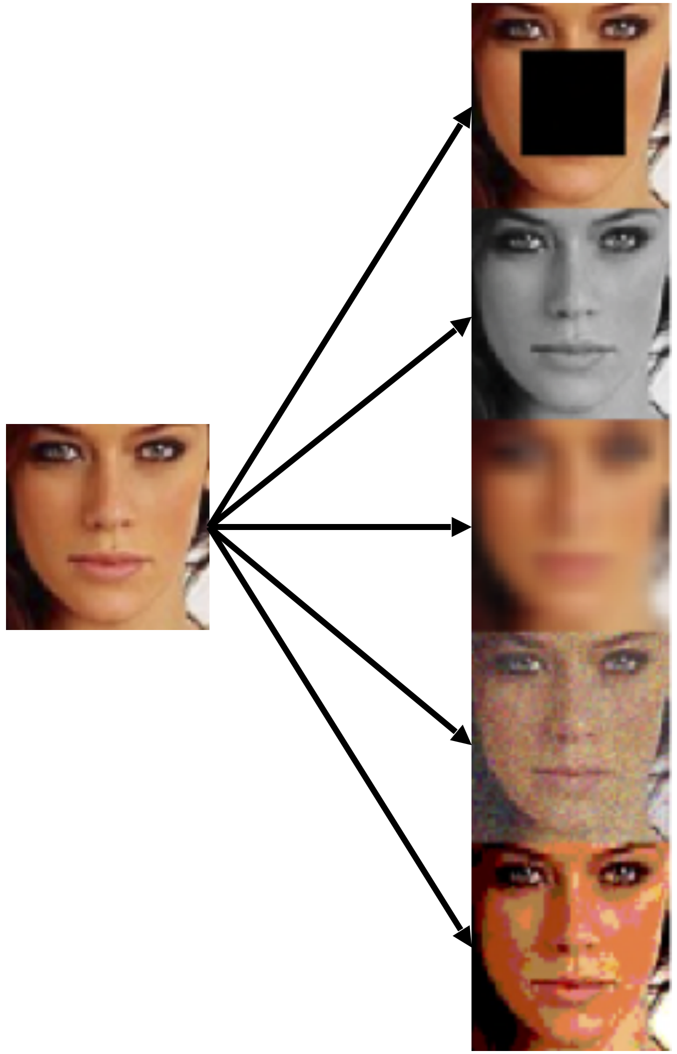
From Statistical Estimation and Generative Modeling to Inverse Problems

LECTURE 3: STATISTICAL ESTIMATION

Prologue

In our three lectures we focus on the following subjects:

- Statistical Estimation methods (Parameter Estimation)
  Measure X but would like to know (estimate) Z
- Random data are described by probability densities
  Not convenient for modern datasets: Usually multi-dimensional but recide on low dimensional surfaces (manifolds)
  Alternative description with Generative Models
- Modification of Statistical Estimation methods to accept generative models in place of probability densities
  Application to inverse problems (mostly from Computer Vision)



Inpainting

Colorization

Super-resolution

De-noising

De-quantization

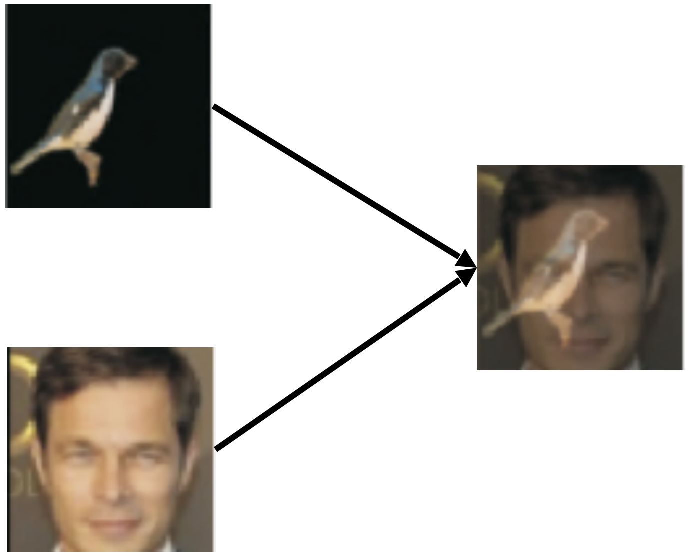


Image Separation

Statistical Estimation

- Problem Definition

- Bayesian Estimation
  - Optimum solution for general cost function
  - Solution for the MSE, MAE, AP

- Non-Bayesian Estimation
  - The ML estimator
  - Asymptotic optimality

Problem Definition

We are given vector X (measurements) but interested in vector Z !
Can we estimate Z from X ?

