ESE 524 - Homework 5

Problems and Solutions
Assigned date: 03/27/19
Due date: 04/09/19
Total Points: 100 + 20 (extra credits)

1) **Bayesian Statistics**
The observed random variable is x. We want to estimate the parameter λ. The probability density of x as a function of λ is

$$p_{x|\lambda}(X|\lambda) = \begin{cases} \lambda e^{-\lambda X}, & X \geq 0, \lambda \geq 0, \\ 0, & X < 0. \end{cases}$$

The prior density of λ depends on two parameters: $n_0 \in \mathbb{Z}_{>0}$ and l_0 such that

$$p_{\lambda|n_0,l_0}(\lambda|n_0, l_0) = \begin{cases} \frac{l_0^{n_0}}{\Gamma(n_0)} e^{-\lambda l_0} \lambda^{n_0-1}, & \lambda \geq 0 \\ 0, & \lambda < 0. \end{cases}$$

a) **(10 pts)** Find $\mathbb{E}(\lambda)$ and $\mathrm{var}(\lambda)$ before any observations are made.
Hint: $\Gamma(n_0) = (n_0 - 1)!$

Soluiton Make a substitution as $\lambda l_0 = t \implies d\lambda l_0 = dt$

$$\mathbb{E}(\lambda) = \int_0^\infty \frac{l_0^{n_0}}{\Gamma(n_0)} e^{-\lambda l_0} \lambda^{n_0} d\lambda = \frac{n_0}{l_0}.$$

Similarly,

$$\mathbb{E}(\lambda^2) = \frac{n_0(n_0+1)}{l_0^2}.$$

Therefore,

$$\mathrm{var}(\lambda) = \mathbb{E}(\lambda^2) - (\mathbb{E}(\lambda))^2 = \frac{n_0}{l_0^2}.$$

b) **(10 pts)** Assume that one observation is made. Find $p_{\lambda|x}(\lambda|X)$. Find the Bayesian MMSE estimate of λ, and denote it as $\hat{\lambda}_{\mathrm{MMSE}}$. Find $\mathbb{E}_{\lambda,x}(\hat{\lambda}_{\mathrm{MMSE}} - \lambda)^2]$

Solution:
The posterior distribution is given as

$$p_{\lambda|x}(\lambda|x) = \frac{p_{x|\lambda}(x|\lambda) p_{\lambda|l_0,n_0}(\lambda|l_0, n_0)}{p_x(x)}$$

Let us first evaluate the expression of the marginal distribution, $p_x(x)$.

$$\begin{aligned} p_x(x) &= \int_0^\infty p_{x,\lambda}(x,\lambda) d\lambda \\ &= \int_0^\infty p_{x|\lambda}(x|\lambda) p_{\lambda|l_0,n_0}(\lambda|l_0,n_0) d\lambda \\ &= \int_0^\infty \lambda e^{-\lambda x} \frac{l_0^{n_0}}{\Gamma(n_0)} e^{-\lambda l_0} \lambda^{n_0-1} \\ &= \frac{l_0^{n_0} n_0}{(x+l_0)^{n_0+1}} \end{aligned}$$

Therefore, the posterior distribution is given as

$$\begin{aligned} p_{\lambda|x}(\lambda|x) &= \lambda e^{-\lambda x}\frac{l_0^{n_0}}{\Gamma(n_0)}e^{-\lambda l_0}\lambda^{n_0-1}\frac{(x+l_0)^{n_0+1}}{l_0^{n_0}n_0} \\ &= \frac{(x+l_0)^{n_0+1}}{\Gamma(n_0+1)}e^{-\lambda(x+l_0)}\lambda^{n_0} \\ &= p_{\lambda|n_0,l_0}(\lambda|n_*,l_*) \end{aligned}$$

where $n_* = n_0 + 1$ and $l_* = l_0 + x$. We see that the posterior density has the same form as prior density. We can find $\hat{\lambda}_{\mathrm{MMSE}}$ by directly evaluating

$$\begin{aligned} \hat{\lambda}_{\mathrm{MMSE}} &= \mathbb{E}_{\lambda|x}(\lambda|x) \\ &= \frac{n_*}{l_*} = \frac{n_0+1}{l_0+x} \end{aligned}$$

