Bayes' Estimation

Different assumptions

→ We observe $x[0], \ldots, x[N-1]$ from the distribution indexed by θ : $p(x;\theta)$ as before

→ But θ is a random variable with dist $p(\theta)$.

New interpretations : $p(x;\theta)$ is joint dist. of two r.V.'s x & θ (both can be vectors in general)

$$p(x) = \int p(x;\theta)\, d\theta$$
$$p(\theta) = \int p(x;\theta)\, dx$$

→ Marginals

(inaccurate) ← our earlier pdf $(p(x;\theta))$ is now $p(x/\theta)$

Given some knowledge about θ in term of a "prior" pdf of θ : $p(\theta)$, find the "best" estimator $(\hat{\theta} = g(x))$ of θ

→ best needs a rethink since the unbiased estimator criteria is not clear : since $\underbrace{E\hat{\theta}}_{\text{Not r.v.}} = \underset{\text{r.v.}}{\theta}$ does not make sense.

→ fall back to MSE.

Define $\underset{\text{Bayesian}}{B}mse(\hat{\theta}) = E\left[(\theta - \underset{g(x)}{\hat{\theta}})^2\right] = \overbrace{\underbrace{\iint (\theta - \hat{\theta})^2 p(x,\theta)\, dx\, d\theta}_{\text{quite diff. from classical mse}}}^{\text{cannot depend on } \theta}$

$\min\limits_{\hat{\theta}} Bmse(\hat{\theta})$ → seems difficult since all $\hat{\theta} = g(x)$ are allowed.

quite diff. from classical mse $= \underbrace{\int (\hat{\theta} - \theta)^2 p(x;\theta)\, dx}_{\text{can depend on } \theta}$

Recall: $p(x;A) = p(A/x)\, p(x)$

Then $\quad B_{mse} = \int \Big[\underbrace{\int_{\theta} (\theta - \hat{\theta})^2 p(\theta/x)\, d\theta}_{I(x)} \Big] p(x)\, dx$

For any given x, $\hat{\theta}$ is just a member (say scalar case)

Note that $[\cdot] \geq 0$ & $p(x) \geq 0$.

So B_{mse} is minimized if $I(x)$ is minimized for each x.

$$\frac{\partial}{\partial \hat{\theta}} \int (\theta - \hat{\theta})^2 p(\theta/x)\, d\theta = \int \frac{\partial}{\partial \hat{\theta}} (\theta - \hat{\theta})^2 p(\theta/x)\, d\theta$$

$$= \int -2(\theta - \hat{\theta})\, p(\theta/x)\, d\theta$$

$$= -2 \int \theta p(\theta/x)\, d\theta + 2\hat{\theta} \overbrace{\int p(\theta/x)\, d\theta}^{=1}$$

Now if we put, $\frac{\partial I(x)}{\partial \hat{\theta}} = 0$, then

$$\underbrace{\hat{\theta} = \int \theta p(\theta/x)\, d\theta = E(\theta/x)}_{\text{the } B_{mse} \text{ estimator}}.$$

Note this is a f^{n} of x.

Q: How to compute this estimator?

$\hat{\theta} = E(\theta/x) \leftarrow$ We need $p(\theta/x)$.

Assumption (as stated above): We know $p(\theta)$ (prior)

$$\boxed{\text{Use Bayes' Rule: } P(A/B) = P(B/A) \frac{P(A)}{P(B)}}$$

i.e. $\quad p(\theta/x) = \dfrac{p(x/\theta)\, p(\theta)}{p(x)}$

$$= \frac{p(x/\theta)\, p(\theta)}{\underbrace{\int p(x/\theta)\, p(\theta)\, d\theta}_{= p(x;\theta)}} \leftarrow \text{marginal}$$

Q) Existence? → Always guaranteed by Baye's rule

Q) How to choose prior pdf $p(\theta)$?

→ Not quite known

→ Usually guided by convenience in getting closed form expression for the posterior.

Ex: $x[n] = A + w[n]$ Assume $A \sim \mathcal{U}[-A_0, A_0]$

$n = 0, \cdots, N-1$ $w[n] \sim N(0, \sigma^2)$

$w[n]$ & A are ind.

