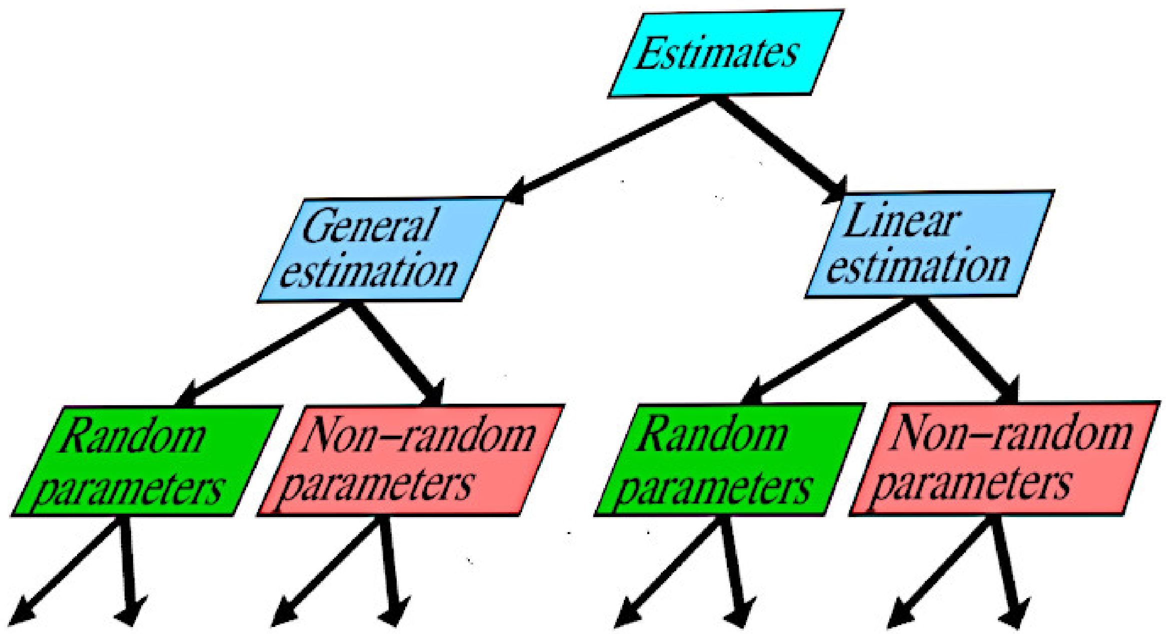
Key Ideas in Parameter Estimation



Pavel Loskot
pavelloskot@intl.zju.edu.cn

ZJU-UIUC INSTITUTE
Zhejiang University-University of Illinois at Urbana-Champaign Institute
浙江大学伊利诺伊大学厄巴纳香槟校区联合学院

The Ninth International Conference on Advances in Signal, Image and Video Processing
SIGNAL 2024
March 10, 2024 to March 14, 2024 - Athens, Greece

About Me

Pavel Loskot joined the ZJU-UIUC Institute as Associate Professor in January 2021. He received his PhD degree in Wireless Communications from the University of Alberta in Canada, and the MSc and BSc degrees in Radioelectronics and Biomedical Electronics, respectively, from the Czech Technical University of Prague. He is the Senior Member of the IEEE, Fellow of the HEA in the UK, and the Recognized Research Supervisor of the UKCGE.

In the past 25 years, he was involved in numerous industrial and academic collaborative projects in the Czech Republic, Finland, Canada, the UK, Turkey, and China. These projects concerned mainly wireless and optical telecommunication networks, but also genetic regulatory circuits, air transport services, and renewable energy systems. This experience allowed him to truly understand the interdisciplinary workings, and crossing the disciplines boundaries.

