Introduction to Estimation Theory

ECE 6279: Spatial Array Processing
Fall 2013
Lecture 20

Prof. Aaron D. Lanterman

School of Electrical & Computer Engineering
Georgia Institute of Technology
AL: 404-385-2548
<lanterma@ece.gatech.edu>

References

- J&D: Section 6.2.4
- Van Trees – “Detection Estimation, and Modulation Theory: Part I” (many parts out of date, but still the best)
- Vince Poor – “An Introduction to Signal Detection and Estimation” (insanely mathematical)
- Stephen Kay – “Fundamentals of Statistical Signal Processing” – two volumes (a bit sprawling)

Setup

- **Model measured data y as a realization of a random variable $\underline{y}$**
- **Assume $\underline{y}$ has a density that is a function of desired parameter(s) ξ**

Vince Poor: $p_{\xi}(y)$

J&D, Van Trees: $p(y \mid \xi)$ **(misleading)**

Aaron's ECE7251: $p(y;\xi)$

Canonical Example: Simple Gaussian

- **Example: i.i.d. samples of a real scalar Gaussian**

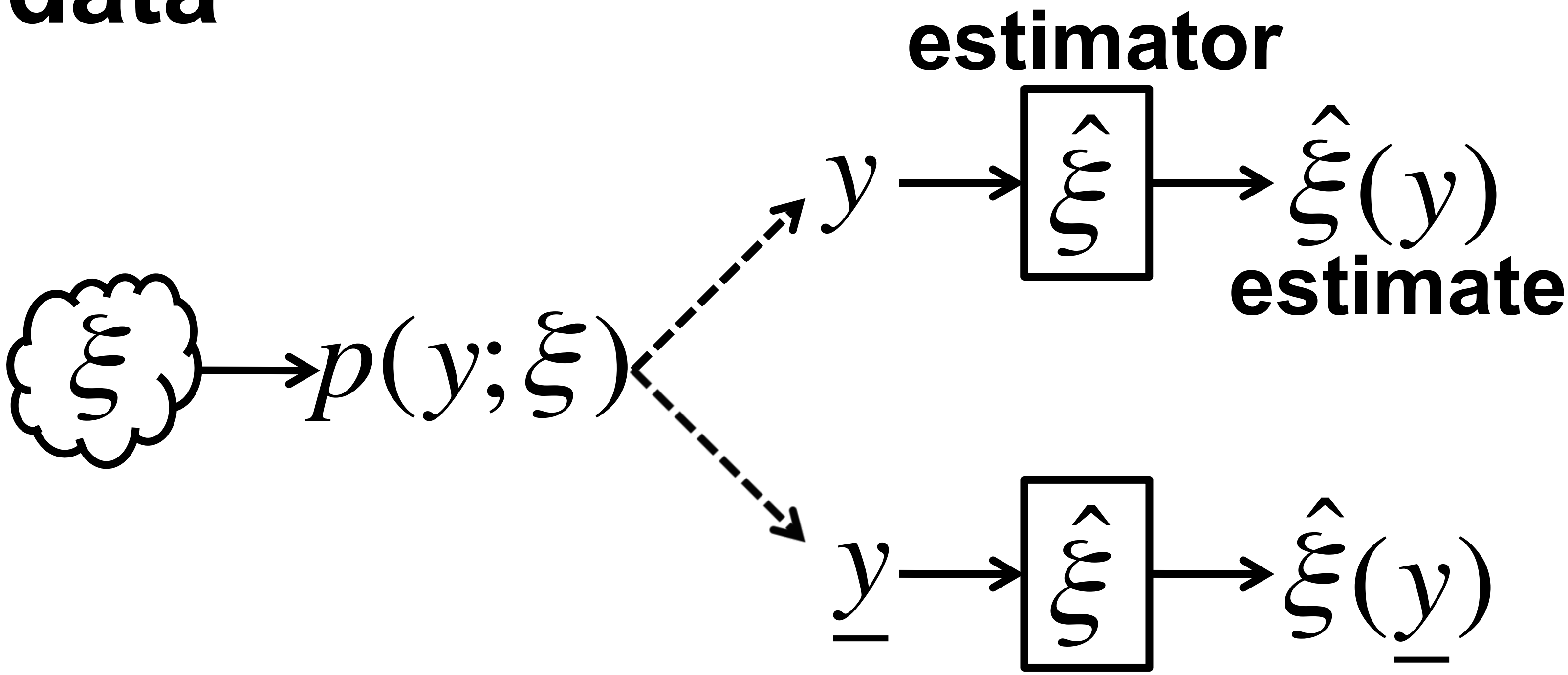
$$\xi = (\mu, \sigma^2)$$

$$p(y;\xi) = \prod_{l=0}^{L-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{[y(l)-\mu]^2}{2\sigma^2} \right\}$$

$$y = \{y(0), \ldots, y(L-1)\}$$

Big Picture

- Estimator $\hat{\xi}$ is a function of the data



Maximum Likelihood Estimators

- **Makes intuitive sense, but not necessarily magical:**

$$\hat{\xi}_{ML}(y) = \arg\max_{\xi} p(y;\xi)$$

$$= \arg\max_{\xi} \underbrace{\ln p(y;\xi)}_{\ell(\xi)}$$

(up to an additive constant)

$$= \arg\max_{\xi} \ell(\xi)$$

Loglikelihood for Gaussian Example

$$p(y;\xi) = \prod_{l=0}^{L-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{[y(l)-\mu]^2}{2\sigma^2} \right\}$$

