ESE 524 - Homework 4

Database Problems and Solutions
Assigned date: 03/05/19
Due Date: xx/yy/zz
Total Points: 100

1) **(Bayesian Estimation)**
Suppose that we have two systems:

$$y_1 = \theta + n_1$$
$$y_2 = \theta + n_2$$

where $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$, $n_1 \sim \mathcal{N}(0, \sigma_{n1}^2)$, and $n_2 \sim \mathcal{N}(0, \sigma_{n2}^2)$. In addition, θ, n_1, and n_2 are mutually independent.

a) Suppose that we have observed y_1, can you infer the output of y_2, i.e., $p(y_2|y_1)$?
(hint: you can determine the joint distribution of y_1 and y_2 at first and then compute $p(y_2|y_1)$.)

b) Suppose we only observe y_1, what is $p(\theta|y_1)$? What is the Bayesian MMSE estimate of θ_1? If we further observe y_2, what is $p(\theta|y_1, y_2)$? What is the Bayesian MMSE estimate of θ_1 now? Compare the Bayesian MSE of these two estimators.

Solution:

a) We can see that y_1 and y_2 are jointly Gaussian and zero mean. The covariance matrix can be found by

$$\boldsymbol{C} = \mathbb{E}\begin{bmatrix} y_1 y_1 & y_1 y_2 \\ y_2 y_1 & y_2 y_2 \end{bmatrix} = \mathbb{E}\begin{bmatrix} (\theta + n_1)(\theta + n_1) & (\theta + n_1)(\theta + n_2) \\ (\theta + n_2)(\theta + n_1) & (\theta + n_2)(\theta + n_2) \end{bmatrix} = \begin{bmatrix} \sigma_\theta^2 + \sigma_{n1}^2 & \sigma_\theta^2 \\ \sigma_\theta^2 & \sigma_\theta^2 + \sigma_{n2}^2 \end{bmatrix}$$

As we know $p(y_2|y_1)$ is also Gaussian, and have

$$\mathbb{E}\left[y_2|y_1\right] = \mu_{y_2} + C_{21}C_{11}^{-1}(y_1 - \mu_{y_1}) = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_{n1}^2} y_1$$
$$\text{var}\left[y_2|y_1\right] = C_{22} - C_{21}C_{11}^{-1}C_{12} = \sigma_\theta^2 + \sigma_{n2}^2 - \frac{\sigma_\theta^4}{\sigma_\theta^2 + \sigma_{n1}^2}$$

b) Easily we have

$$[\theta, y_1]^T \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_\theta^2 & \sigma_\theta^2 \\ \sigma_\theta^2 & \sigma_\theta^2 + \sigma_{n1}^2 \end{bmatrix}\right),$$

and

$$\theta|y_1 \sim \mathcal{N}\left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_{n1}^2} y_1, \sigma_\theta^2 - \frac{\sigma_\theta^4}{\sigma_\theta^2 + \sigma_{n1}^2}\right)$$

Note that the Bayesian MMSE estimate is $\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_{n1}^2} y_1$, and the Bayesian MSE is

$$\text{MSE}_1 = \mathbb{E}_{y_1,\theta}[\theta - \mathbb{E}[\theta|y_1]]^2 = \mathbb{E}_{y_1}\left[\mathbb{E}_{\theta|y_1}\left[\theta - \mathbb{E}[\theta|y_1]\right]^2\right] = \mathbb{E}_{y_1}[\text{var}[\theta|y_1]] = \sigma_\theta^2 - \frac{\sigma_\theta^4}{\sigma_\theta^2 + \sigma_{n1}^2} = \frac{\sigma_\theta^2 \sigma_{n1}^2}{\sigma_\theta^2 + \sigma_{n1}^2}.$$

Further, we have

$$[\theta, y_1, y_2]^T \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \sigma_\theta^2 & \sigma_\theta^2 & \sigma_\theta^2 \\ \sigma_\theta^2 & \sigma_\theta^2 + \sigma_{n1}^2 & \sigma_\theta^2 \\ \sigma_\theta^2 & \sigma_\theta^2 & \sigma_\theta^2 + \sigma_{n2}^2 \end{bmatrix}\right),$$