Z and X must be related otherwise impossible

The relationship can be "soft" (not deterministic) and expressed with a joint probability density f(X,Z)

So Z and X are random and have a random relationship captured by f(X,Z)

The density f(X,Z) also expresses prior knowledge about each individual random quantity Z and X

Prior knowledge about individual X and Z is captured by the marginal densities $\mathsf{g}(X)$ and $\mathsf{h}(Z)$ where

$$\mathsf{g}(X) = \int \mathsf{f}(X, Z)\, dZ, \quad \mathsf{h}(Z) = \int \mathsf{f}(X, Z)\, dX$$

When measurements X become available, knowledge about Z changes to

$$\mathsf{f}(Z|X) = \frac{\mathsf{f}(X, Z)}{\mathsf{g}(X)} = \frac{\mathsf{f}(X, Z)}{\int \mathsf{f}(X, Z)\, dZ}$$

If X and Z independent then $\mathsf{f}(X{,}Z){=}\mathsf{g}(X)\mathsf{h}(Z)$ and

$$\mathsf{f}(Z|X) = \frac{\mathsf{f}(X, Z)}{\mathsf{g}(X)} = \frac{\mathsf{g}(X)\mathsf{h}(Z)}{\mathsf{g}(X)} = \mathsf{h}(Z)$$

X has no information about Z

What is an estimator ?

Given f(X, Z) and measurements X suppose we select “estimate” $\hat{Z}$

If measurements X do not change no reason to change $\hat{Z}$

Estimate $\hat{Z} = \hat{Z}(X)$ is any deterministic function of X

There are “good” estimators and there are “bad” estimators

Statistical Estimation theory identifies the “best”

Bayesian Estimation

There is cost function $\mathsf{C}(\cdot,\cdot)$ such that if
true value is Z and estimate $\hat{Z}$ then cost is $\mathsf{C}(\hat{Z}, Z)$

Interested in average cost

$$\mathscr{C}(\hat{Z}) = \mathbb{E}_{X,Z}\big[\mathsf{C}\big(\hat{Z}(X), Z\big)\big] = \iint \mathsf{C}\big(\hat{Z}(X), Z\big)\mathsf{f}(X, Z)\, dX\, dZ$$

Depends on the estimator function $\hat{Z}(X)$

We like to minimize it

$$\min_{\hat{Z}} \mathscr{C}(\hat{Z}) = \min_{\hat{Z}} \iint \mathsf{C}\big(\hat{Z}(X), Z\big)\mathsf{f}(X, Z)\, dX\, dZ$$

Prior knowledge about X and Z

$$\mathsf{g}(X) = \int \mathsf{f}(X, Z)\, dZ, \quad \mathsf{h}(Z) = \int \mathsf{f}(X, Z)\, dX$$

After we take measurements X our knowledge about Z becomes

$$\mathsf{f}(Z|X) = \frac{\mathsf{f}(X, Z)}{\mathsf{g}(X)} = \frac{\mathsf{f}(X, Z)}{\int \mathsf{f}(X, Z)\, dZ}$$

Posterior probability density

and we can write $\mathsf{f}(X, Z) = \mathsf{f}(Z|X)\mathsf{g}(X)$

$$\mathscr{C}(\hat{Z}) = \int \left(\int \mathsf{C}\big(\hat{Z}(X), Z\big)\mathsf{f}(Z|X)\, dZ \right) \mathsf{g}(X)\, dX$$