We evaluate $\mathbb{E}[(\hat{\lambda}_{\mathrm{MMSE}} - \lambda)^2]$ in two steps.

$$\mathbb{E}[(\hat{\lambda}_{\mathrm{MMSE}} - \lambda)^2] = \mathbb{E}_x[\mathbb{E}_{\lambda|x}(\hat{\lambda}_{\mathrm{MMSE}} - \lambda)^2]$$

The inner expression is equivalent to (from part (a))

$$\mathbb{E}_{\lambda|x}(\hat{\lambda}_{\mathrm{MMSE}} - \lambda)^2 = \frac{n_0+1}{(l_0+x)^2}$$

Using the expressions derived in the above equations

$$\mathbb{E}_x[\mathbb{E}_{\lambda|x}(\hat{\lambda}_{\mathrm{MMSE}} - \lambda)^2] = \frac{n_0(n_0+1)}{(n_0+2)l_0^2}.$$

c) **(10 pts)** Now assume that n independent observations are made. Denote these n observations by the vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T$. Verify that

$$p_{\lambda|x}(\lambda|\boldsymbol{X}) = \begin{cases} \frac{l_* n_*}{\Gamma(n_*)}e^{-\lambda l_*}\lambda^{n_*-1}, & \lambda \geq 0 \\ 0, & \lambda < 0. \end{cases}$$

where $l_* = l + l_0$ and $n_* = n + n_0$ with $l = \sum_{i=1}^n x_i$. Find $\hat{\lambda}_{\mathrm{MMSE}}$.

Solution: The result can be shown by induction. The a priori density for the second observation is the a posterior density of the first observation. Thus,

$$p_{\lambda|x_1,x_2}(\lambda|x_1,x_2) = p_{\lambda|n_0,l_0}(\lambda|n'_*,l'_*)$$

where $n'_* = 1 + n_* = 2 + n_0$ and $l'_* = x_2 + l_* = x_2 + x_1 + l_0$. Proceeding to n observations gives desired result.

d) **(10 pts)** Also find the maximum a posterior estimate of λ assuming that there are n i.i.d observations as in part (c). Denote this estimator as $\hat{\lambda}_{\mathrm{MAP}}$. Is $\hat{\lambda}_{\mathrm{MAP}} = \hat{\lambda}_{\mathrm{MMSE}}$? Justify your answer.

Solution: Differentiating the posterior distribution we find

$$\hat{\lambda}_{\mathrm{MAP}} = \frac{n_0}{l_0+x}$$

for one observation. For n observations

$$\hat{\lambda}_{\mathrm{MAP}} = \frac{n_0+n-1}{l_0+\sum_{i=1}^n x_i}.$$

We see that $\hat{\lambda}_{\mathrm{MAP}}$ and $\hat{\lambda}_{\mathrm{MMSE}}$ are not equal but they are essentially equal for large n.

2) **Lecture Notes, l4.pdf. Page #60**:
Define the loss function

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\boldsymbol{x})) = \boldsymbol{c}^T \boldsymbol{a}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\boldsymbol{x}))$$

where $\boldsymbol{c} = [c_1, \ldots, c_p]^T$, $c_i > 0 \quad \forall i$ and a_i is the absolute error between i-th component of $\boldsymbol{\theta}$ and the i-th component of $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$, i.e., $a_i(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\boldsymbol{x})) = |\theta_i - \hat{\theta}_i(\boldsymbol{x})|$.