and

$$\theta|y_1,y_2 \sim \mathcal{N}\left([\sigma_\theta^2,\sigma_\theta^2]\begin{bmatrix}\sigma_\theta^2+\sigma_{n1}^2 & \sigma_\theta^2\\ \sigma_\theta^2 & \sigma_\theta^2+\sigma_{n2}^2\end{bmatrix}^{-1}\begin{bmatrix}y_1\\y_2\end{bmatrix}, \sigma_\theta^2-[\sigma_\theta^2,\sigma_\theta^2]\begin{bmatrix}\sigma_\theta^2+\sigma_{n1}^2 & \sigma_\theta^2\\ \sigma_\theta^2 & \sigma_\theta^2+\sigma_{n2}^2\end{bmatrix}^{-1}\begin{bmatrix}\sigma_\theta^2\\ \sigma_\theta^2\end{bmatrix}\right)$$

$$\theta|y_1,y_2 \sim \mathcal{N}\left(\frac{\sigma_\theta^2\sigma_{n2}^2y_1+\sigma_\theta^2\sigma_{n1}^2y_2}{\sigma_\theta^2\sigma_{n1}^2+\sigma_\theta^2\sigma_{n2}^2+\sigma_{n1}^2\sigma_{n2}^2}, \frac{\sigma_{n1}^2\sigma_{n2}^2\sigma_\theta^2}{\sigma_\theta^2\sigma_{n1}^2+\sigma_\theta^2\sigma_{n2}^2+\sigma_{n1}^2\sigma_{n2}^2}\right)$$

Thus, the Bayesian MMSE estimate is $\frac{\sigma_\theta^2\sigma_{n2}^2y_1+\sigma_\theta^2\sigma_{n1}^2y_2}{\sigma_\theta^2\sigma_{n1}^2+\sigma_\theta^2\sigma_{n1}^2+\sigma_{n1}^2\sigma_{n2}^2}$, and the Bayesian MSE is

$$\begin{aligned}\text{MSE}_2 = \mathbb{E}_{y_1,y_2,\theta}[\theta-\mathbb{E}[\theta|y_1,y_2]]^2 &= \mathbb{E}_{y_1,y_2}\left[\mathbb{E}_{\theta|y_1,y_2}[\theta-\mathbb{E}[\theta|y_1,y_2]]^2\right]\\ &= \mathbb{E}_{y_1,y_2}[\text{var}[\theta|y_1,y_2]] = \frac{\sigma_{n1}^2\sigma_{n2}^2\sigma_\theta^2}{\sigma_\theta^2\sigma_{n1}^2+\sigma_\theta^2\sigma_{n2}^2+\sigma_{n1}^2\sigma_{n2}^2}.\end{aligned}$$

Note that

$$\text{MSE}_1 = \frac{1}{\frac{1}{\sigma_{n1}^2}+\frac{1}{\sigma_\theta^2}}$$
$$\text{MSE}_2 = \frac{1}{\frac{1}{\sigma_{n1}^2}+\frac{1}{\sigma_{n2}^2}+\frac{1}{\sigma_\theta^2}} < \text{MSE}_1.$$

Thus, we can see the contribution from a further observation.

2) **Posterior Distribution)**

Suppose that X follows a binomial distribution with parameters n and θ. Suppose that the prior distribution of the parameter θ has the beta p.d.f.:

$$\boldsymbol{h(\theta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

where 0 <θ <1.

a) Find the posterior p.d.f of θ given $X=x$

b) Find the Bayesian Minimum Mean Squared Estimator of θ.

Solution:

First, we find the joint p.d.f. of the statistic Y and the parameter θ by multiplying the prior p.d.f. $h(\theta)$ and the conditional p.m.f. of Y given θ.

$$= p(x,\theta) = \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

Then, we find the marginal p.d.f. of Y by integrating $p(y,\theta)$ over the parameter space of θ:

$$p(x) = \int_0^1 p(x,\theta)d\theta$$
$$= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+x)\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

Then, the posterior p.d.f. of θ, given that Y = y is

$$\boldsymbol{p(\theta|x)} = \frac{p(x,\theta)}{p(x)} = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

We can see that the posterior p.d.f. of θ is a beta p.d.f. with parameters $x+\alpha$ and $n-x+\beta$.

In order to minimize the squared error loss function, we should use the conditional mean

$$E[\theta \mid x]$$

as an estimate of the parameter θ.

In our case, the posterior p.d.f. of is the beta p.d.f. with parameters $x+\alpha$ and $n-x+\beta$.
Therefore, the conditional mean is:

$$\boldsymbol{E[\theta|x]} = \frac{\alpha+x}{\alpha+x+n-x+\beta} = \frac{\alpha+x}{\alpha+n+\beta}$$

It is the Bayesian's best estimate of θ when using the squared error loss function.

