APPM 5720

Solutions to Final Exam Review Problems

1. The likelihood is

$$f(\vec{x}|\lambda) = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^n (x_i!)} \prod_{i=1}^n I_{\{0,1,2,\ldots\}}(x_i).$$

The prior is

$$f(\lambda) = \frac{1}{\Gamma(\alpha)}\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda} I_{(0,\infty)}(\lambda).$$

The posterior is

$$\begin{aligned} f(\lambda|\vec{x}) &\propto f(\vec{x}|\lambda)\cdot f(\lambda) \\ &\propto e^{-n\lambda}\lambda^{\sum x_i}\cdot \lambda^{\alpha-1}e^{-\beta\lambda} I_{(0,\infty)}(\lambda) \\ &= \lambda^{\sum x_i+\alpha-1}e^{-(n+\beta)\lambda} I_{(0,\infty)}(\lambda) \end{aligned}$$

$$\Rightarrow \boxed{\lambda|\vec{x} \sim \Gamma\left(\sum x_i+\alpha, n+\beta\right)}$$

---

2. (a) The likelihood is

$$f(\vec{x}|\theta) = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \prod_{i=1}^n I_{\{0,1\}}(x_i.)$$

The prior is

$$f(\theta) = \frac{1}{\mathcal{B}(a,b)}\theta^{a-1}(1-\theta)^{b-1} I_{(0,1)}(\theta).$$

The posterior is

$$\begin{aligned} f(\theta|\vec{x}) &\propto f(\vec{x}|\theta)\cdot f(\theta) \\ &\propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\cdot \theta^{a-1}(1-\theta)^{b-1} I_{(0,1)}(\theta) \\ &= \theta^{\sum x_i+a-1}(1-\theta)^{n-\sum x_i+b-1} I_{(0,1)}(theta) \end{aligned}$$

$$\Rightarrow \boxed{\theta|\vec{x} \sim Beta\left(\sum x_i+a, n-\sum x_i+b\right)}$$

The posterior Bayes estimator is

$$\begin{aligned} \widehat{\theta}_{PBE} &= \mathsf{E}[\Theta|\vec{X}] = \frac{\sum X_i+a}{\sum X_i+a+n-\sum X_i+b} = \frac{\sum X_i+a}{a+b+n} \\ &= \frac{\sum X_i}{a+b+n} + \frac{a}{a+b+n} \\ &= \frac{n}{a+b+n}\underbrace{\overline{X}}_{\substack{\text{sample}\\ \text{mean}}} + \frac{a+b}{a+b+n}\underbrace{\frac{a}{a+b}}_{\substack{\text{prior}\\ \text{mean}}} \end{aligned}$$

(b) Let $\vec{x} = (x_1, x_2, \ldots, x_n)$. Let $a^* = \sum_{i=1}^n x_i + a$ and $b^* = n - \sum_{i=1}^n x_i + b$.
The predictive distribution is given by

$$\begin{aligned} f(x_{n+1}|\vec{x}) &= \int_0^1 f(x_{n+1}|\theta, \vec{x}) \cdot f(\theta|\vec{x})\, d\theta \\ &= \int_0^1 f(x_{n+1}|\theta) \cdot f(\theta|\vec{x})\, d\theta \\ &= \int_0^1 \theta^{x_{n+1}}(1-\theta)^{1-x_{n+1}} \frac{1}{\mathcal{B}(a^*,b^*)} \theta^{a^*-1}(1-\theta)^{b^*-1}\, d\theta \\ &= \frac{1}{\mathcal{B}(a^*,b^*)} \int_0^1 \theta^{a^*+x_{n+1}-1}(1-\theta)^{b^*-x_{n+1}}\, d\theta \\ &= \frac{\mathcal{B}(a^*+x_{n+1}, b^*-x_{n+1}+1)}{\mathcal{B}(a^*,b^*)} \end{aligned}$$