(a) **(10 pts)** Show that the conditional risk may be written as

$$\int p(\boldsymbol{\theta}|\boldsymbol{x}) \boldsymbol{c}^T \boldsymbol{a}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\boldsymbol{x})) d\boldsymbol{\theta} = \sum_{i=1}^{p} c_i \beta$$

where

$$\beta = -\int_{-\infty}^{\hat{\theta}_i(\boldsymbol{x})} p(\theta_i|\boldsymbol{x})(\theta_i - \hat{\theta}_i(\boldsymbol{x})) d\theta_i + \int_{\hat{\theta}_i(\boldsymbol{x})}^{\infty} p(\theta_i|\boldsymbol{x})(\theta_i - \hat{\theta}_i(\boldsymbol{x})) d\theta_i.$$

Solution: Conditional risk is defined as

$$\begin{aligned}\int L(\theta, \hat{\theta}(\boldsymbol{x})) f_{\theta|\boldsymbol{x}}(\theta|\boldsymbol{x}) d\theta &= \int \boldsymbol{\pi}^T a(\theta, \hat{\theta}(\boldsymbol{x})) f(\theta|\boldsymbol{x}) d\theta \\ &= \sum_{i=1}^{P} \pi_i \int |\theta_i - \hat{\theta}_i(\boldsymbol{x})| f(\theta_i|\boldsymbol{x}) d\theta_i\end{aligned}$$

We know that

$$|\theta_i - \hat{\theta}_i(\boldsymbol{x})| = \begin{cases} -(\theta_i - \hat{\theta}_i(\boldsymbol{x})) & if \theta_i - \hat{\theta}_i(\boldsymbol{x}) \leq 0 \\ (\theta_i - \hat{\theta}_i(\boldsymbol{x})) & if \theta_i - \hat{\theta}_i(\boldsymbol{x}) > 0 \end{cases}$$

Therefore,

$$\begin{aligned}\int_\theta L(\theta, \hat{\theta}(\boldsymbol{x})) f_{\theta|\boldsymbol{x}}(\theta|\boldsymbol{x}) d\theta &= \sum_{i=1}^{P} \pi_i \left(\int_{-\infty}^{\hat{\theta}_i(\boldsymbol{x})} -(\theta_i - \hat{\theta}_i(\boldsymbol{x})) f(\theta_i|\boldsymbol{x}) d\theta_i + \int_{-\hat{\theta}_i(\boldsymbol{x})}^{\infty} (\theta_i - \hat{\theta}_i(\boldsymbol{x})) f(\theta_i|\boldsymbol{x}) d\theta_i \right) \\ &= \sum_{i=1}^{P} \pi_i \beta\end{aligned}$$

(b) **(10 pts)** Show that the minimizing estimator $\hat{\theta}_i(\boldsymbol{x})$ is the median of the conditional density $f(\theta_i|\boldsymbol{x})$.

Solution: Minimizing the expected loss gives us

$$\frac{\partial}{\partial \hat{\theta}_i} \sum_{i=1}^{P} \pi_i \beta = \sum_{i=1}^{P} \pi_i \frac{\partial \beta}{\partial \hat{\theta}_i} = 0$$

Expanding the partial derivative and setting to zero we get

$$\frac{\partial \beta}{\partial \hat{\theta}_i} = -\frac{\partial}{\partial \hat{\theta}_i} \int_{-\infty}^{\hat{\theta}_i(\boldsymbol{x})} d\theta_i f(\theta_i|\boldsymbol{x})(\theta_i - \hat{\theta}_i(\boldsymbol{x})) + \frac{\partial}{\partial \hat{\theta}_i} \int_{\hat{\theta}_i(\boldsymbol{x})}^{\infty} d\theta_i f(\theta_i|\boldsymbol{x})(\theta_i - \hat{\theta}_i(\boldsymbol{x}))$$

On further simplifying we get

$$\int_{-\infty}^{\hat{\theta}_i(\boldsymbol{x})} f(\theta_i|\boldsymbol{x}) d\theta_i = \int_{\hat{\theta}_i(\boldsymbol{x})}^{\infty} f(\theta_i|\boldsymbol{x}) d\theta_i$$

Hence, $\hat{\theta}_i(\boldsymbol{x})$ is the median of the conditional density.