3) (**Source: l4.pdf, Pages 16-19**) Let $X_1,\ldots,X_n$ be independent and identically distributed as a Gaussian distribtuion given as

$$p(X_k|\theta) \sim \mathcal{N}(\theta,\sigma^2).$$

The parameter of interest θ is also follows a Gaussian distribution given as

$$\pi(\theta) \sim \mathcal{N}(m,\sigma_\theta^2).$$

Find

(a) The conditional distribution of $\theta|x$, i.e., $p(\theta|x)$?

Solution: We can write the posterior distribution as

$$\begin{aligned}
p(\theta|\boldsymbol{x}) &= \frac{p(\boldsymbol{x}|\theta)p(\theta)}{p(\boldsymbol{x})} \\
&\propto p(\boldsymbol{x}|\theta)p(\theta) \\
&= \prod_{i=1}^{N} p(x_i|\theta)p(\theta) \\
&\propto c_1 \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2\right\}\exp\left\{-\frac{1}{2\sigma_\theta^2}(\theta-m)^2\right\} \quad \text{where } c_1 \text{ is the proportionality constant} \\
&= c_1 \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\bar{x}+\bar{x}-\theta)^2\right\}\exp\left\{-\frac{1}{2\sigma_\theta^2}(\theta-m)^2\right\} \quad \text{where} \quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i
\end{aligned}$$

The first term can be simplified to

$$\sum_{i=1}^{N}(x_i-\bar{x}+\bar{x}-\theta)^2 = Ns^2 + N(\bar{x}-\theta)^2 \quad \text{where} \quad s = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2$$

Therefore,

$$
\begin{aligned}
p(\theta|\boldsymbol{x}) &\propto c_1 \exp\left\{-\frac{1}{2\sigma^2}[Ns^2 + N(\bar{x}-\theta)^2] - \frac{1}{2\sigma_\theta^2}(\theta-m)^2\right\} \\
&\propto c_2 \exp\left\{-\frac{N}{2\sigma^2}(\bar{x}^2 - 2\bar{x}\theta + \theta^2) - \frac{1}{2\sigma^2\theta}(\theta^2 + m^2 - 2m\theta)\right\} \\
&\propto c_3 \exp\left\{-\frac{N}{2\sigma^2}(\theta^2 - 2\bar{x}\theta) - \frac{1}{2\sigma_\theta^2}(\theta^2 - 2\theta m)\right\} \\
&= c_3 \exp\left\{-\theta^2\left(\frac{1}{2\sigma^2/N} + \frac{1}{2\sigma_\theta^2}\right) + \theta\left(\frac{m}{\sigma_\theta^2} + \frac{\bar{x}}{\sigma^2/N}\right)\right\}
\end{aligned}
$$

Let

$$
\frac{1}{\sigma_x^2} = \frac{1}{\sigma^2/N} + \frac{1}{\sigma_\theta^2}
$$

Then

$$
\begin{aligned}
p(\theta|\boldsymbol{x}) &= c_3 \exp\left\{-\frac{\theta^2}{2\sigma_x^2} + \theta\left(\frac{m}{\sigma_\theta^2} + \frac{\bar{x}}{\sigma^2/N}\right)\right\} \\
&\propto c_4 \exp\left\{-\frac{1}{2\sigma_x^2}\left(\theta - \left(\frac{\sigma_x^2 m}{\sigma_\theta^2} + \frac{\sigma_x^2\bar{x}}{\sigma^2/N}\right)\right)^2\right\}
\end{aligned}
$$

Therefore, the posterior distribution is a Gaussian distribution which is distributed as

$$
p(\theta|\boldsymbol{x}) \sim \mathcal{N}\left(\frac{\sigma_x^2 m}{\sigma_\theta^2} + \frac{\sigma_x^2\bar{x}}{\sigma^2/N}, \sigma_x^2\right) \tag{1}
$$

(b) What are the conditional mean and variance of $f(\theta|x)$?

Solution: The conditional mean of the posterior distribution is given as

$$
\begin{aligned}
\mathbb{E}(\theta|\boldsymbol{x}) &= \frac{\sigma_x^2 m}{\sigma_\theta^2} + \frac{\sigma_x^2\bar{x}}{\sigma^2/N} \\
&= \frac{\frac{m}{\sigma_\theta^2} + \frac{\bar{x}}{\sigma^2/N}}{\frac{1}{\sigma^2/N} + \frac{1}{\sigma_\theta^2}}
\end{aligned}
$$

The variance of the posterior distribution is given as

$$
\mathrm{var}(\theta|\boldsymbol{x}) = \sigma_x^2 = \frac{\sigma_\theta^2\sigma^2/N}{\sigma_\theta^2 + \sigma^2/N}
$$

(c) Compare the $\mathbb{E}(\theta|x)$ to $\hat{\theta}_{\mathrm{ML}}$, the maximum likelihood estimator for this problem.