$$\ell(\xi) = \cancel{-\frac{L}{2}\ln 2\pi} - \frac{L}{2}\ln\sigma^2 - \sum_{l=0}^{L-1} \frac{[y(l)-\mu]^2}{2\sigma^2}$$

Customary to omit terms that do not contain parameters

ML Estimator for Mean

$$\ell(\xi) = -\frac{L}{2}\ln\sigma^2 - \sum_{l=0}^{L-1}\frac{[y(l)-\mu]^2}{2\sigma^2}$$

- "Usual" trick "usually" works:

$$\frac{\partial}{\partial\mu}\ell(\xi) = \sum_{l=0}^{L-1}\frac{y(l)-\mu}{\sigma^2} = 0$$

$$\sum_{l=0}^{L-1} y(l) = L\mu \Rightarrow \hat{\mu}_{ML}(y) = \frac{1}{L}\left[\sum_{l=0}^{L-1} y(l)\right]$$

ML Estimator for Variance (1)

$$\ell(\xi) = -\frac{L}{2}\ln\sigma^2 - \sum_{l=0}^{L-1}\frac{[y(l)-\mu]^2}{2\sigma^2}$$

- "Usual" trick "usually" works:

$$\frac{\partial}{\partial\sigma^2}\ell(\xi) = -\frac{L}{2\sigma^2} + \sum_{l=0}^{L-1}\frac{[y(l)-\mu]^2}{2(\sigma^2)^2} = 0$$

$$L\sigma^2 = \sum_{l=0}^{L-1}[y(l)-\mu]^2$$

ML Estimator for Variance (2)

$$\sigma^2 = \frac{1}{L}\sum_{l=0}^{L-1}[y(l)-\mu]^2$$

- **Here we lucked out:**

$$\hat{\sigma}^2_{ML}(y) = \frac{1}{L}\sum_{l=0}^{L-1}[y(l)-\hat{\mu}_{ML}(y)]^2$$

$$\hat{\mu}_{ML}(y) = \frac{1}{L}\left[\sum_{l=0}^{L-1} y(l)\right]$$

Bias

$$b(\xi) = E_{\xi}[\hat{\xi}(\underline{y})] - \xi$$

- **If $b(\xi) = 0$, i.e. $E_{\xi}[\hat{\xi}(\underline{y})] = \xi$, we say estimator is unbiased**

Bias of ML Estimate of the Mean

$$b(\mu) = E_\xi[\hat{\mu}(\underline{y})] - \mu$$

$$= E_\xi\left[\frac{1}{L}\sum_{l=0}^{L-1}\underline{y}(l)\right] - \mu$$

$$= \frac{1}{L}\sum_{l=0}^{L-1}E_\xi[\underline{y}(l)] - \mu = \frac{1}{L}L\mu - \mu = 0$$

