Bayesian Inference

- Conjugate priors,
- Computing the posterior distribution using Bayes' rule,
- Bayesian MMSE and MAP estimation, LMMSE estimation,
- Bayesian detection.

Example 0. We are given the following two conditional and marginal (prior and likelihood) pdfs:

$$\begin{aligned} f_{X \mid \Theta}(x \mid \boldsymbol{\theta}) &= \mathcal{N}(x \mid \boldsymbol{h}^T \boldsymbol{\theta}, \sigma^2) \\ f_{\Theta}(\boldsymbol{\theta}) &= \mathcal{N}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0, C). \end{aligned}$$

Recall that the marginal pdf $f_X(x)$ is Gaussian as well. Find it.

We just need to find the marginal mean and covariance matrix of x

$$\begin{aligned} \mathrm{E}_X(X) &= \mathrm{E}_{\Theta}[\mathrm{E}_{X \mid \Theta}(X \mid \boldsymbol{\Theta})] = \mathrm{E}_{\Theta}(\boldsymbol{h}^T \boldsymbol{\Theta}) = \boldsymbol{h}^T \boldsymbol{\theta}_0 \\ \mathrm{cov}_X(X) &= \mathrm{E}_{\Theta}[\mathrm{var}_{X \mid \Theta}(X \mid \boldsymbol{\Theta})] + \mathrm{var}_{\Theta}[\mathrm{E}_{X \mid \Theta}(X \mid \boldsymbol{\Theta})] \\ &= \underbrace{\mathrm{E}_{\Theta}(\sigma^2)}_{\sigma^2} + \underbrace{\mathrm{var}_{\Theta}(\boldsymbol{h}^T \boldsymbol{\theta})}_{\boldsymbol{h}^T C\, \boldsymbol{h}} = \sigma^2 + \boldsymbol{h}^T C\, \boldsymbol{h}. \end{aligned}$$

Example 1 (from HW # 6). Recall that, if F is a class of measurement models and P a class of prior distributions, then P is *conjugate* for F if $\pi(\theta) \in P$ and $f_{X \,|\, \Theta}(x \,|\, \theta) \in F$ implies $f_{\Theta \,|\, X}(\theta \,|\, x) \in P$.

If P is a family of probability density functions (pdfs), define the family of *M-term mixture densities based on P*, denoted P_{mix} as the set of all densities of the form

$$f_\Theta(\theta) = \sum_{i=1}^{M} p_i \, f_i(\theta)$$

where $f_i(\theta) \in P$ and p_i, $i = 1, 2, \ldots, M$ are weights satisfying

$$p_i \geq 0, \quad \sum_{i=1}^{M} p_i = 1.$$

(a) Show that, if P is conjugate for F, then P_{mix} is conjugate for F.

(b) Now take $M = 2$ and consider the model

$$X = \Theta + W$$

where

$$W \sim \mathcal{N}(0, \sigma^2)$$

and

$$\Theta \sim p\,\mathcal{N}(\theta \,|\, 0, \tau_1^2) + (1-p)\,\mathcal{N}(\theta \,|\, 0, \tau_2^2).$$

Assume that $\sigma^2, \tau_1^2, \tau_2^2$, and p are known. Find the posterior pdf $f_{\Theta \,|\, X}(\theta \,|\, x)$.

Example 2. We observe X and model it as follows:

$$X = (1 - B) \cdot G_1 + B \cdot G_2$$

where the parameter B is Bernoulli with prior probability mass function (pmf)

$$\Pr\{B = 0\} = \Pr\{B = 1\} = \tfrac{1}{2}$$

and G_1 and G_2 are random variables distributed as $\mathcal{N}(0, \sigma^2)$ and $\mathcal{N}(\mu, \sigma^2)$, respectively. Assume that μ and σ^2 are known and that B, G_1, and G_2 are independent.