Plugging in our expressions for a^* and b^*, this becomes

$$f(x_{n+1}|\vec{x}) = \frac{\mathcal{B}\left(\sum_{i=1}^{n+1} x_i + a, n - \sum_{i=1}^{n+1} x_i + b + 1\right)}{\mathcal{B}\left(\sum_{i=1}^{n} x_i + a, n - \sum_{i=1}^{n} x_i + b\right)}$$

for $x_{n+1} \in \{0, 1\}$.
Note that this discrete pdf represents the probability $P(X_{n+1} = x_{n+1}|\vec{X} = \vec{x})$. A point estimate for X_{n+1} is given by

$$\begin{aligned} \widehat{X}_{n+1} &= \mathsf{E}[X_{n+1}|\vec{X} = \vec{x}] = 0 \cdot P(X_{n+1} = 0|\vec{X} = \vec{x}) + 1 \cdot P(X_{n+1} = 1|\vec{X} = \vec{x}) \\ &= P(X_{n+1} = 1|\vec{X} = \vec{x}) \\ &= \frac{\mathcal{B}\left(\sum_{i=1}^{n} x_i + 1 + a, n - \sum_{i=1}^{n} x_i - 1 + b + 1\right)}{\mathcal{B}\left(\sum_{i=1}^{n} x_i + a, n - \sum_{i=1}^{n} x_i + b\right)} \end{aligned}$$

While you can stop there, this simplifies to

$$\boxed{\widehat{X}_{n+1} = \frac{\sum_{i=1}^n x_i + a}{a + b + n}}$$

---

3. (a) Let $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n \stackrel{iid}{\sim} MVN_p(\vec{0}, V)$ for $n \geq p$ and V positive deinite.
Define the $p \times n$ matrix $\mathbb{X}$ so that its ith column is $\vec{X}_i$.
Define the p
timesp matrix A as

$$A := \mathbb{X}\mathbb{X}^t.$$

Then A is a random matrix that has a Wishart distribution with parameters n, p, and V.
n is called the "degrees of freedom" of the distribution.
V is a variance/covariance parameter but is not the variance/covariance matrix for A. (In fact, we have not even defined the variance for matrix random variables.)

(b) The Wishart distribution is a conjugate prior for the inverse of the variance/covariance matrix of the multivariate normal distribution.

---

4. (a) The likelihood is

$$f(\vec{x}|\lambda) = \lambda^5 e^{-\lambda \sum_{i=1}^5 x_i} \prod_{i=1}^{5} I_{(0,\infty)}(x_i).$$

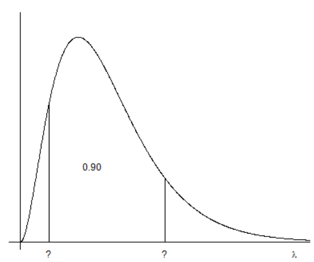
The prior is

$$f(\lambda) = 2e^{-2\lambda} I_{(0,\infty)}(\lambda).$$

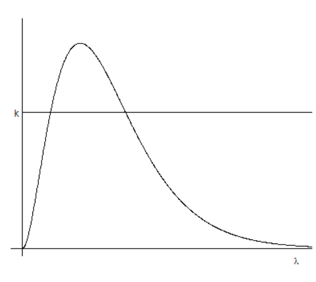
The posterior is

$$\begin{aligned} f(\lambda|\vec{x}) &\propto f(\vec{x}|\lambda) \cdot f(\lambda) \\ &\propto \lambda^5 e^{-\lambda \sum_{i=1}^5 x_i} \cdot e^{-2\lambda} I_{(0,\infty)}(\lambda) \\ &= \lambda^5 e^{(\sum x_i + 2)\lambda} I_{(0,\infty)}(\lambda) \\ \Rightarrow \quad \lambda|\vec{x} &\sim \Gamma\left(6, \sum_{i=1}^5 x_i + 2\right) = \Gamma(6,5) \end{aligned}$$

since $\sum_{i=1}^5 x_i = 3$.



(b) Consider a horizontal line at some height k.



We want to choose the maximum k for which

$$\{\lambda : f(\lambda|\vec{x}) > k\}$$

gives a 90% credible region. This will necessarily be the shortest interval in λ.

(c) Let $\vec{x} = (x_1, x_2, x_3, x_4, x_5)$. Then

$$\begin{aligned} f(x_6|\vec{x}) &= \int_0^\infty f(x_6|\lambda, \vec{x}) f(\lambda|\vec{x})\, d\lambda \\ &= \int_0^\infty f(x_6|\lambda) f(\lambda|\vec{x})\, d\lambda \\ &= \int_0^\infty \lambda e^{-\lambda x_6} \cdot \tfrac{1}{\Gamma(6)} 5^6 \lambda^5 e^{-5\lambda}\, d\lambda \\ &= \tfrac{5^6}{\Gamma(6)} \int_0^\infty \lambda^6 e^{-(x_6+5)\lambda}\, d\lambda \end{aligned}$$