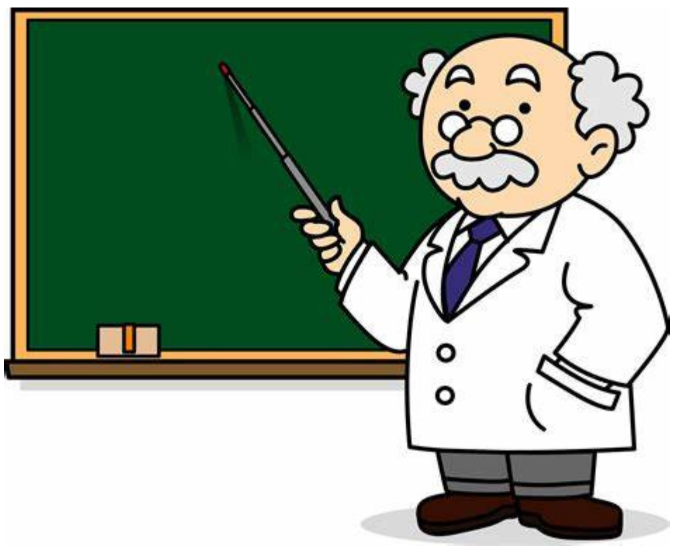
His current research focuses mathematical and probabilistic modeling, statistical signal processing and classical machine learning for multi-sensor data in biomedicine, computational molecular biology, and wireless communications.

Objectives

- estimation theory is important even at the age of machine learning
  → best possible, interpretable, computationally efficient (usually)
- understanding estimation of time-invariant parameters is a good start
  → then move on to time varying parameters i.e. signals
- review fundamental principles of parameter estimation
  → many topics not covered

Outline

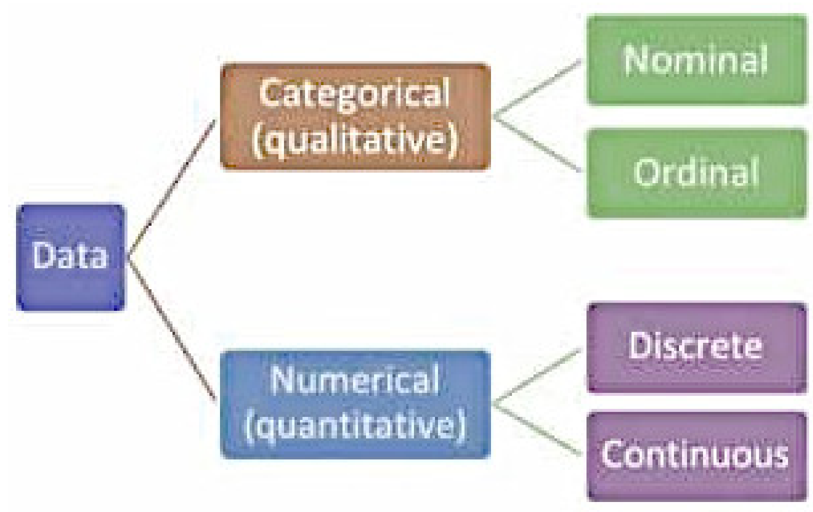
- general estimation of random and non-random parameters
- linear estimation of random and non-random parameters



Big Picture

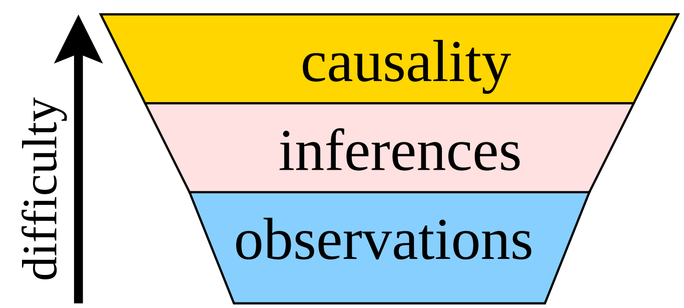
Systems appear random

- uncertainty
- complexity
- limited knowledge
- measurement noise
  → statistical description
  → random variables and processes