3) **Bayesian MSE**
(20 pts) In fitting a line through experimental data, we assume the model

$$x[n] = A + Bn + w[n], \quad \forall - M \leq n \leq M$$

where $w[n]$ is white Gaussian noise with variance σ^2. If we have some prior knowledge of the slope B and intercept A, such as

$$\begin{bmatrix} A \\ B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} A_0 \\ B_0 \end{bmatrix}, \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix} \right),$$

find the MMSE estimator of A and B as well as the Bayesian MSE. Assume that A and B are independent of $w[n]$. Which parameter will benefit most from the prior knowledge?

Solution:
As the model is a Bayesian linear model, we get

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mu_\theta} + \left(\boldsymbol{C}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{H}^T \boldsymbol{C}_{\boldsymbol{c}}^{-1} \boldsymbol{H} \right)^{-1} \boldsymbol{H}^T \boldsymbol{C}_{\boldsymbol{w}}^{-1} (\boldsymbol{x} - \boldsymbol{H}\boldsymbol{\mu_\theta})$$

where

$$\begin{aligned} \boldsymbol{\mu_\theta} &= [A_0, B_0]^T \\ \boldsymbol{C_\theta} &= \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix} \\ \boldsymbol{H} &= \begin{bmatrix} 1 & -M \\ \vdots & \vdots \\ 1 & M \end{bmatrix} \\ \boldsymbol{C_w} &= \sigma^2 \boldsymbol{I} \end{aligned}$$

Substituting and further simplifying we get

$$\begin{aligned} \hat{A} &= A_0 + \frac{\frac{N}{\sigma^2}}{\frac{1}{\sigma_A^2} + \frac{N}{\sigma^2}} (\bar{x} - A_0), \quad N = 2M + 1 \\ \hat{B}_0 &= B_0 + \frac{\frac{\sum_{n=-M}^{M} n^2}{\sigma^2}}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^{M} n^2}{\sigma^2}} \left[\frac{\sum_{n=-M}^{M} n x[n]}{\sum_{n=-M}^{M} n^2} - B_0 \right] \end{aligned}$$

Similarly,

$$\begin{aligned} \text{BMSE}(\hat{A}) &= \left(\frac{1}{\sigma_A^2} + \frac{N}{\sigma^2} \right)^{-1} \\ \text{BMSE}(\hat{B}) &= \left(\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^{M} n^2}{\sigma^2} \right)^{-1} \end{aligned}$$

The intercept $\hat{A}$ will benefit most from prior knowledge since the reduction in BMSE due to the data is much larger for the slope than for the intercept.

4) **Kalman Filter (requires MATLAB)**
In Kalman filter, at every step, we update the estimation of the current state β_k using the estimate from the previous state β_{k-1} and only the observation from the current state y_k. Therefore, a Kalman filter is also a recursive filter. Now suppose that we have the following state and measurement equations:

$$\text{State Model}: \beta_k = \beta_{k-1}, \quad k = 1, 2, \ldots$$
$$\text{Measurement Model}: y_k = \beta_k + w_k, \quad k = 1, 2, \ldots$$

where β_k is a scalar state at time k and it does not vary, y_k is the observation at time k, and $w_k \sim \mathcal{N}(0, \sigma^2_{w_k})$. Here $\{w_k,\ k = 1, 2, \ldots\}$ are mutually independent. The prior pdf for the initial state is $\beta_0 \sim \mathcal{N}(0, \sigma^2_{\beta_0})$.

a) (**5 pts**) Write the Kalman filter equations for this linear system.