call $\mathsf{G}\big(\hat{Z}(X), X\big)$

$$\mathsf{G}(U,X) = \int \mathsf{C}(U,Z)\mathsf{f}(Z|X)\,dZ$$

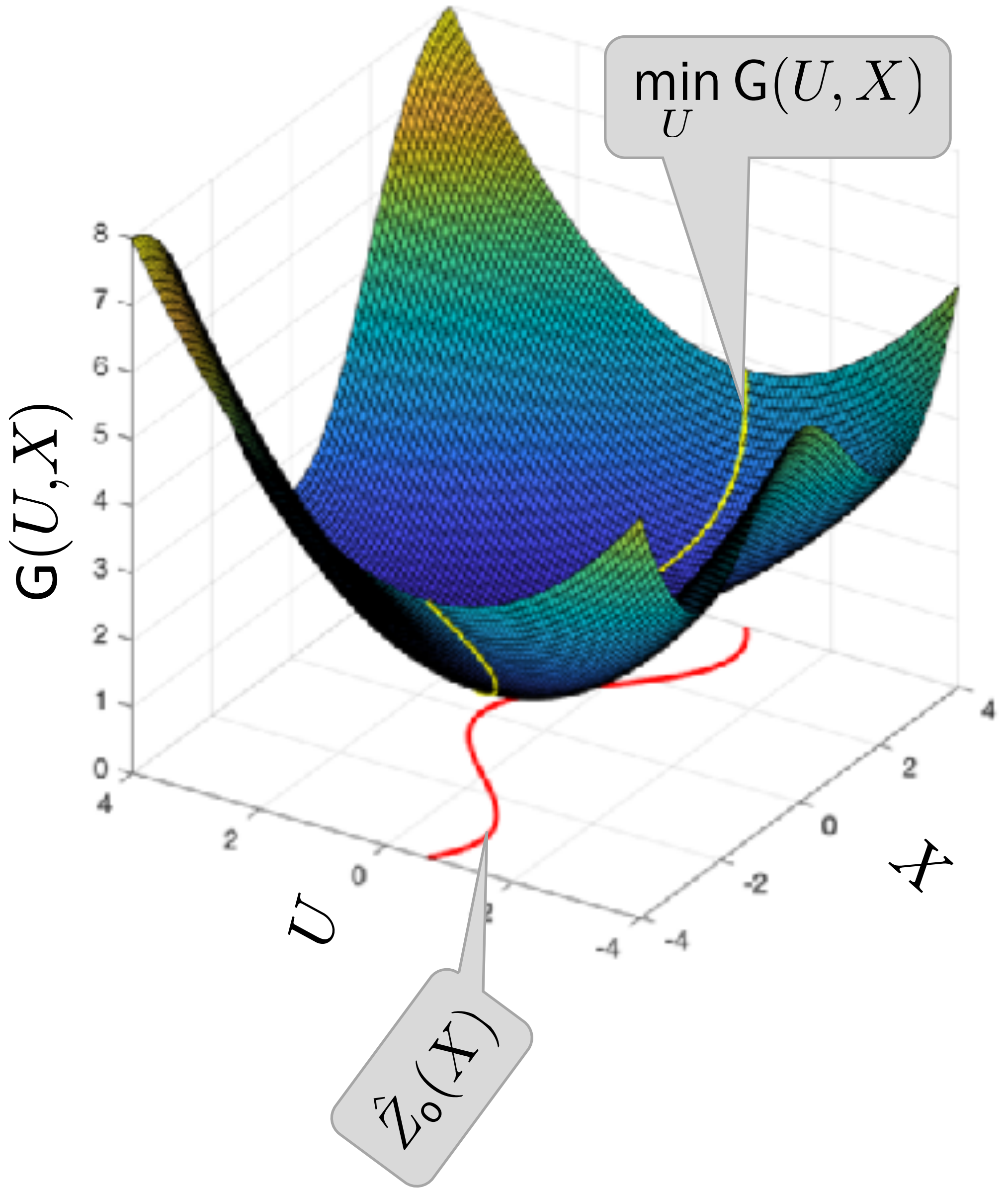
$$\mathscr{C}(\hat{Z}) = \int \mathsf{G}(\hat{Z}(X),X)\mathsf{g}(X)\,dX$$

For each X perform minimization

$$\min_U \mathsf{G}(U,X) = \Phi(X)$$

$$\hat{Z}_o(X) = \arg\min_U \mathsf{G}(U,X)$$

$$\mathsf{G}(\hat{Z}_o(X),X) = \min_U \mathsf{G}(U,X) = \Phi(X)$$



$$\min_U \mathsf{G}(U, X) = \Phi(X) \;\Rightarrow\; \mathsf{G}(??, X) \geq \Phi(X) \;\Rightarrow\; \mathsf{G}\big(\hat{Z}(X), X\big) \geq \Phi(X)$$

$$\mathscr{C}(\hat{Z}) = \int \mathsf{G}\big(\hat{Z}(X), X\big)\mathsf{g}(X)\, dX \;\geq \int \Phi(X)\mathsf{g}(X)\, dX$$