The integrand is looking like a $\Gamma(7, x_6 + 5)$ pdf. Thus

$$\begin{aligned} f(x_6|\vec{x}) &= \tfrac{5^6}{\Gamma(6)} \int_0^\infty \lambda^6 e^{-(x_6+5)\lambda}\, d\lambda \\ &= \tfrac{5^6}{\Gamma(6)} \tfrac{\Gamma(7)}{(x_6+5)^7} \underbrace{\int_0^\infty \frac{1}{\Gamma(7)} (x_6+5)^7 \lambda^6 e^{-(x_6+5)\lambda}\, d\lambda}_{1} \\ &= \tfrac{6 \cdot 5^6}{(x_6+5)^7} \end{aligned}$$

for $x_6 > 0$.

In order to estimate $P(X_6 > 0.25)$, we could simulate several values of x_6 from the target density

$$f(x_6|\vec{x}) = \frac{6 \cdot 5^6}{(x_6+5)^7} I_{(0,\infty)}(x_6)$$

and compute the proportion of times we get a result greater than 0.25. There are many ways to do this. I would simply use the inverse cdf method since the cdf is compoutable and invertible for this problem.

---

5. (a)

$$f(\vec{x}|M_1) = \prod_{i=1}^n \left(\frac{1}{2}\right)^{x_i} \left(1 - \frac{1}{2}\right)^{1-x_i} = \left(\frac{1}{2}\right)^{\sum x_i} \left(\frac{1}{2}\right)^{n - \sum x_i} = \left(\frac{1}{2}\right)^n$$

$$\begin{aligned} f(\vec{x}|M_2) &= \int_0^1 f(\vec{x}|\theta)\, f(\theta)\, d\theta \\ &= \int_0^1 \theta^{\sum x_i} (1-\theta)^{n - \sum x_i} \cdot \tfrac{1}{\mathcal{B}(a,b)} \theta^{a-1} (1-\theta)^{b-1}\, d\theta \\ &\vdots \\ &= \tfrac{\mathcal{B}(\sum x_i + a, n - \sum x_i + b)}{\mathcal{B}(a,b)} \end{aligned}$$

The Bayes factor is

$$B_{12} = \frac{f(\vec{x}|M_1)}{f(\vec{x}|M_2)} = \frac{\mathcal{B}(a,b)}{2^n \mathcal{B}(\sum x_i + a, n - \sum x_i + b)}$$

(b)

$$\begin{aligned} f(\vec{x}|M_1) &= \int_0^{1/2} f(\vec{x}|\theta)\, f(\theta)\, d\theta \\ &= \tfrac{1}{\mathcal{B}(a,b)} \int_0^{1/2} \theta^{a + \sum x_i - 1} (1-\theta)^{n - \sum x_i + b - 1}\, d\theta \end{aligned}$$

and

$$\begin{aligned} f(\vec{x}|M_1) &= \int_{1/2}^{1} f(\vec{x}|\theta)\, f(\theta)\, d\theta \\ &= \frac{1}{\mathcal{B}(a,b)} \int_{1/2}^{1} \theta^{a+\sum x_i - 1}(1-\theta)^{n-\sum x_i + b - 1}\, d\theta \end{aligned}$$

These would have to be numerically integrated. (Use Monte Carlo integration!)
The Bayes factor is then

$$B_{12} = \frac{f(\vec{x}|M_1)}{f(\vec{x}|M_2)}.$$

(c) A "large" Bayes factor would support model M_1 over model M_2.

---

6. (a) We need to specify two probabilities $P(M_1)$ and $P(M_2)$ (Really only one since they must add up to 1!) representing our prior belief in Model 1 versus Model 2.
We also need to set up priors for β_1, σ_1^2, β_2, and σ_2^2.

(b)

$$\begin{aligned} f(\vec{y}|M_j) &= \int_0^\infty \int_{-\infty}^{\infty} f(\vec{y}|\beta_j, \sigma_j^2, M_j) \cdot f(\beta_j, \sigma_j^2|M_j)\, d\beta_j\, d\sigma_j^2 \\ &= \int_0^\infty \int_{-\infty}^{\infty} f(\vec{y}|\beta_j, \sigma_j^2) \cdot f(\beta_j, \sigma_j^2)\, d\beta_j\, d\sigma_j^2 \end{aligned}$$

(c)

$$P(M_j|\vec{y}) = \frac{f(\vec{y}|M_j) \cdot P(M_j)}{f(\vec{y})}$$

(d)

$$PO_1 = \frac{P(M_1|\vec{y})}{P(M_2|\vec{y})}$$

Note that, if there were more than 2 models, it should be

$$PO_1 = \frac{P(M_1|\vec{y})}{1 - P(M_1|\vec{y})}$$

A large value of PO_1 gives support for Model 1.