Statistical analysis

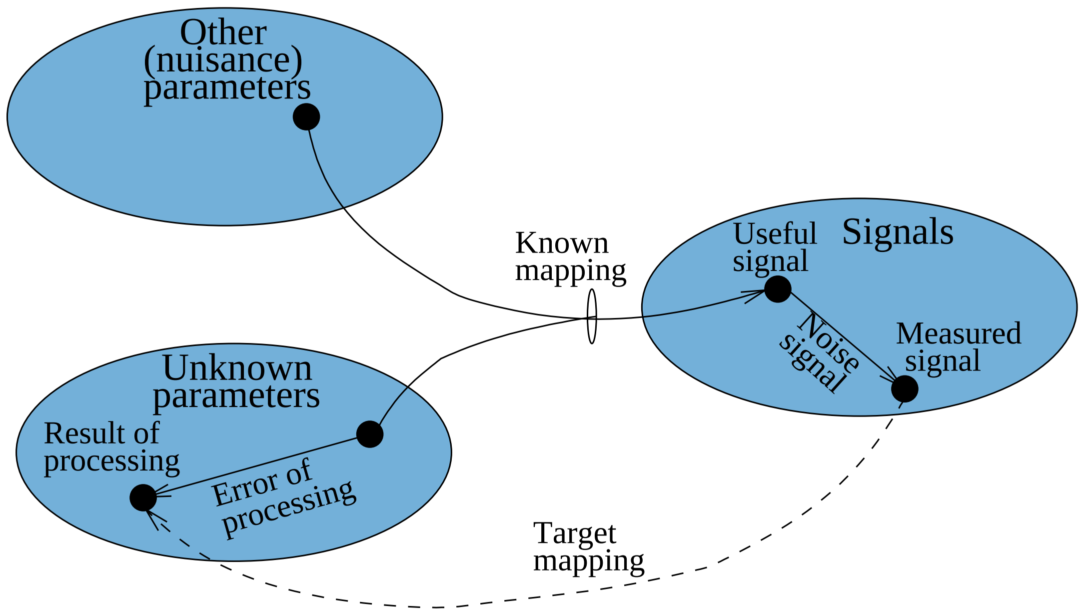
- descriptive inferences
  → parametric and non-parametric statistics
- statistical inferences
  → model-based and model-free
- causal inferences
  → cause-effect relationships



Key factors

- what is known, available measurements, task/application, batch or streaming

STATISTICAL INFERENCE



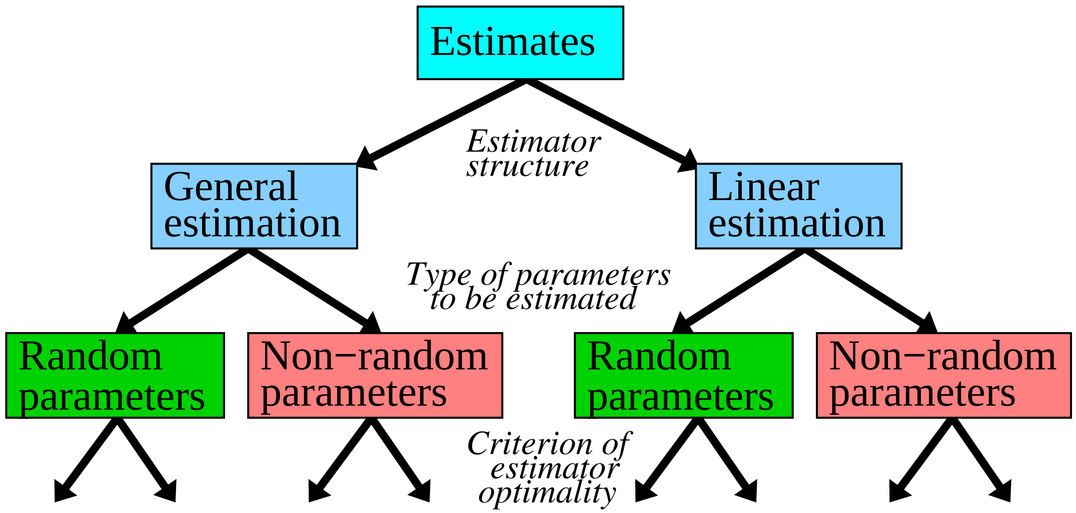
Two types of problems

- Detection: hypothesis testing, “*Which value from the set?*”
- Estimation: point estimation, “*How big is the value?*”
  → also interval estimation, posterior estimation etc.