Lower bound

If estimator attains lower bound then optimum

Consider $\hat{Z}_{\mathsf{o}}(X) = \arg\min_U \mathsf{G}(U, X)$

$$\mathscr{C}(\hat{Z}_{\mathsf{o}}) = \int \mathsf{G}\big(\hat{Z}_{\mathsf{o}}(X), X\big)\mathsf{g}(X)\, dX \;= \int \Phi(X)\mathsf{g}(X)\, dX$$

$= \Phi(X)$

$\hat{Z}_{\mathsf{o}}(X) = \arg\min_U \mathsf{G}(U, X)$ is optimum

Examples

Minimum Mean Square Error (MMSE): $\mathsf{C}(\hat{Z}, Z) = \|\hat{Z} - Z\|^2$

$$\mathsf{G}(U, X) = \int \mathsf{C}(U, Z)\mathsf{f}(Z|X)\, dZ = \int \|U - Z\|^2 \mathsf{f}(Z|X)\, dZ$$

Must compute $\min_{U} \mathsf{G}(U, X)$

$$\nabla_U \mathsf{G}(U, X) = 0 \;\Rightarrow\; \int \nabla_U \|U - Z\|^2 \mathsf{f}(Z|X)\, dZ = 0$$

$$\int 2(U - Z)\mathsf{f}(Z|X)\, dZ = 0 \;\Rightarrow\; U = \int Z\mathsf{f}(Z|X)\, dZ$$

$$\boxed{\hat{Z}_{\mathsf{MMSE}}(X) = \mathbb{E}[Z|X]} = \int Z\, \mathsf{f}(Z|X)\, dZ = \frac{\int Z\, \mathsf{f}(X, Z)\, dZ}{\int \mathsf{f}(X, Z)\, dZ}$$

Minimum Mean Absolute Error (MMAE): $\mathsf{C}(\hat{Z}, Z) = |\hat{z}_1 - z_1| + \cdots + |\hat{z}_L - z_L|$

$$\hat{Z} = [\hat{z}_1 \cdots \hat{z}_L]^{\mathsf{T}}, \qquad Z = [z_1 \cdots z_L]^{\mathsf{T}}$$

Treat each coordinate separately

Minimize with respect to u : $\mathsf{G}(u, X) = \int |u - z| \, \mathsf{f}(z|X) \, dz$

$$\frac{\partial \mathsf{G}(u, X)}{\partial u} = 0 \;\Rightarrow\; \int \frac{\partial |u - z|}{\partial u} \mathsf{f}(z|X) \, dz = \int \mathsf{sign}(u - z) \, \mathsf{f}(z|X) \, dz = 0$$

$$\Rightarrow \; \int_{-\infty}^{u} \mathsf{f}(z|X) \, dz = \int_{u}^{\infty} \mathsf{f}(z|X) \, dz \;\Rightarrow\; 2 \int_{-\infty}^{u} \mathsf{f}(z|X) \, dz = 1$$

$$\int_{-\infty}^{\hat{z}_{\mathsf{MAE}}} \mathsf{f}(z|X) \, dz = \frac{1}{2},$$ Conditional Median

Maximum Aposteriory Probability (MAP):

$$\mathsf{C}(\hat{Z}, Z) = \begin{cases} 1 & \text{when } \|\hat{Z} - Z\| > \delta \\ 0 & \text{when } \|\hat{Z} - Z\| \leq \delta \end{cases} \qquad \delta \to 0$$