Bias of ML Estimate of the Variance (1)

$$b(\sigma^2) = E_\xi[\hat{\sigma}^2(\underline{y})] - \sigma^2$$

$$E_\xi[\hat{\sigma}^2(\underline{y})] = E_\xi\left\{\frac{1}{L}\sum_{l=0}^{L-1}[\underline{y}(l) - \hat{\mu}_{ML}(\underline{y})]^2\right\}$$

$$= E_\xi\left\{\frac{1}{L}\sum_{l=0}^{L-1}\left[\underline{y}(l) - \frac{1}{L}\sum_{k=0}^{L-1}\underline{y}(k)\right]^2\right\}$$

Bias of ML Estimate of the Variance (2)

$$E_\xi[\hat{\sigma}^2(\underline{y})] = E_\xi\left\{\frac{1}{L}\sum_{l=0}^{L-1}\left[\underline{y}(l) - \frac{1}{L}\sum_{k=0}^{L-1}\underline{y}(k)\right]^2\right\}$$

$$= \frac{1}{L}E_\xi\left\{\sum_{l=0}^{L-1}\left[\underline{y}^2(l) - 2\underline{y}(l)\left(\frac{1}{L}\sum_{k=0}^{L-1}\underline{y}(k)\right) + \left(\frac{1}{L}\sum_{k=0}^{L-1}\underline{y}(k)\right)^2\right]\right\}$$

First Term

$$E_{\xi}\left\{\sum_{l=0}^{L-1} \underline{y}^2(l)\right\} = \sum_{l=0}^{L-1} E_{\xi}\left\{\underline{y}^2(l)\right\}$$

$$= L(\sigma^2 + \mu^2)$$

Second Term (1)

$$-2E_{\xi}\left\{\sum_{l=0}^{L-1}\underline{y}(l)\left(\frac{1}{L}\sum_{k=0}^{L-1}\underline{y}(k)\right)\right\}$$

$$=-\frac{2}{L}\sum_{l=0}^{L-1}\sum_{k=0}^{L-1}E_{\xi}\{\underline{y}(l)\underline{y}(k)\}$$

$$=-\frac{2}{L}\sum_{l=0}^{L-1}\left(E_{\xi}\{\underline{y}^{2}(l)\}+\sum_{k\neq l}E_{\xi}\{\underline{y}(l)\underline{y}(k)\}\right)$$

Second Term (2)

$$-\frac{2}{L}\sum_{l=0}^{L-1}\left(E_{\xi}\{\underline{y}^2(l)\}+\sum_{k\neq l}E_{\xi}\{\underline{y}(l)\underline{y}(k)\}\right)$$

$$=-2\left([\sigma^2+\mu^2]+[L-1]\mu^2\right)$$

$$=-2(\sigma^2+L\mu^2)$$

Third Term (1)

$$E_{\xi}\left\{\sum_{l=0}^{L-1}\left(\frac{1}{L}\sum_{k=0}^{L-1}y(k)\right)^2\right\}$$

$$= E_{\xi}\left\{\sum_{l=0}^{L-1}\left(\frac{1}{L}\sum_{k=0}^{L-1}y(k)\right)\left(\frac{1}{L}\sum_{j=0}^{L-1}y(j)\right)\right\}$$

Third Term (2)

$$E_\xi\left\{\sum_{l=0}^{L-1}\left(\frac{1}{L}\sum_{k=0}^{L-1}y(k)\right)\left(\frac{1}{L}\sum_{j=0}^{L-1}y(j)\right)\right\}$$

$$=\frac{1}{L}E_\xi\left\{\sum_{k=0}^{L-1}\sum_{j=0}^{L-1}y(k)y(j)\right\}$$

$$=\frac{1}{L}\left\{\sum_{k=0}^{L-1}\left(E_\xi[y^2(k)]+\sum_{j\neq k}E_\xi[y(k)y(j)]\right)\right\}$$

