Statistical Signal Processing
Formula Collection

Lukas Schüttler

July 24, 2023

i.i.d: independent and identically distributed

1 Math Basics

Binome, Trinome

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

Sequences and Series

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

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

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

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

1.3.1 Quadratic Matrices

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

regular/invertible/non-singular: $\det(\mathbf{A}) \neq 0$, $\operatorname{rank}(\mathbf{A}) = n$
singular/non-invertible: $\det(\mathbf{A}) = 0$, $\operatorname{rank}(\mathbf{A}) < n$
$\mathbf{A}^{-1}$ 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 $\operatorname{rank}(\mathbf{A}) < n$, then $\det(\mathbf{A}) = 0$

1.3.3 Eigenvalues and Eigenvectors

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

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 $\underline{u}_i \in \ker(\mathbf{A}^T)$

1.3.5 Helpful Tricks

$$tr(\mathbf{A}\mathbf{B}^T + \mathbf{B}\mathbf{A}^T) = 2tr(\mathbf{A}\mathbf{B}^T)$$
$$\underline{a}^T \underline{b} = \underline{b}^T \underline{a} \implies \underline{a}^T \mathbf{M} \underline{b} = \underline{b}^T \mathbf{M}^T \underline{a}$$

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

2.5 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:

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

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

2.6 Independence

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

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

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

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

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 $\sim \mathbf{Exp}(\lambda)$

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

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

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.

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

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

$$E[aX+b] = aE[X]+b$$

$$E[X+Y] = E[X]+E[Y]$$

$$E[XY] = E[X]E[Y] \qquad \text{if } X \text{ and } Y \text{ are independent}$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\mathrm{d}x$$

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.

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

5 Estimation

5.1 Estimator Quality

Consistency: $\lim_{n\to\infty} P(|\,\hat{\theta}_n-\theta\,| > \epsilon) = 0$
Unbiasedness: $E[\hat{\theta}] = \theta$ with $\mathrm{Bias}[\hat{\theta}] = E[\hat{\theta}]-\theta$
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.

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

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

5.1.2 Minimum Mean Squared Error (MMSE)

The MMSE is the minimum MSE over all possible estimators.

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

5.2 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:

$$T_{ML} = \arg\max_\theta\{L(x;\theta)\} = \arg\max_\theta\{l(x;\theta)\}$$

$$\frac{\delta L(x;\theta)}{\delta\theta} = \frac{\delta l(x;\theta)}{\delta\theta} \stackrel{!}{=} 0$$

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

5.3 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 twice differentiable in 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: $I_F^{(N)}(x,\theta) = N \cdot I_F(x,\theta)$

5.3.1 Exponential Models

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

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