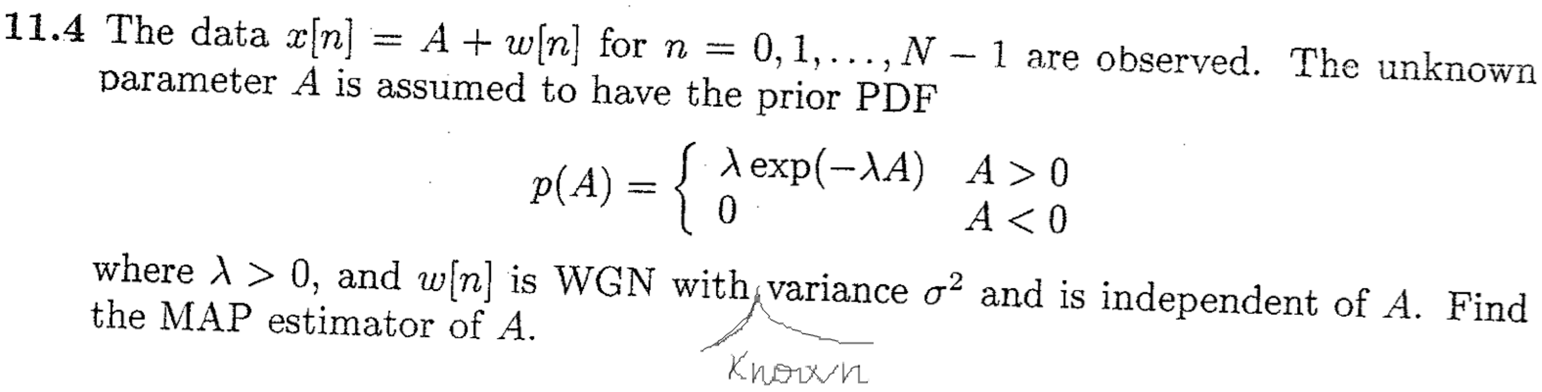
(a) Find the posterior distribution $f_{B\,|\,X}(b\,|\,x)$ of B.

(b) Find the posterior mean (MMSE estimator) of B.

(c) Find the posterior mode (MAP estimator) of B.

(d) Find the LMMSE estimate of B.

Example 3. Problem 11.4 in Kay-II.



11.4 The data $x[n] = A + w[n]$ for $n = 0, 1, \ldots, N-1$ are observed. The unknown parameter A is assumed to have the prior PDF

$$p(A) = \begin{cases} \lambda \exp(-\lambda A) & A > 0 \\ 0 & A < 0 \end{cases}$$

where $\lambda > 0$, and $w[n]$ is WGN with Known variance σ^2 and is independent of A. Find the MAP estimator of A.

Example 4. Consider conditionally i.i.d. measurements $X[0], X[1], \ldots, X[N-1]$ given the parameter θ, following the exponential pdf:

$$f_{X \,|\, \Theta}(x[n] \,|\, \theta) = \mathrm{Expon}(x[n] \,|\, \theta) = \theta \, \exp(-\theta \, x[n]) \, i_{(0,+\infty)}(x[n]).$$

We adopt the Bayesian approach with a gamma prior pdf for the unknown parameter θ:

$$\pi(\theta) = \mathrm{Gamma}(\theta \,|\, \alpha, \beta) \propto \theta^{\alpha-1} \exp(-\beta \, \theta) \, i_{(0,+\infty)}(\theta)$$

where $\alpha > 0$ and $\beta > 0$ are known constants. Define $\boldsymbol{x} = [x[0], x[1], \ldots, x[N-1]]^T$.

(a) Find the posterior pdf

$$f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x}).$$

(b) Compute the posterior mean

$$\widehat{\theta}_{\mathrm{MMSE}}(\boldsymbol{x}) = \mathrm{E}_{\,\Theta \,|\, \boldsymbol{X}}(\Theta \,|\, \boldsymbol{x})$$

and mode

$$\widehat{\theta}_{\mathrm{MAP}}(\boldsymbol{x}) = \arg\max_{\theta} f_{\Theta \,|\, \boldsymbol{X}}(\theta \,|\, \boldsymbol{x}).$$

Example 5. Consider a scalar observation X coming from the following probability density function (pdf):

$$f_{X\,|\,\Theta}(x\,|\,\theta) = (2\,\theta\,x + 1 - \theta)\cdot i_{(0,1)}(x)$$

where θ is the parameter and the parameter space is

$$-1 \le \theta \le 1.$$

Assign a uniform prior pdf to Θ:

$$\pi(\theta) = \mathrm{U}(\theta\,|\,-1,1) = \tfrac{1}{2}\,i_{(-1,1)}(\theta).$$

- **(a) (HW 3)** Find and plot the maximum-likelihood (ML) estimate of θ as a function of x.
- **(a)** Find and plot the MMSE estimator of θ as a function of x.
- **(b)** Compute the (classical) bias and MSE for each of the above two estimators. Recall:

$$\begin{aligned}\mathrm{bias}\{\widehat{\theta}\} &= b(\theta) = \mathrm{E}_X(\widehat{\theta}) - \theta \\ \mathrm{MSE}\{\widehat{\theta}\} &= \mathrm{E}_{\boldsymbol{X}}[(\widehat{\theta} - \theta)^2] = \mathrm{var}_{\boldsymbol{X}}(\widehat{\theta}) + b^2(\theta).\end{aligned}$$

  Does one estimator perform uniformly better than the other?

Example 6. Consider the hypothesis-testing problem:

$$\begin{aligned} \mathcal{H}_0: &\quad X = W \quad \text{versus} \\ \mathcal{H}_1: &\quad X = S + W \end{aligned}$$

where S and W are two independent random variables with densities

$$\begin{aligned} f_S(s) &= \mathrm{Expon}(s \,|\, a) = a \exp(-a\, s)\, i_{(0,+\infty)}(s) \\ f_W(w) &= \mathrm{Expon}(w \,|\, b) = b \exp(-b\, w)\, i_{(0,+\infty)}(w). \end{aligned}$$

Find the Bayes' decision rule for this problem.