Solution:
The Kalman filter:
Predict:

$$\hat{\beta}(k|k-1) = \hat{\beta}(k-1|k-1)$$
$$P(k|k-1) = P(k-1|k-1)$$

Update:

$$\hat{\beta}(k|k) = \hat{\beta}(k|k-1) + \frac{P(k|k-1)}{\sigma^2_{w_k} + P(k|k-1)} \left(y_k - \hat{\beta}(k|k-1) \right)$$
$$P(k|k) = P(k|k-1) - \frac{P(k|k-1)^2}{\sigma^2_{w_k} + P(k|k-1)}.$$

Note that it means that $\beta_k | y_1, y_2, \ldots, y_k \sim \mathcal{N}(\hat{\beta}(k|k), P(k|k))$.

b) (**5 pts**) From another perspective, in this problem, the state is unchanged. It means that we actually keep estimating β_0 at every time step k, when we recevied the measurement y_k. Now, suppose that we are at time step k, i.e., we have observations $\boldsymbol{y}_{1:k} \stackrel{\text{def}}{=} \{y_1, y_2, \ldots, y_k\}$. Give the posterior distribution of β_0 (i.e., β_k).

Hint: Write it as a Gaussian linear model $\boldsymbol{y}_{1:k} = \boldsymbol{H}_k \beta_0 + \boldsymbol{W}_k$, and specify the $\boldsymbol{H}_k$ and $\boldsymbol{W}_k$, then you can use the formula from the lectures.

Solution:
Write the linear model

$$\boldsymbol{y}_{1:k} = \boldsymbol{H}_k \beta_0 + \boldsymbol{W}_k$$

where $\boldsymbol{H}_k = [\underbrace{1, 1, \ldots, 1}_{k}]^T$, and $\boldsymbol{W}_k = [w_1, \ldots, w_k]^T$. Using the formula, we have that

$$\hat{\beta}_{0,k} = \frac{\sum_{i=1}^k y_i / \sigma^2_{w_i}}{\sum_{i=1}^k \frac{1}{\sigma^2_{w_i}} + \frac{1}{\sigma^2_{\beta_0}}},$$
$$C_{\beta_0 | \boldsymbol{y}_{1:k}} = \frac{1}{\sum_{i=1}^k \frac{1}{\sigma^2_{w_i}} + \frac{1}{\sigma^2_{\beta_0}}}.$$

Note that the results on page 54(l4.pdf) are used. It is fine if other formulas are used.

c) (**10 pts**) Now let $\sigma^2_{\beta_0} = 100$, $\sigma^2_{w_k} = 25$, and

$$\boldsymbol{y}_{1:10} = \{16.7549, 25.9058, 16.2077, 14.4519, 15.7722, 17.1367, 17.2066, 20.8919, 19.0157, 22.9322\}.$$

For time $k = 1, 2, 3, \ldots, 10$, use MATLAB to compute the MMSE estimates of β_k using the formulas in a) and b). Compare whether they are equal or not.

Solution:
See “Kalmanest.m”. Here are estimation results from Kalman filter:

Here are estimation results computed from the Gaussian linear model:

time k	1	2	3	4	5	6	7	8	9	10
$\hat{\beta}(k\|k)$	13.4039	18.9603	18.1134	17.2518	16.9700	16.9967	17.0256	17.4943	17.6587	18.1732

time k	1	2	3	4	5	6	7	8	9	10
$\hat{\beta}_{0,k}$	13.4039	18.9603	18.1134	17.2518	16.9700	16.9967	17.0256	17.4943	17.6587	18.1732

They are equal as expected.

5) **Extra Credit**:
(**20 pts**) Come up with an example and solution illustrating one or more concepts from class so far. This example should be something you believe would be good to present in class to help other students understand a concept from the lectures. MATLAB (or other software) simulations are encouraged. Problems can be inspired by or explore applications from literature, but should not just copy the results of a paper. We will also accept larger "projects" which may take more time to complete but will be worth more points.