Solution: The ML estimate of θ is $\bar{x}$. Based on the expression of the conditional expectation, we can write

$$
\begin{aligned}
\mathbb{E}(\theta|\boldsymbol{x}) &= \frac{\frac{m}{\sigma_\theta^2} + \frac{\bar{x}}{\sigma^2/N}}{\frac{1}{\sigma^2/N} + \frac{1}{\sigma_\theta^2}} \\
&= \frac{\frac{m}{\sigma_\theta^2} + \frac{\hat{\theta}}{\sigma^2/N}}{\frac{1}{\sigma^2/N} + \frac{1}{\sigma_\theta^2}}
\end{aligned}
$$

As N becomes larger, the prior information is neglected and the maximum likelihood portion dominates.

4) **Jeffreys' Prior**

For a Gaussian distribution, we have

$$p(\boldsymbol{x}|\sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2\right]$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]^T$, θ is the mean, which is known. Suppose that we have known the Jeffreys' prior for σ^2, which is

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2}.$$

a) Regard σ as the parameter for the Gaussian distribution rather than σ^2, and find $\sqrt{\mathcal{I}(\sigma)}$, where $\mathcal{I}(\sigma)$ is the Fisher information for σ.
b) Use $\phi = \log\sigma^2$ to re-parameterize the Gaussian distribution, and find $\sqrt{\mathcal{I}(\phi)}$, where $\mathcal{I}(\phi)$ is the Fisher information for ϕ.
c) Use change-of-variables transformation to compute the prior $\pi(\sigma)$ based on $\pi(\sigma^2)$.
d) Use change-of-variables transformation to compute the prior $\pi(\phi)$ based on $\pi(\sigma^2)$.
e) Check if $\pi(\sigma) \propto \sqrt{\mathcal{I}(\sigma)}$ and $\pi(\phi) \propto \sqrt{\mathcal{I}(\phi)}$.

Note: the change-of-variables transformation for pdfs is referred to that $p_Y(y) = \{p_X(x) \cdot |dx/dy|\}|_{x=h^{-1}(y)}$ if $Y = h(X)$. The above problems shows that we have two alternative ways to compute the Jeffreys' priors for transformed parameters.

Solution:

a) We have that

$$p(\boldsymbol{x}|\sigma) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2\right]$$

$$\begin{aligned}
\mathcal{I}(\sigma) &= -\mathrm{E}\left[\frac{d^2 \log p(\boldsymbol{x}|\sigma)}{d\sigma d\sigma}\right] \\
&= -\mathrm{E}\left[\frac{d^2(-N/2\log(2\pi) - N\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2)}{d\sigma d\sigma}\right] \\
&= -\mathrm{E}\left[\frac{d^2(-N\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\theta)^2)}{d\sigma d\sigma}\right] \\
&= -N\frac{1}{\sigma^2} + \frac{3}{\sigma^4}\cdot N\sigma^2 \\
&= \frac{2N}{\sigma^2}
\end{aligned}$$

Thus, $\sqrt{\mathcal{I}(\sigma)} = \sqrt{\frac{2N}{\sigma^2}} \propto \frac{1}{\sigma}$.

b) we have

$$p(\boldsymbol{x}|\phi) = \frac{1}{(2\pi)^{N/2}e^{\phi N/2}} \exp\left[-\frac{1}{2e^{\phi}}\sum_{i=1}^{N}(x_i-\theta)^2\right]$$

Thus,

$$\begin{aligned}
\mathcal{I}(\log\sigma^2) = \mathcal{I}(\phi) &= -\mathrm{E}\left[\frac{d^2 \log p(\boldsymbol{x}|\phi)}{d\phi d\phi}\right] \\
&= -\mathrm{E}\left[\frac{d^2(-N/2\log(2\pi) - \phi N/2 - \frac{1}{2e^\phi}\sum_{i=1}^N (x_i-\theta)^2)}{d\phi d\phi}\right] \\
&= \mathrm{E}\left[\frac{1}{2e^\phi}\sum_{i=1}^N (x_i-\theta)^2\right] \\
&= \frac{1}{2e^\phi}\sum_{i=1}^N \mathrm{E}(x_i-\theta)^2 \\
&= \frac{1}{2e^\phi}\cdot N\sigma^2|_{\sigma^2=e^\phi} \\
&= \frac{N}{2}
\end{aligned}$$