Third Term (3)

$$\frac{1}{L}\left\{\sum_{k=0}^{L-1}\left(E_{\xi}[y^2(k)]+\sum_{j\neq k}E_{\xi}[y(k)y(j)]\right)\right\}$$

$$=(\sigma^2+\mu^2)+(L-1)\mu^2$$

$$=\sigma^2+L\mu^2$$

Putting it Back Together

$$E_\xi[\hat{\sigma}^2(\underline{y})] =$$

$$\frac{1}{L} E_\xi \left\{ \sum_{l=0}^{L-1} \left[y^2(l) - 2y(l)\left(\frac{1}{L}\sum_{k=0}^{L-1} y(k)\right) + \left(\frac{1}{L}\sum_{k=0}^{L-1} y(k)\right)^2 \right] \right\}$$

$$= \frac{1}{L}\left\{ L(\sigma^2 + \mu^2) - 2(\sigma^2 + L\mu^2) + (\sigma^2 + L\mu^2) \right\}$$

$$= \frac{L-1}{L}\sigma^2$$

Almost Forgot What We Were Computing

$$b(\sigma^2) = E_\xi[\hat{\sigma}^2(\underline{y})] - \sigma^2$$

$$= \frac{L-1}{L}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{L}$$

- **Notice** $b(\sigma^2) \to 0$ **as** $L \to \infty$**; we say estimator is asymptotically unbiased**

Tweaking the Estimator

- **Previous computations suggest an unbiased estimator:**

$$\hat{\sigma}^2_{UB}(y) = \frac{1}{L-1}\sum_{l=0}^{L-1}[y(l) - \hat{\mu}_{ML}(y)]^2$$

Variance of ML Est. of Gaussian Mean

$$\mathrm{var}_\xi[\hat{\mu}_{ML}(\underline{y})] = \mathrm{var}_\xi\left[\frac{1}{L}\sum_{l=0}^{L-1}\underline{y}(l)\right]$$

$$= \frac{1}{L^2}\sum_{l=0}^{L-1}\mathrm{var}_\xi[\underline{y}(l)] = \frac{1}{L^2}L\sigma^2 = \frac{\sigma^2}{L}$$

Recall: $b_\xi(\hat{\mu}_{ML}) = 0$

Estimators of Gaussian Variance

$$\hat{\sigma}^2_{ML}(y) = \frac{1}{L}\sum_{l=0}^{L-1}[y(l) - \hat{\mu}_{ML}(y)]^2$$

$$b_\xi(\hat{\sigma}^2_{ML}) = -\sigma^2 / L$$

$$\sigma^2_{UB}(y) = \frac{1}{L-1}\sum_{l=0}^{L-1}[y(l) - \hat{\mu}_{ML}(y)]^2$$

$$b_\xi(\hat{\sigma}^2_{UB}) = 0$$

Vars. of Estimators of Gaussian Vars.

$$\mathrm{var}_{\xi}[\hat{\sigma}^2_{ML}(\underline{y})] = 2\frac{(\sigma^2)^2}{L}\left(\frac{L-1}{L}\right)$$

$$\mathrm{var}_{\xi}[\hat{\sigma}^2_{UB}(\underline{y})] = 2\frac{(\sigma^2)^2}{L}\left(\frac{L}{L-1}\right)$$

(from estimation theory notes by Al Hero, U. of Michigan)

Notion of Mean Squared Error

We will explore the idea of bias/ variance tradeoffs in the MSE:

$$E_{\xi}[(\hat{\xi} - \xi)^2] = \mathrm{var}_{\xi}(\hat{\xi}) + b_{\xi}^2(\hat{\xi})$$

Bias Matrices

$$\mathbf{b}_\xi(\hat{\xi}) = E_\xi[\hat{\xi}(\underline{y})] - \xi$$

$$= E_\xi\left\{\begin{bmatrix} \hat{\mu}(\underline{y}) \\ \hat{\sigma}^2(\underline{y}) \end{bmatrix}\right\} - \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$$