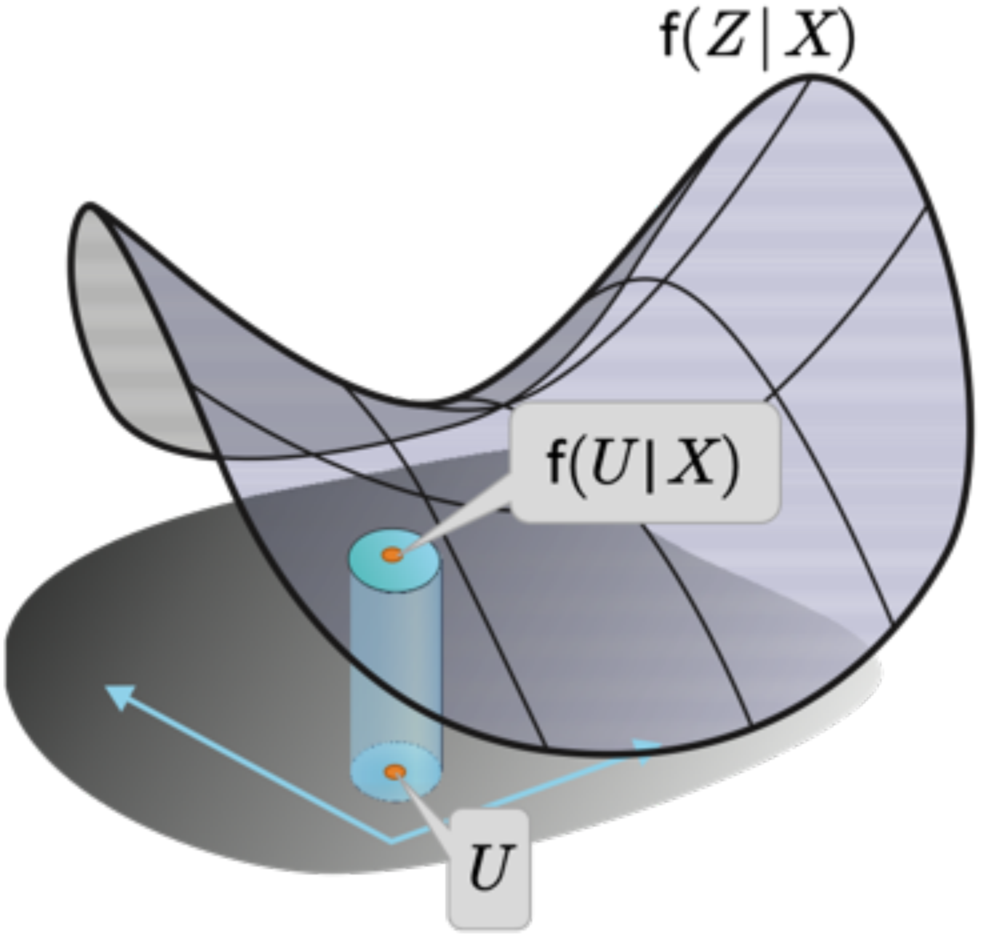
$$\mathsf{G}(U, X) = \int \mathsf{C}(U, Z)\mathsf{f}(Z|X)\, dZ$$

$$= \int_{\|U-Z\|>\delta} \mathsf{f}(Z|X)\, dZ$$

$$= 1 - \int_{\|U-Z\|\leq\delta} \mathsf{f}(Z|X)\, dZ \approx 1 - \mathsf{Ball}(\delta)\mathsf{f}(U|X)$$

$$\min_U \mathsf{G}(U, X) \approx 1 - \mathsf{Ball}(\delta)\max_U \mathsf{f}(U|X)$$

$$\hat{Z}_{\mathsf{MAP}}(X) = \arg\max_Z \mathsf{f}(Z|X)$$



Non-Bayesian Estimation

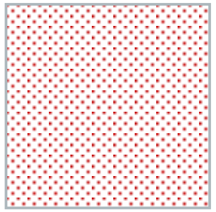
Measurements X and desired Z assumed to be "weakly" related through joint probability density f(X,Z) considered known

From f(X, Z) computed posterior probability density

$$\text{f}(Z|X) = \frac{\text{f}(X, Z)}{\text{g}(X)} = \frac{\text{f}(X, Z)}{\int \text{f}(X, Z)\, dZ}$$

In many applications no access to Z translates in no availability of f(X,Z) and f$(Z|X)$