Thus, $\sqrt{\mathcal{I}(\phi)} = \sqrt{N/2} \propto 1$

c) For $\pi(\sigma)$, we have

$$\pi(\sigma) = \pi(\sigma^2)\cdot|d\sigma^2/d\sigma| = \pi(\sigma^2)\cdot 2\sigma \propto \frac{1}{\sigma^2}\cdot 2\sigma \propto \frac{1}{\sigma}$$

d)

$$\begin{aligned}
\pi(\log\sigma^2) = \pi(\phi) &= \pi(\sigma^2)\cdot|d\sigma^2/d\log\sigma^2| \\
&= \pi(\sigma^2)\cdot|d\log\sigma^2/d\sigma^2|^{-1} \\
&\propto \frac{1}{\sigma^2}\cdot|d\log\sigma^2/d\sigma^2|^{-1} \\
&= \frac{1}{\sigma^2}\left(\frac{1}{\sigma^2}\right)^{-1} \\
&= 1
\end{aligned}$$

e) Yes. Already shown above.

5) **MATLAB Problem**: Assume you are interested in examining the proportion of defective products coming out of a production line. Denote $\theta = \frac{\text{\# of defective items}}{\text{\# of total items}}$.

(i) Let $x_i \sim \mathrm{Bern}(\theta)$, $i = 0, 1, ..., N-1$ be the (i.i.d.) examination results for the first N products off the line. Assume there were n_f defective products in the sample. Compute two posterior distributions $\mathrm{p}(\theta|x)$, one with a uniform prior, and one with Jeffrey's prior. (Hint: The uniform distribution is a special case of the Beta distribution with both parameters equal to 1.)

Solution: Let $x = 1$ represent failure. We can write the uniform prior as a Beta distribution with parameters $\alpha = \beta = 1$, i.e. $p(\theta) = \frac{1}{\mathrm{B}(1,1)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$.

The likelihood function of the data is $\mathrm{p}(x|\theta) = \Pi_{n=0}^{N-1}\theta^{x_n}(1-\theta)^{1-\mathrm{x_i}} = \theta^{n_f}(1-\theta)^{N-n_f}$

Then the posterior is proportional to:

$$\mathrm{p}(\theta|x) \propto \mathrm{p}(x|\theta)\mathrm{p}(\theta) = \theta^{n_f}(1-\theta)^{N-n_f}\frac{1}{\mathrm{B}(1,1)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \propto$$

$$\theta^{\alpha+n_f-1}(1-\theta)^{beta+N-n_f-1} \sim \mathrm{Beta}(\alpha+n_f, \beta+N-n_f)$$

This is just another beta distribution with parameters shifted by the number of defective and nondefective products.

Jeffrey's prior is given by the square root of the fisher information matrix. For a single bernoulli random variable,

$$I(\theta) = \mathbb{E}\left(\frac{\partial^2 log(p(x|\theta))}{\partial\theta^2}\right) = \mathbb{E}(log(\theta^x(1-\theta)^{1-x})) =$$

$$\mathbb{E}(\frac{\partial^2}{\partial\theta^2} x \log\theta + (1-x)\log(1-\theta)) =$$

$$\mathbb{E}(\frac{X}{\theta^2} + \frac{1-x}{(1-\theta)^2}) =$$

$$\frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta(1-\theta)}$$

For n i.i.d. samples, the fisher information in multiplied. So the jeffrey's prior is $p(\theta) = n\theta^{-1/2}(1-\theta)^{-1/2} \propto \text{Beta}(1/2, 1/2)$ Therefore the posterior is another beta distribution with parameters $1/2 + n_f,\ 1/2 + N - n_f$

(ii) Let the "true" value of $\theta_{true} = 0.25$ and $N = 100$. Generate N random samples using the true value of theta. For each prior from the previous part, plot the likelihood, prior, and posterior as functions of θ. What are the values of the MAP and MLE estimators for this problem? Which one is closer to θ? (Hint: We are not asking you to compute formulas for these estimators, but to get the answers from your plots).
solution: See HW4.m for solutions. The posterior should be a beta with degree of freedom shifted by the number of defective samples from your sample.

(iii) While you've been computing these posteriors, your production line has cranked out another 50 products. Generate another set of N random samples based on the true value. Using your results from (ii) for the Jeffrey's prior, treat this as a sequential Bayesian inference problem. Plot the original posterior and the second posterior together. Then repeat this process several times. Do the posterior distributions appear to converge? Is your final sequential MAP estimate better than the original MAP estimate?
Solution: See HW4.m for solutions. Each iteration slightly shifts the Beta distributions, as iterations increase they should spike around the correct value.