Invert mapping?

- may not be easy to obtain
- not optimum, may amplify measurement noise

Parameter Estimation



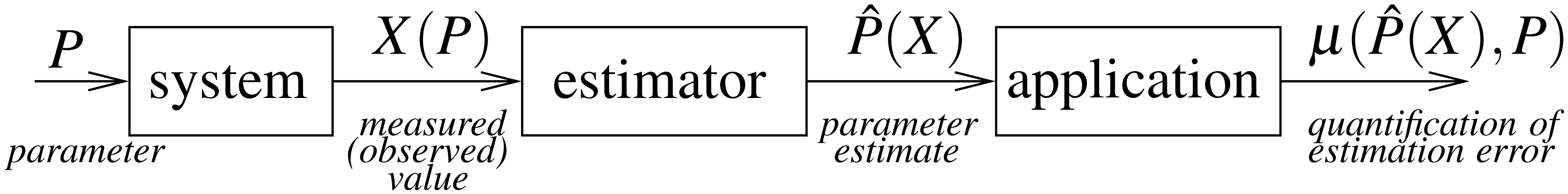
Parameters to be estimated

- parameters are unknown, so must be estimated
- if their prior distribution is known, they appear as being random
- no prior distribution, treat them as non-random (deterministic)
- often only some statistics known (mean and variance)

Estimator structure

- general (unconstrained)
- linear (linear filter)

General Estimation of Random Parameters



Requires

- knowing prior $f_P(p)$ or $\Pr_P(p)$
- knowing statistical dependence of X on P: $f_{X|P}(x|p)$ or $\Pr_{X|P}(x|p)$
- quantifying the estimation error $(\hat{P} - P)$ as $\mu(\hat{P}, P)$

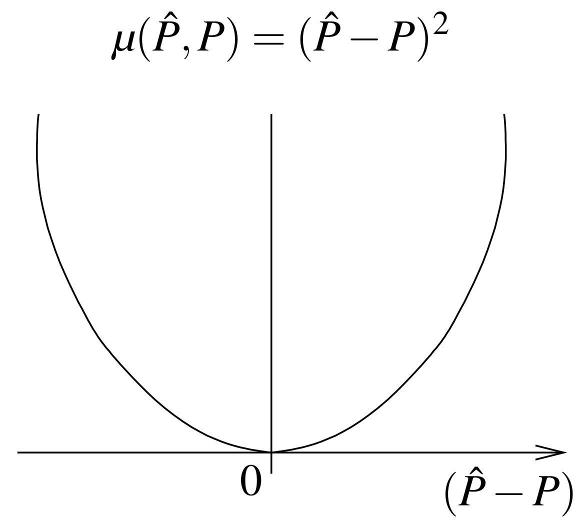
Optimum estimator

- minimize the risk $\mathrm{E}_{X,P}\left[\mu(\hat{P}(X), P)\right]$

$$\hat{P}_{\mathrm{opt}} = \mathrm{argmin}_{\hat{P}(x)} \mathrm{E}_{X,P}\left[\mu(\hat{P}(x), P)|X = x\right]$$

- Minimum Mean Square Error (MMSE): $\mathrm{E}_{X,P}\left[(\hat{P}(X) - P)^2\right]$

$$\hat{P}_{\mathrm{MMSE}}(x) = \mathrm{E}_P[P|X = x] = \int_{\{P\}} p\, f_{P|X}(p|x)\, \mathrm{d}p$$



General Estimation of Random Parameters (cont.)

