Statistical signal processing
5CTA0

Estimation theory - Bayesian estimation

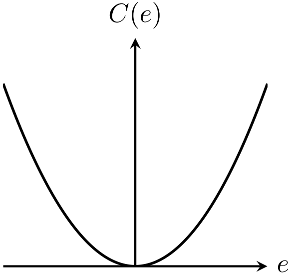
Bayesian estimation

- Classical estimation: parameter θ is a deterministic but unknown constant
- Bayesian estimation: parameter θ is a random variable with known prior probability
- Cannot evaluate performance using mean and variance as criteria
- Instead, define a cost $C(\mathbf{e})$ which is a function of the estimation error $\mathbf{e} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$
- Baysian estimators are found by minimizing the average cost or Bayes risk:

$$\mathrm{E}[C(\mathbf{e})] = \int \int C(\mathbf{e}) p(\boldsymbol{\theta}, \mathbf{x}) d\boldsymbol{\theta} d\mathbf{x}$$

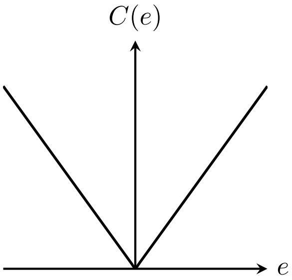
Cost: quadratic error

$$C(\mathbf{e}) = \|\mathbf{e}\|_2^2 = \sum_{k=0}^{K-1} e_k^2.$$



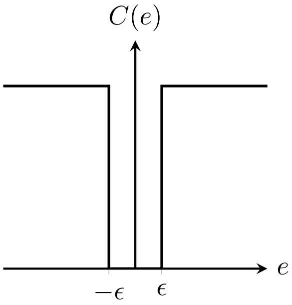
Cost: absolute error

$$C(\mathbf{e}) = \|\mathbf{e}\|_1 = \sum_{k=0}^{K-1} |e_k|.$$



Cost: uniform error

$$C(\mathbf{e}) = \begin{cases} 0 & \|\mathbf{e}\|_\infty = \max\limits_{0 \le k \le K-1} |e_k| < \epsilon, \\ 1 & \text{otherwise.} \end{cases}$$



Bayes estimate

- Bayes risk:

$$\begin{aligned}
\mathrm{E}[C(\mathbf{e})] &= \int\int C(\mathbf{e})p(\boldsymbol{\theta},\mathbf{x})d\boldsymbol{\theta}d\mathbf{x} \\
&= \int\int C(\mathbf{e})p(\boldsymbol{\theta}|\mathbf{x})p(\mathbf{x})d\boldsymbol{\theta}d\mathbf{x} \\
&= \int\left[\int C(\mathbf{e})p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}\right]p(\mathbf{x})d\mathbf{x}
\end{aligned}$$

- Minimize risk

$$\hat{\boldsymbol{\theta}} = \arg\min_{\hat{\boldsymbol{\theta}}} \int\left[\int C(\mathbf{e})p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}\right]p(\mathbf{x})d\mathbf{x} = \arg\min_{\hat{\boldsymbol{\theta}}} \int C(\mathbf{e})p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}$$

Quadratic error - minimum mean square error

- Gradient:

$$\frac{\partial}{\partial \hat{\boldsymbol{\theta}}} \int \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} = 2 \int (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}$$

- MMSE estimator:

$$\hat{\boldsymbol{\theta}}_{\mathsf{MMSE}} = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}$$

- The minimum mean square error estimate $\hat{\boldsymbol{\theta}}_{\mathsf{MAE}}$ is the *mean* of the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$

Absolute error - minimum absolute error

- kth component of gradient:

$$\frac{\partial}{\partial \theta_k} \int \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_1 p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} = \int \operatorname{sign}(\hat{\theta}_k - \theta_k) p(\theta_k|\mathbf{x}) d\theta_k,$$

- $$\operatorname{sign}(x) = \begin{cases} 1 & x \geq 0, \\ -1 & x < 0. \end{cases}$$

- $$\int \operatorname{sign}(\hat{\theta}_k - \theta_k) p(\theta_k|\mathbf{x}) d\theta_k = \int_{-\infty}^{\hat{\theta}_k} p(\theta_k|\mathbf{x}) d\theta_k - \int_{\hat{\theta}_k}^{\infty} p(\theta_k|\mathbf{x}) d\theta_k.$$

- The minimum absolute error estimate $\hat{\boldsymbol{\theta}}_{\mathsf{MAE}}$ is the *median* of the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$.

Uniform error - maximum a posteriori estimate

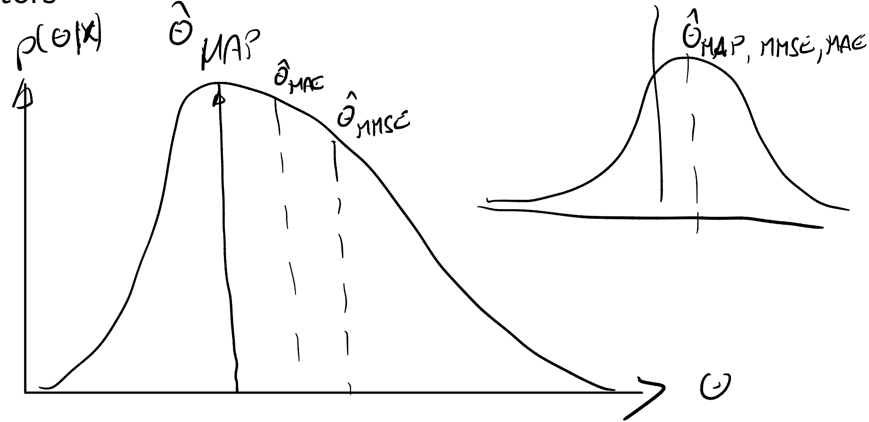
- $$\int C(\mathbf{e})p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} = 1 - \int_{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_\infty<\epsilon} p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}.$$

- $$\int_{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_\infty<\epsilon} p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}$$

- $$\hat{\boldsymbol{\theta}}_{\mathsf{MAP}} = \arg\max_{\boldsymbol{\theta}\in\mathbb{R}^K} p(\boldsymbol{\theta}|\mathbf{x}).$$

- The minimum uniform error estimate $\hat{\boldsymbol{\theta}}_{\mathsf{MAP}}$ is the *max* of the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$.

Bayesian estimators



Example

$$p(\theta|x) = \frac{p(x|\theta)\,p(\theta)}{p(x)}$$

- Prior probability:

$$p(\theta) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{1}{2\sigma_\theta^2}(\theta-\mu_\theta)^2\right)$$

- Likelihood:

$$p(\mathbf{x}|\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-\theta)^2\right)$$

-

$$p(\theta|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma_{\theta|\mathbf{x}}^2}} \exp\left(-\frac{1}{2\sigma_{\theta|\mathbf{x}}^2}\left(\theta-\mu_{\theta|\mathbf{x}}\right)^2\right)$$

$$\sigma_{\theta|\mathbf{x}}^2 = \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_\theta^2}} \qquad \mu_{\theta|\mathbf{x}} = \left(\frac{N\bar{x}}{\sigma^2}+\frac{\mu_\theta}{\sigma_\theta^2}\right)\sigma_{\theta|\mathbf{x}}^2$$

Example