03/20/18. Bayesian estimation: "$x|y$" viewpoint

ex. $y = Ax + e$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$

$e \sim N(0, \Sigma)$ $\Sigma = \mathbb{E}\, ee^T$

prior $x \sim N(0, \Sigma_p)$ uniform: $\Sigma_p \to \infty$

$\hookrightarrow$ indep. e

likelihood $p(y|x) \sim \exp\left[-\frac{1}{2}(y-Ax)^T \Sigma^{-1}(y-Ax)\right]$

posterior $p(x|y) \sim p(x,y) = p(y|x)\, p(x)$

$\sim \exp\left[-\frac{1}{2}(y-Ax)^T\Sigma^{-1}(y-Ax) - \frac{1}{2}x^T \Sigma_p^{-1} x\right]$

Maximum a posteriori (MAP) estimator:

$\nabla_x(-\ln p(x|y)) = A^T\Sigma^{-1}(Ax-y) + \Sigma_p^{-1} = 0$

$$\boxed{x_{MAP} = \left(A^T\Sigma^{-1}A + \Sigma_p^{-1}\right)^{-1} A^T \Sigma^{-1} y}$$

usual LS when $\Sigma_p^{-1} \to 0$, $\Sigma^{-1} = \sigma^2 I$

$= R A^T \Sigma^{-1} y$, $R^{-1} = A^T\Sigma^{-1}A + \Sigma_p^{-1}$

$= \hat{x}_{MAP}(y)$, an estimator in the Bayesian sense, random like y. ($)

Complete square in $p(x|y)$:

$$p(x|y) \sim \exp\left[-\frac{1}{2}(x - RA^T\Sigma^{-1}y) R^{-1} (x - RA^T\Sigma^{-1}y)\right]$$

$\mathbb{E}[x|y] := \mathbb{E}_{x|y}\, x = RA^T\Sigma^{-1}y = \hat{x}_{MAP}$ (posterior mean)

$\mathrm{Cov}[x|y] := \mathbb{E}_{x|y}(x - \mathbb{E}[x|y])(x - \mathbb{E}[x|y])^T = R$

or, more suggestively, $R = \Sigma_{posterior}$,

$$\boxed{\Sigma_{posterior}^{-1} = A^T\Sigma^{-1}A + \Sigma_{prior}^{-1}}$$

$f(x) = -\ln p(x|y)$, $\nabla\nabla f(x) = R^{-1}$ uniform over y

<u>Def.</u> MSE (Bayesian), or Bayes risk of an estimator $\hat{x}(y)$.

$$MSE = \mathbb{E}_{x,y} \|\hat{x}(y) - x\|^2$$

↳ over the <u>joint</u> distribution of x, y:

$$\mathbb{E}_{x,y} f(x,y) = \int f(x,y) \underbrace{p(x,y)}_{} dx\,dy$$

↳ $p(y|x)p(x)$ or $p(x|y)p(y)$

$$= \mathbb{E}_x \left[\mathbb{E}_{y|x} f(x,y)\right]$$
$$= \mathbb{E}_y \left[\mathbb{E}_{x|y} f(x,y)\right]$$

<u>Def</u> $\hat{x} = \operatorname{argmin} MSE$ is called minimum MSE (MMSE) estimator.

<u>Prop.</u> $\hat{x}(y) = \mathbb{E}_{x|y}\, x = \int x\, p(x|y)\, dx = \mathbb{E}[x|y]$ (posterior mean)

<u>Pf.</u> $MSE = \mathbb{E}_y \mathbb{E}_{x|y} (\hat{x}(y) - x)^T (\hat{x}(y) - x)$

$$= \mathbb{E}_y \underbrace{\int (\hat{x}(y) - x)^T (\hat{x}(y) - x)\, p(x|y)\, dx}$$

$$\nabla_{\hat{x}(y)} = 2 \int (\hat{x}(y) - x)\, p(x|y)\, dx = 0$$

$$\Rightarrow \hat{x}(y) = \int x\, p(x|y)\, dx = \mathbb{E}_{x|y}\, x$$

is a minimizer for fixed y

$\Rightarrow$ the random variable $\hat{x}(y) = \mathbb{E}_{x|y}\, x$, $y \sim p(y)$ (marginal), gives rise to the smallest $\mathbb{E}_y[\ldots]$.