It is possible in some applications to know the conditional density f$(X|Z)$

 $=$  $+$ 

X Z W

$$X = Z + W$$

Conditional density $f(X \mid Z)$ requires only $g_w(W)$

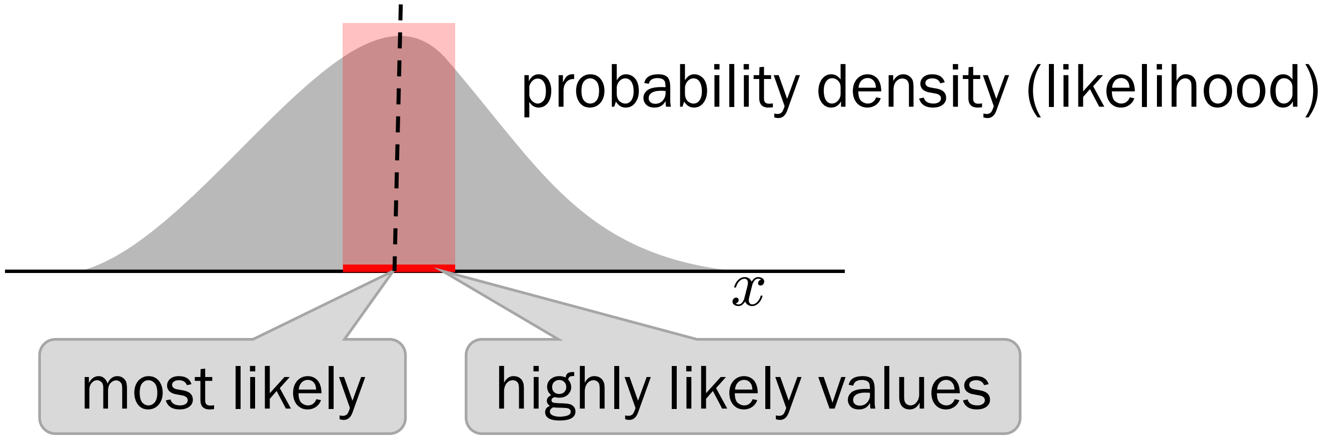
$$f(X|Z) = g_w(X - Z)$$

Joint density $f(X,Z)$ requires $g_w(W)$ and $h(Z)$

$$f(X, Z) = f(X|Z)h(Z) = g_w(X - Z)h(Z)$$

Knowing noise density $g_w(W)$ possible. Knowing $h(Z)$ difficult !

Maximum Likelihood Estimator (MLE):



Density $f(X|Z)$ is likelihood of X (given Z).

For measurements X what is Z that makes X most likely ?

We must solve $\max_{Z} f(X|Z)$

$$\hat{Z}_{\text{MLE}} = \arg\max_{Z} f(X|Z)$$

$$\hat{Z}_{\text{MAP}} = \arg\max_{Z} \text{f}(Z|X) = \arg\max_{Z} \frac{\text{f}(X,Z)}{\text{g}(X)} = \arg\max_{Z} \text{f}(X,Z)$$

$$= \arg\max_{Z} \text{f}(X|Z)\text{h}(Z)$$

if $\text{h}(Z)$ constant (degenerate uniform)

$$\hat{Z}_{\text{MLE}} = \arg\max_{Z} \text{f}(X|Z)$$

Optimality of MLE

Unbiased Estimator

$$\mathbb{E}_X\left[\hat{Z}(X)|Z\right] = Z$$

Interested in Error covariance matrix: $\mathbb{E}_X\left[\left(\hat{Z}(X) - Z\right)\left(\hat{Z}(X) - Z\right)^{\intercal}|Z\right]$

Cramer-Rao Lower Bound (CRLB)

From unbiased property

$$0 = \mathbb{E}_X\left[\hat{Z}(X) - Z|Z\right] = \int \Big(\hat{Z}(X) - Z\Big)\mathsf{f}(X|Z)\, dX$$

Derivative (Jacobian) with respect to Z

$$0 = \int \Big\{ - I\mathsf{f}(X|Z) + \Big(\hat{Z}(X) - Z\Big)\Big(\nabla_Z \mathsf{f}(X|Z)\Big)^{\mathsf{T}}\Big\}\, dX$$

$$I = \int \Big\{\Big(\hat{Z}(X) - Z\Big)\Big(\frac{\nabla_Z \mathsf{f}(X|Z)}{\mathsf{f}(X|Z)}\Big)^{\mathsf{T}}\Big\}\mathsf{f}(X|Z)\, dX$$

$$= \mathbb{E}_X\left[\Big(\hat{Z}(X) - Z\Big)\Big(\frac{\nabla_Z \mathsf{f}(X|Z)}{\mathsf{f}(X|Z)}\Big)^{\mathsf{T}}\Big| Z\right]$$