(e) First, we will need the predictive density.

$$\begin{aligned} f(y_{n+1}|\vec{y}) &= \int_0^\infty \int_{-\infty}^{\infty} f(y_{n+1}|\beta_1, \sigma_1^2, \vec{y}) f(\beta_1, \sigma_1^2|\vec{y})\, d\beta_1\, d\sigma_1^2 \\ &= \int_0^\infty \int_{-\infty}^{\infty} f(y_{n+1}|\beta_1, \sigma_1^2) f(\beta_1, \sigma_1^2|\vec{y})\, d\beta_1\, d\sigma_1^2 \end{aligned}$$

The prediction is then

$$\widehat{Y}_{n+1} = \mathsf{E}[Y_{n+1}|\vec{y}] = \int_{-\infty}^{\infty} y_{n+1} f(y_{n+1}|\vec{y})\, dy_{n+1}.$$

(f) The model averaging prediction is

$$\widehat{Y}_{n+1} = \mathsf{E}[Y_{n+1}|\vec{y}] = \mathsf{E}[Y_{n+1}|M_1, \vec{y}] \cdot P(M_1|\vec{y}) + \mathsf{E}[Y_{n+1}|M_2, \vec{y}] \cdot P(M_2|\vec{y})$$

The expectation $\mathsf{E}[Y_{n+1}|M_1, \vec{y}]$ is exactly the expectation computed in part (e) above. Working out part (e) using Model 2 parameters would give us the other expectation.

---

7. Suppose that $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} f(x|\theta)$.

   Let $f_\theta(\theta)$ denote the prior for θ.

   The likelihood may be written as $f_{x|\theta}(\vec{x}|\theta)$ and then the posterior is

$$f_{\theta|x}(\theta|\vec{x}) \propto f_{x|\theta}(\vec{x}|\theta) f_\theta(\theta)$$

   Suppose we wish to reparameterize the model in terms of τ where $\tau = g(\theta)$ for some function g.

   The prior for θ induces a prior for τ, given by

$$f_\tau(\tau) = f_\theta(g^{-1}(\tau)) \cdot \left| \frac{d}{d\tau} g^{-1}(\tau) \right|.$$

   The likelihood, as a function of τ, is written as $f_{x|\tau}(\vec{x}|\tau)$.

   The posterior, in terms of τ is then

$$\begin{aligned} f_{\tau|x}(\tau|\vec{x}) &\propto f_{x|\tau}(\vec{x}|\tau) f_\tau(\tau) \\ &= f_{x|\tau}(\vec{x}|\tau) f_\theta(g^{-1}(\tau)) \cdot \left| \tfrac{d}{d\tau} g^{-1}(\tau) \right|. \end{aligned} \tag{1}$$

   However, we are not in general going to get the same result if we reparameterize the posterior at the end:

$$f_{\tau|x}(\tau|\vec{x}) = f_{\theta|x}(g^{-1}(\tau)|\vec{x}) \cdot \left| \frac{d}{d\tau} g^{-1}(\tau) \right|. \tag{2}$$

   That is, we are not guaranteed that (1) is the same as (2). We will have this guarantee if we use the Jeffreys' prior for $f_\theta(\theta)$!

---

8. The Bayes risk is

$$\begin{aligned} R_\delta &= \mathsf{E}[L(\Theta, \delta(\vec{X}))] \\ &= \textstyle\int\int L(\theta, \delta(\vec{x})) f(\vec{x}, \theta)\, d\theta\, d\vec{x} \\ &= \textstyle\int\int (\delta(\vec{x}) - \theta)^2 f(\theta|\vec{x}) f(\vec{x})\, d\theta\, d\vec{x} \end{aligned}$$