Properties of MMSE estimator

- it is unbiased
- estimation error uncorrelated with any function of X (orthogonality)

Gauss-Markov theorem

- $\mathbf{P}$ is vector of parameters, $\mathbf{X}$ is vector of measurements
- if $\mathbf{P}$ and $\mathbf{X}$ are jointly Gaussian with means $\bar{\mathbf{P}}$ and $\bar{\mathbf{X}}$, and the covariance matrices $\mathrm{var}[\mathbf{P}]$ and $\mathrm{var}[\mathbf{X}]$, then

$$\hat{\mathbf{P}}_{\mathrm{MMSE}}(\mathbf{x}) = \bar{\mathbf{P}} + \mathbf{H}(\mathbf{x} - \bar{\mathbf{X}}) \quad \text{where} \quad \mathbf{H} = \mathrm{cov}[\mathbf{P}, \mathbf{X}]\,\mathrm{var}^{-1}[\mathbf{X}]$$

Nuisance parameters

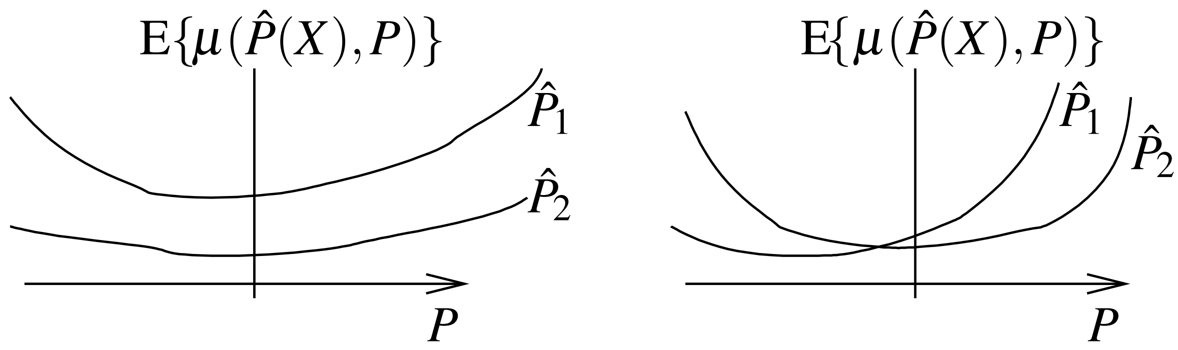
- estimated and then ignored, or averaged out

Maximum a posteriori probability (MAP) estimator

- for discrete P: $\mu(\hat{P}, P) = 1$, if $\hat{P} \neq P$, and 0 otherwise
- then the risk is equal to the probability of error

$$\hat{P}_{\mathrm{MAP}}(X) = \mathrm{argmax}_{p_i} \Pr(P = p_i | X = x)$$

General Estimation of Non-Random Parameters



Caveat

- the estimator may not exist, or is difficult to find

Minimum Variance Unbiased (MVUB) estimator

- unbiased and minimizes the MSE equal to variance of $\hat{P}$

Cramer-Rao lower bound

- lower bounds variance of unbiased estimator of non-random parameter
- the lower bounds is achieved by efficient estimators
- for consistent estimators, variance decreases with # measurements
- idea:

allow a small bias to further reduce the variance?

General Estimation of Non-Random Parameters (cont.)

Maximum Likelihood (ML) estimator

$$\hat{P}_{\mathrm{ML}}(X) = \mathrm{argmax}_{\hat{P}}\, f_X(x, \hat{P}) \quad \text{or} \quad \hat{P}_{\mathrm{ML}}(X) = \mathrm{argmax}_{\hat{P}}\, \mathrm{Pr}_X\big(X = x, \hat{P}\big)$$

Properties of ML estimator

- if efficient estimate exists, then it is ML estimate
- if efficient estimate does not exists, then ML estimate is neither guaranteed to have minimum variance, nor to be unbiased
- asymptotically unbiased and efficient, and invariant to any function $g(P)$

Least Squares (LS) estimation of non-random parameter

- if cannot obtain distribution of measurements, but can approximate $X \approx g(P)$
- LS estimator corresponds to ML estimator if noise is AWGN

$$\hat{P}_{\mathrm{opt}}(X) = \mathrm{argmin}_{\hat{P}}\, \mu\Big(X, g(\hat{P})\Big), \qquad \mu\Big(X, g(\hat{P})\Big) = \sum_{i=1}^{N} v_i \Big(X_i - g(\hat{P})\Big)^2$$