Example 7. Additive shot-noise channel, from HW # 6. Consider the signal-plus-noise model

$$X = \Theta + W$$

where the signal Θ follows

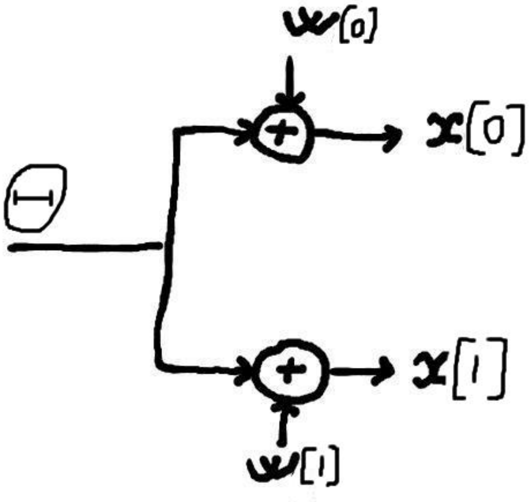
$$\Theta \sim \mathrm{U}(0, 1)$$

and noise W is conditionally Gaussian given $\Theta = \theta$, with variance proportional to the signal, i.e.

$$\{W \mid \Theta = \theta\} \sim \mathcal{N}(0, c\,\theta)$$

for some known constant $c > 0$. Find the LMMSE estimate of Θ based on X. Your answer should be in terms of X and c only.

Example 8. Laplacian noise. One of two signals $\Theta = \theta_0 = -1$ or $\Theta = \theta_1 = 1$ is transmitted over the channel shown in the figure below,

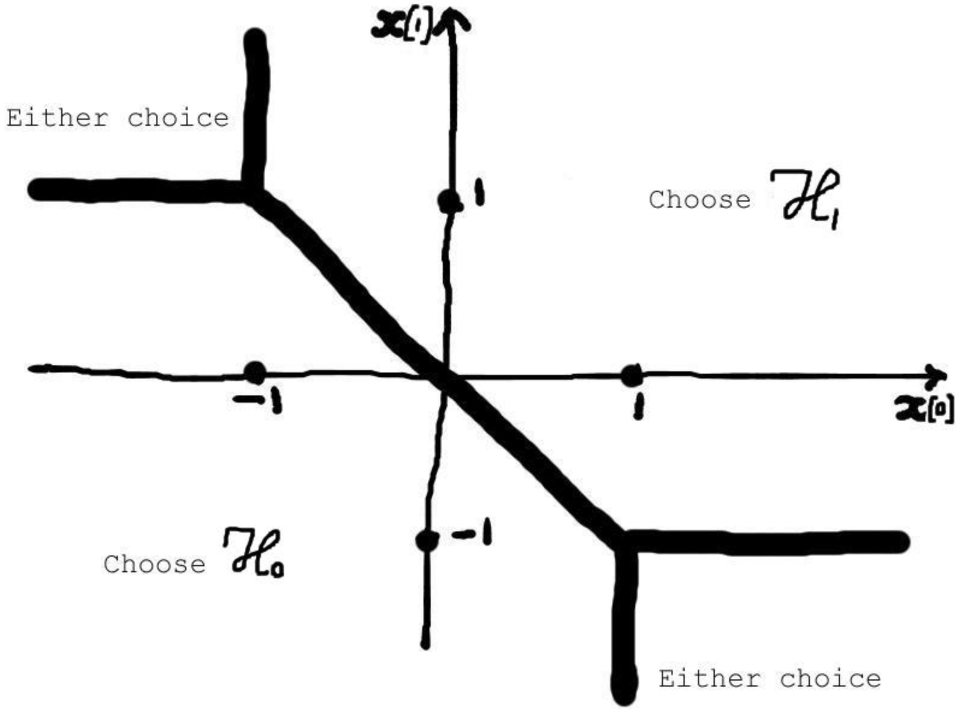


$$X[0] = \Theta + W[0], \quad X[1] = \Theta + W[1], \quad \boldsymbol{X} = \left[\begin{array}{c} X[0] \\ X[1] \end{array} \right]$$

where $W[0]$ and $W[1]$ are independent realizations of Laplacian noise with pdf

$$f_W(w) = \tfrac{1}{2} \cdot \exp(-|w|)$$

and Θ and $W[0], W[1]$ are independent. Show that the optimal decision regions (i.e. minimizing the average error probability) for equally likely messages are as shown below


x[1]
x[0]
1
-1
1
-1
Either choice
Choose $\mathcal{H}_1$
Choose $\mathcal{H}_0$
Either choice