   To minimize this with respect to $\delta = \delta(\vec{x})$, we need only to minimize the inner integral ($f(\vec{x})$ is pulled out of the integral)

$$\int (\delta(\vec{x}) - \theta)^2 f(\theta|\vec{x})\, d\theta$$

   This is equal to

$$\mathsf{E}[(\delta(\vec{x}) - \Theta)^2 | \vec{X} = \vec{x}].$$

   Expanding, we get

$$\delta^2 - 2\delta \mathsf{E}[\Theta|\vec{X} = \vec{x}] + \mathsf{E}[\Theta^2|\vec{X} = \vec{x}]$$

   Taking the derivative WRT δ and setting it equal to zero gives

$$2\delta - 2\mathsf{E}[\Theta|\vec{X} = \vec{x}] \stackrel{set}{=} 0$$

   Thus, we get that

$$\delta = \delta(\vec{x}) = \mathsf{E}[\Theta|\vec{X} = \vec{x}]$$

   which is the posterior Bayes estimator for θ.

---

9. (a)

$$\begin{aligned} R_\delta(\lambda) &= \mathsf{E}[(\delta(X)-\lambda)^2] = \mathsf{E}[(cX-\lambda)^2] \\ &= c^2\mathsf{E}[X^2] - 2c\lambda\mathsf{E}[X] + \lambda^2 \\ &= c^2(\lambda+\lambda^2) - 2c\lambda\cdot\lambda + \lambda^2 \\ &= c^2(\lambda+\lambda^2) - 2c\lambda^2 + \lambda^2 \end{aligned}$$

(b) Let $\delta_c = cX$. Note that

$$\begin{aligned} R_{\delta_c}(\lambda) &= c^2(\lambda+\lambda^2) - 2c\lambda^2 + \lambda^2 \\ &= (c-1)^2\lambda^2 + c^2\lambda \\ &= \geq c^2\lambda > \lambda \end{aligned}$$

if $c > 1$.

So, any δ_c with $c > 1$ would be dominated by δ_1.

(c) A decision rule δ^* is minimax if

$$\sup_\lambda R_{\delta^*}(\lambda) = \inf_\delta \sup_\lambda R_\delta(\lambda)$$

Remember, we will be taking the infimum over all decision rules of the form $\delta(X) = c(X)$. So, we can rewrite this as

$$\sup_\lambda R_{\delta^*}(\lambda) = \inf_c \sup_\lambda R_{\delta_c}(\lambda)$$

Note that

$$\sup_\lambda R_{\delta_c}(\lambda) = \sup_\lambda[(c-1)^2\lambda^2 + c^2\lambda] = \infty$$

Thus, there is no minimax rule.

By the way, minimizing the result of part (a) with respect to c gives $c = \frac{\lambda}{1+\lambda}$. Call the resulting decision rule δ^*. That is, define

$$\delta^*(X) = \frac{\lambda}{1+\lambda}X$$

which is 1 when "supped" over λ. However, you may recall from analysis that

$$\sup_x \inf_y f(x,y) \leq \inf_y \sup_x f(x,y).$$

This is what we are seeing ($1 < \infty$) in this case!

(d) **For this part, you were not supposed to be limited to a certain class of decision functions! I will continue using $\delta(\mathbf{x})$ of the form $\delta(\mathbf{x}) = \mathbf{cx}$ but we will not end up getting c to be constant!**

A standard computation shows that the posterior distribution of λ given x is $\Gamma(x+1,2)$. The Bayes risk is

$$R_\delta = \int\left[\int (cx-\lambda)^2 f(\lambda|x)\,d\lambda\right] f(x)\,dx$$

The inner integral is

$$\mathsf{E}[(cx-\Lambda)^2|X=x] = c^2x^2 - 2cx\mathsf{E}[\Lambda|X=x] + \mathsf{E}[\Lambda^2|X=x]$$

Taking the derivative WRT c and setting it equal to 0 gives us

$$c = \frac{\mathsf{E}[\Lambda|X=x]}{x^2} = \frac{(x+1)/2}{x^2} = \frac{x+1}{2x^2}.$$

(A second derivative shows that we are in fact minimizing here.)
Thus, the Bayes rule is

$$\delta(X) = cX = \frac{X+1}{2X^2}X = \frac{X+1}{2X}.$$

(e) The posterior Bayes estimator is

$$\widehat{\lambda} = \mathsf{E}[\Lambda|X] = \frac{X+1}{2}.$$

This should only match our decision rule, under squared error loss, if the possible decision rules were unrestricted. In this problem they were restricted to ones of the form $\delta(x) = cx$.