Bias and Covariance Matrices

$$\mathrm{cov}_\xi[\hat{\xi}(\underline{y})] = E_\xi[\hat{\xi}\hat{\xi}^{\mathrm{T}}] - E_\xi[\hat{\xi}]E_\xi[\hat{\xi}^{\mathrm{T}}]$$

$$= \begin{bmatrix} E[\hat{\mu}^2] & E[\hat{\mu}\hat{\sigma}^2] \\ E[\hat{\mu}\hat{\sigma}^2] & E[(\hat{\sigma}^2)^2] \end{bmatrix}$$

$$- \begin{bmatrix} E^2[\hat{\mu}] & E[\hat{\mu}]E[\hat{\sigma}^2] \\ E[\hat{\mu}]E[\hat{\sigma}^2] & E^2[\hat{\sigma}^2] \end{bmatrix}$$

Decomposition of MSE Matrix

$$E_\xi\{[\hat{\xi}(\underline{y})-\xi][\hat{\xi}(\underline{y})-\xi]^T]\}$$

$$=E_\xi[\hat{\xi}(\underline{y})\hat{\xi}^T(\underline{y})-\hat{\xi}(\underline{y})\xi^T-\xi\hat{\xi}^T(\underline{y})+\xi\xi^T]$$

$$=E[\hat{\xi}\hat{\xi}^T]-E[\hat{\xi}]\xi^T-\xi E[\hat{\xi}^T]+\xi\xi^T$$
$$+E[\hat{\xi}]E[\hat{\xi}^T]-E[\hat{\xi}]E[\hat{\xi}^T]$$

$$=E[\hat{\xi}\hat{\xi}^T]-E[\hat{\xi}]E[\hat{\xi}^T]$$
$$+E[\hat{\xi}]E[\hat{\xi}^T]-E[\hat{\xi}]\xi^T-\xi E[\hat{\xi}^T]+\xi\xi^T$$

Decomposition of MSE Matrix

$$E_{\boldsymbol{\xi}}\{[\hat{\boldsymbol{\xi}}(\underline{y})-\boldsymbol{\xi}][\hat{\boldsymbol{\xi}}(\underline{y})-\boldsymbol{\xi}]^T]\}$$

$$= E[\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}^T]-E[\hat{\boldsymbol{\xi}}]E[\hat{\boldsymbol{\xi}}^T]$$

$$+\underbrace{E[\hat{\boldsymbol{\xi}}]E[\hat{\boldsymbol{\xi}}^T]-E[\hat{\boldsymbol{\xi}}]\boldsymbol{\xi}^T-\boldsymbol{\xi}E[\hat{\boldsymbol{\xi}}^T]+\boldsymbol{\xi}\boldsymbol{\xi}^T}_{(E[\hat{\boldsymbol{\xi}}]-\boldsymbol{\xi})(E[\hat{\boldsymbol{\xi}}]-\boldsymbol{\xi})^T}$$

$$= \mathrm{cov}_{\boldsymbol{\xi}}(\hat{\boldsymbol{\xi}})+\mathbf{b}_{\boldsymbol{\xi}}(\hat{\boldsymbol{\xi}})\mathbf{b}_{\boldsymbol{\xi}}^T(\hat{\boldsymbol{\xi}})$$

In scalar case: $E_{\xi}[(\hat{\xi}-\xi)^2]=\mathrm{var}_{\xi}(\hat{\xi})+b_{\xi}^2(\hat{\xi})$

Composite MSE

$$MSE = \mathrm{tr}\left\{ E_{\xi}\{[\hat{\xi}(\underline{y}) - \xi][\hat{\xi}(\underline{y}) - \xi]^T\}\right\}$$

$$= \mathrm{tr}\{\mathrm{cov}_{\xi}(\hat{\xi}) + \mathbf{b}_{\xi}(\hat{\xi})\mathbf{b}_{\xi}^T(\hat{\xi})\}$$

$$= \mathrm{tr}\{\mathrm{cov}_{\xi}(\hat{\xi})\} + \mathrm{tr}\{\mathbf{b}_{\xi}(\hat{\xi})\mathbf{b}_{\xi}^T(\hat{\xi})\}$$