Then $p_x(x[n]/A) = p_w(x[n] - A/A)$

$p(B/A) = p(B)$ if A, B are ind

$$= \begin{cases} \cancel{p_w(x[n]-A)} & \cancel{\text{if } A \in [-A_0, A_0]} \\ \cancel{0} & \cancel{\text{otherwise}} \end{cases}$$

$$= \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[n]-A)^2\right] \right.$$

(if $A \in [-A_0, A_0]$ no need to say since A cannot be outside)

Then $$p(x/A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \underbrace{\exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}_{\text{Looks same as earlier}}$$

conditional (not same as $p(x;A)$)

Now using Baye's Rule:

$$p(A/x) = \begin{cases} \dfrac{\overbrace{\dfrac{1}{2A_0}}^{p(A)}\dfrac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right]}{\int_{-A_0}^{A_0} p(x/A)\,p(A)\,dA} & |A| \le A_0 \\ 0 & |A| > A_0 \end{cases}$$

→ No closed form expression.
So no closed form for $\hat{A} = E(A/x)$

On the other hand, consider a Gaussian prior:

$$p(A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp\left[-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right]$$

(Assume σ_A, μ_A are known)

$$p(x/A) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum(x[n]-A)^2\right]$$

$\Big\downarrow$ Steps (Exercise)

$$p(A/x) = \frac{1}{\sqrt{2\pi\sigma_{A/x}^2}} \exp\left[-\frac{1}{2\sigma_{A/x}^2}(A-\mu_{A/x})^2\right]$$

where $\sigma_{A/x}^2 = \dfrac{1}{\dfrac{N}{\sigma^2} + \dfrac{1}{\sigma_A^2}}$ & $\mu_{A/x} = \left(\dfrac{N}{\sigma^2}\bar{x} + \dfrac{\mu_A}{\sigma_A^2}\right)\sigma_{A/x}^2$

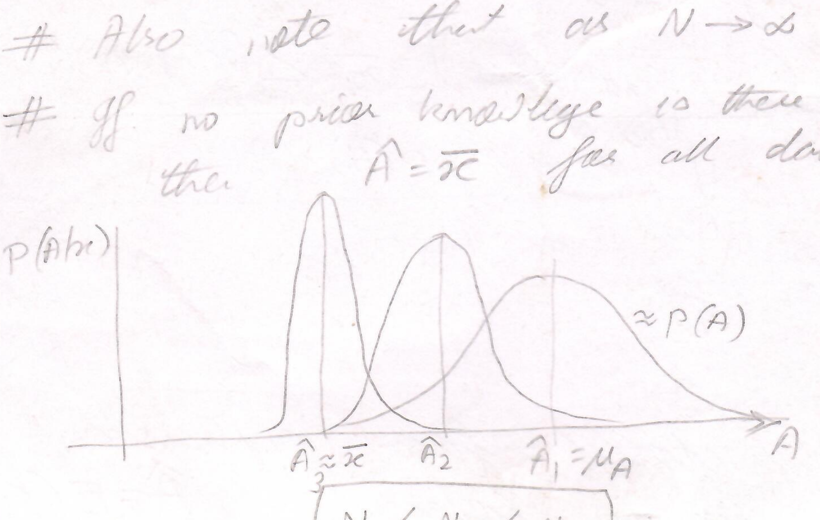
so $\hat{A} = \dfrac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}}\bar{x} + \dfrac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N}\mu_A$

<u>Note</u>: # If we have little data so that $\sigma_A^2 \ll \frac{\sigma^2}{N}$ then $\hat{A} \approx \mu_A$

As N grows, $\hat{A} \to \bar{x}$.

Also note that as $N \to \infty$, $\sigma_{A/x}^2 \to 0$

If no prior knowledge is there i.e. $\sigma_A^2 = \infty$, then $\hat{A} = \bar{x}$ for all data lengths



$\boxed{N_1 < N_2 < N_3}$

<u>Exercise:</u> Check <u>Bmse</u> for this example.