Define

$$\mathcal{E}(X) = \hat{Z}(X) - Z, \qquad \Delta(X) = \frac{\nabla_Z \mathsf{f}(X|Z)}{\mathsf{f}(X|Z)}$$

Compute covariance matrix

$$\mathbb{E}_X\left[\begin{bmatrix} \mathcal{E}(X) \\ \Delta(X) \end{bmatrix} \left[\mathcal{E}^\intercal(X)\ \Delta^\intercal(X)\right]\Big|Z\right] = \begin{bmatrix} \mathbb{E}_X[\mathcal{E}(X)\mathcal{E}^\intercal(X)|Z] & \mathbb{E}_X[\mathcal{E}(X)\Delta^\intercal(X)|Z] \\ \mathbb{E}_X[\Delta(X)\mathcal{E}^\intercal(X)|Z] & \mathbb{E}_X[\Delta(X)\Delta^\intercal(X)|Z] \end{bmatrix}$$

$$= \begin{bmatrix} \mathbb{E}_X[\mathcal{E}(X)\mathcal{E}^\intercal(X)|Z] & I \\ I & \mathbb{E}_X[\Delta(X)\Delta^\intercal(X)|Z] \end{bmatrix}$$

$$\mathbb{E}_X[\mathcal{E}(X)\mathcal{E}^\intercal(X)|Z] \geq (\mathsf{FI})^{-1}$$

Fisher Information matrix

$$\mathsf{FI} = \mathbb{E}_X[\Delta(X)\Delta^\intercal(X)|Z] = \mathbb{E}_X\left[\left(\frac{\nabla_Z \mathsf{f}(X|Z)}{\mathsf{f}(X|Z)}\right)\left(\frac{\nabla_Z \mathsf{f}(X|Z)}{\mathsf{f}(X|Z)}\right)^\intercal \Big| Z\right]$$

Estimation error covariance matrix of any unbiased estimator is lower bounded by inverse of Fisher Information matrix

$$\text{CRLB} = (\text{FI})^{-1}$$

Estimation error power of any estimator cannot go below a certain level

Call n the size of measurement data X

Theorem: Under general conditions we have

$$\lim_{n\to\infty} \mathbb{E}_X\left[\left(\hat{Z}_{\text{MLE}}(X) - Z\right)\left(\hat{Z}_{\text{MLE}}(X) - Z\right)^{\mathsf{T}} | Z\right] \times \text{FI} = I$$

Asymptotic optimality (for large data size)

Summary

Minimum Mean Square Error: $\hat{Z}_{\text{MMSE}}(X) = \mathbb{E}[Z|X]$

Minimum Mean Absolute Error: $\hat{Z}_{\text{MMAE}}(X) = \text{arg}\left\{u : \int_{-\infty}^{u} \text{f}(z|X)\,dz = \frac{1}{2}\right\}$

Maximum Aposteriori Probability: $\hat{Z}_{\text{MAP}}(X) = \text{arg}\max_{Z} \text{f}(Z|X)$

Optimum for any measurement datasize.

Maximum Likelihood Estimator: $\hat{Z}_{\text{MLE}}(X) = \text{arg}\max_{Z} \text{f}(X|Z)$

Asymptotically optimum, “large” measurement datasize.

Special Case

Let $Z = \{Z_1, Z_2\}$ and there is prior for Z_1 but not for Z_2

Treat non-existing priors as degenerate uniforms starting with MAP estimator

$$\begin{aligned}\arg\max_{Z_1,Z_2} \mathsf{f}(Z_1, Z_2|X) &= \arg\max_{Z_1,Z_2} \mathsf{f}(X, Z_1, Z_2) \\ &= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1, Z_2)\mathsf{h}(Z_1, Z_2) \\ &= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1, Z_2)\mathsf{h}_1(Z_1|Z_2)\mathsf{h}_2(Z_2)\end{aligned}$$

$\mathsf{h}_2(Z_2)$ degenerate uniform $= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1, Z_2)\mathsf{h}_1(Z_1|Z_2)$

If Z_1 and Z_2 independent and interested only in estimating Z_1

$$\hat{Z}_1 = \arg\max_{Z_1} \left\{ \left(\max_{Z_2} \mathsf{f}(X|Z_1, Z_2) \right) \mathsf{h}_1(Z_1) \right\}$$