Statistical moments based estimation

- no regard to optimality, but simple and low complexity
- unbiased and consistent estimate of the n-th moment: $\hat{g}_n(P) = \frac{1}{N}\sum_{i=1}^{N} x^n(i)$
- the estimate is the inverse

$$\hat{P} = g_n^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x^n(i)\right)$$

Linear Estimation

Advantage

- easy to implement (linear filter)
- only knowledge of basic statistics required

Linear MMSE estimator

- cf. Gauss-Markov theorem

Properties of LMMSE estimator

- it is unbiased
- estimation error and measurements are uncorrelated (orthogonality)
- estimation error and estimates are uncorrelated
- if the estimator is linear, unbiased and orthogonal, then it is LMMSE

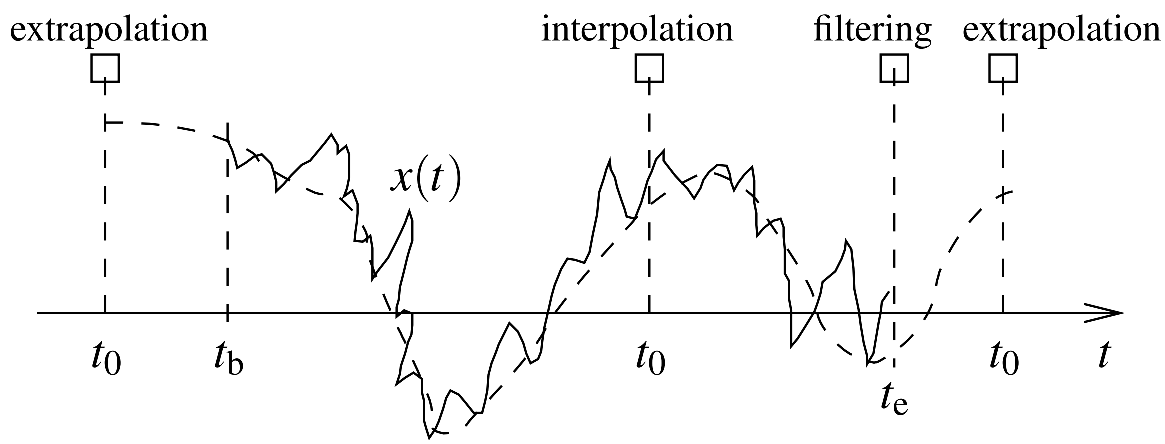
Non-random parameters

- for linear estimator that is also unbiased

$$\hat{\mathbf{P}} = \mathbf{H}(\mathbf{X}-\mathbf{r}), \quad \text{where } \bar{\mathbf{X}} = \mathbf{D}\mathbf{P}+\mathbf{r}, \text{ and, } \mathbf{H}\mathbf{D} = \mathbf{I}$$

- Best Linear Unbiased Estimator (BLUE) minimizes covariance matrix of estimation error

Estimating Time-Varying Parameters



Tasks

- extrapolation: $t_0 < t_\mathrm{b}$ or $t_0 > t_\mathrm{e}$
- interpolation: $t_\mathrm{b} \leq t_0 < t_\mathrm{e}$
- filtration: $t_0 = t_\mathrm{e}$

Wiener filter

- $X(t)$ and $P(t)$ are stationary and Gaussian
- auto-covariances and cross-covariance of $X(t)$ and $P(t)$ known
- LMMSE of $P(t)$

Kalman filter

- generalizes Wiener filter to non-stationary Gaussian signals
- fast adaptation to changes in statistics
- several modifications for non-Gaussian signals

Take-Home Messages

Key points

- well-established and understood
- estimators are optimum, interpretable, computationally efficient
- general and linear estimators
- estimating random and non-random parameters

Machine learning

- more universal, no assumptions
- replace model with labeled samples
- exchange efficiency and interpretability for performance

$\Rightarrow$ need for incorporating ideas from estimation theory

In the paper

- more details
- more explanations
- number of examples
- list of textbooks

Topics not covered

- statistical filtering
- Bayesian inference
- adaptive estimation
- interval estimation

Thank you!

pavelloskot@intl.zju.edu.cn