$\square$

Rmk: $\mathbb{E}_{x,y}\left[\mathbb{E}_{x|y} x - x\right] = \mathbb{E}_y \mathbb{E}_{x|y} x - \mathbb{E}_{x,y} x = 0 \rightarrow$ unbiased.

then $MSE = \mathbb{E}_{x,y} (x - \mathbb{E}_{x|y} x)^T (x - \mathbb{E}_{x|y} x)$

$= \mathbb{E}_y \; \underbrace{tr\, Cov[x|y]}_{\text{posterior covariance}}$

Concl: MMSE = posterior mean $\mathbb{E}[x|y]$
= MAP in the case of the example.

ex. (Same!) $y = Ax + e \qquad A \in \mathbb{R}^{m \times n}$.

Postulate $\hat{x}(y) = By$ for some $B \in \mathbb{R}^{n \times m}$

$MSE = \mathbb{E}_{x,y} (By - x)^T (By - x)$.

$\nabla_B MSE = 2\, \mathbb{E}_{x,y} (By - x) y^T = 0$

$\Rightarrow \boxed{B = (\mathbb{E}_{x,y}\, xy^T)(\mathbb{E}_{x,y}\, yy^T)^{-1}}$

with $\mathbb{E}_{x,y}\, xy^T = \underbrace{\mathbb{E}_{x,y}\, xx^T A^T}_{\int xx^T p(x) dx \,=\, \Sigma_p} + \cancel{\mathbb{E}_{x,y}\, xe^T}$

Σ_p: marginal = prior

$\mathbb{E}_{x,y}\, yy^T = \mathbb{E}_{x,y}\, Axx^T A^T + 0 + \mathbb{E}_{x,y}\, ee^T$

$= A \Sigma_p A^T + \Sigma$

$\Rightarrow \boxed{B = (\mathbb{E}\, xx^T) A^T (A\, \mathbb{E}\, xx^T A^T + \mathbb{E}\, ee^T)^{-1}}$

$= \boxed{\Sigma_p A^T (A \Sigma_p A^T + \Sigma)^{-1}}$

this is in fact the same as $(A^T \Sigma^{-1} A + \Sigma_p^{-1})^{-1} A^T \Sigma^{-1}$
that we had seen earlier!

better when $\Sigma_p^{-1} \to 0$

Pf. $(A^T\Sigma^{-1}A+\Sigma_p^{-1})B(A\Sigma_p A^T+\Sigma)$

$= (A^T\Sigma^{-1}A+\Sigma_p^{-1})\Sigma_p A^T$

$= A^T\Sigma^{-1}A\Sigma_p A^T + A^T$

$= A^T\Sigma^{-1}(A\Sigma_p A^T+\Sigma)$

$\Rightarrow B = (A^T\Sigma^{-1}A+\Sigma_p^{-1})^{-1}A^T\Sigma^{-1}$ $\square$

Wiener filter: Apply to signal processing situation where

$Ax[j] = \sum g[j-k]x[k] = g * x[j]$

$(\mathbb{E}\, yy^T)_{jk} = \mathbb{E}\, y[j]y[k] = R_{yy}[j-k]$, autocorrelation, if stationary

$(\mathbb{E}\, xy^T)_{jk} = R_{xy}[j-k]$, cross-correlation

$B_{kl} = h[k-l]$, $\quad By[k] = \sum h[k-l]\,y[l]$ (filtering)

$\sum_k (\mathbb{E}\, yy^T)_{jk} B_{kl} = (\mathbb{E}\, xy^T)_{jl}$

$\Rightarrow \sum_k R_{yy}[j-k]\,h[k-l] = R_{xy}[j-l]$ pick $l=0$

$$\boxed{\sum_k R_{yy}[j-k]\,h[k] = R_{xy}[j]}$$

(normal equations of MMSE)

If no boundary issues (IIR filters), pass to the z-domain:

$S_{yy}(z) = \sum_{j\in\mathbb{Z}} R_{yy}[j]\,z^j$ $\qquad H(z) = \sum h(j) z^j$

$S_{xy}(z) = \sum_{j\in\mathbb{Z}} R_{xy}[j]\,z^j$

then $S_{yy}(z) H(z) = S_{xy}(z)$

$$\boxed{H(z) = S_{xy}(z) / S_{yy}(z)}$$

or, if starting from

$$B = \mathbb{E}xx^T A^T \left(A \, \mathbb{E}xx^T A^T + \mathbb{E}ee^T\right)^{-1},$$

$$H(z) = \frac{S_{xx}(z) G(1/z)}{S_{xx}(z) G(z) G(1/z) + S_{ee}(z)}$$

<u>Concl</u>: Fourier domain (or z-domain) diagonalizes the matrices in the expression of the MAP/MMSE estimator, in the case of stationary processes and filtering.

<u>Ref</u>: 6.011, Signals, Systems, and Inference, Chap. 11, as of 2010 on OCW