unscented Kalman-Filter (UKF).

7.3 Extended Kalman-Filter

Linear approximation of non-linear functions for every step.

$$\underline{x}_n = \underline{g}_n\underline{x}_{n-1} + \underline{v}_n \qquad \underline{y}_n = \underline{h}_n\underline{x}_n + \underline{w}_n$$

$$\text{with} \quad \underline{g}_n = \frac{\delta \underline{g}_n}{\delta \underline{x}}\Big|_{x_n} \qquad \text{and} \quad \underline{h}_n = \frac{\delta \underline{h}_n}{\delta \underline{x}}\Big|_{x_n}$$

7.4 Unscented Kalman-Filter

Approximation of desired PDF with gaussian PDF.

7.5 Kalman-Filter

Recursively estimate the next state of a Gaussian Markov sequence.

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

With $\underline{v}_n \sim \mathcal{N}(\underline{\mu}_{v_n}, \mathbf{C}_{v_n})$ and $\underline{w}_n \sim \mathcal{N}(\underline{\mu}_{w_n}, \mathbf{C}_{w_n})$

0. Step: Initialization

$$\hat{\underline{x}}_{0|-1} = E[\underline{x}_0] \qquad \mathbf{C}_{x_{0|-1}} = \text{Var}[\underline{x}_0]$$

1. Step: Prediction

$$\hat{\underline{x}}_{n|n-1} = \mathbf{G}_n\hat{\underline{x}}_{n-1|n-1} \quad \mathbf{C}_{x_{n|n-1}} = \mathbf{G}_n\mathbf{C}_{x_{n-1|n-1}}\mathbf{G}_n^T + \mathbf{C}_{v_n}$$

2. Step: Update

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

$$\mathbf{C}_{x_{n|n}} = \mathbf{C}_{x_{n|n-1}} - \mathbf{K}_n\mathbf{H}_n\mathbf{C}_{x_{n|n-1}}$$

$$\mathbf{K}_n = \mathbf{C}_{x_{n|n-1}}\mathbf{H}_n^T(\underbrace{\mathbf{H}_n\mathbf{C}_{x_{n|n-1}}\mathbf{H}_n^T + \mathbf{C}_{w_n}}_{\mathbf{C}_{y_{n|n-1}}})^{-1} = \mathbf{C}_{x_n\Delta y_n}\mathbf{C}_{\Delta y_n}^{-1}$$

Inovation: Closness of the measurement to the prediction

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

7.6 Particle Filter

For non linear and non Gaussian problems. Approximate the PDF with a set of particles.

$$\underline{x}_n = \underline{g}_n(\underline{x}_{n-1}, \underline{v}_n) \qquad \underline{y}_n = \underline{h}_n(\underline{x}_n, \underline{w}_n)$$

N particles $\underline{x}_n^{(i)}$ with weights $w_n^{(i)}$ at step n. **Monte-Carlo-Integration:**

$$\int g(x)f(x)\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}^N g(x^{(i)}) \qquad \text{with} \quad x^{(i)} \sim f(x)$$

Importance Sampling: Instead of sampling from $f(x)$ sample from $q(x)$ (**Importance Density**)

$$\int g(x)f(x)\mathrm{d}x = \int g(x)\frac{f(x)}{q(x)}q(x)\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}^N g(x^{(i)})\frac{f(x^{(i)})}{q(x^{(i)})}$$

7.6.1 Wheight Update

$$\tilde{w}_n^{(i)} = \frac{f(x_n^{(i)})}{q(x_n^{(i)})} \quad w_n^{(i)} = \frac{\tilde{w}_n^{(i)}}{\sum_{j=1}^N \tilde{w}_n^{(j)}} \quad \sum_{i=1}^N w^{(i)}\delta(x - x^{(i)}) \approx f(x)$$

$$\tilde{w}_n^{(i)} = \frac{f_{X_n,X_{(n-1)}|Y_{(n)}}(x_n^{(i)}, x_{n-1}^{(i)})}{q_{X_n,X_{(n-1)}|Y_{(n)}}(x_n^{(i)}, x_{n-1}^{(i)})}$$

$$\approx \tilde{w}_{n-1}^{(i)}\frac{f_{Y_n|X_n}(y_n \mid x_n^{(i)})f_{X_n|X_{n-1}}(x_n^{(i)} \mid x_{n-1}^{(i)})}{q_{X_n|X_{n-1},Y_n}(x_n^{(i)} \mid x_{n-1}^{(i)}, y_n)}$$

Degeneracy: Monotonic increase of the weights over time. $\implies$ only some particles have a significant weight.

$$\frac{\max\{\sigma_{\text{est}}^2\}}{\sigma_{\text{est}^2}} = \frac{1}{\sum_{i=1}^N (w_n^{(i)})^2} \leq w_{\text{thr}}$$

Resampling: Particles with low weight are replaced by particles with high weight at the position of the high weight particles (with some noise).