APPM 5720: Computational Bayesian Statistics

Solutions to Exam I Review Problems

1. A **conjugate prior**, for a particular model, is one such that the posterior distribution is from the same family of distributions as the prior.

   A **natural conjugate prior** is a conjugate prior with the additional property of having the same form as the likelihood when considered as a function of the parameter.

   A **non-informative** (also "uninformative" or "flat") **prior** for a parameter θ assigns equal probability to all possibilities for θ.

   A **mixture prior** is a weighted average of two or more densities.

   An **expert prior** is a prior that may not be computationally nice but reflects the actual opinon of an expert in the field that the Bayesian model is modeling!

   An **improper prior** is a prior that is not a proper probability density function in that it does not integrate (or sum) to 1. It is usually used to reflect euqally likely parameter possibilities over an infinite range.

   A **Jeffreys prior**, for a one-dimensional parameter θ, is defined as proportional to $\sqrt{I_n(\theta)}$ where $I_n(\theta)$ is the Fisher information. It is constructed so that the posterior is invariant under parameter transformations.

---

2. (a) An infinite sequence of binary random variables, $\{X_n\}_{n=1}^{\infty}$, is exchangeable if and only if there exists a cdf F on $[0,1]$ such that

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\, dF(\theta).$$

   (b) The pdf for any one of the X_i is

$$\begin{aligned} f(x|\theta_1,\theta_2) &= \tfrac{1}{2}\theta_1^x(1-\theta_1)^{1-x}\, I_{\{0,1\}}(x) + \tfrac{1}{2}\theta_2^x(1-\theta_2)^{1-x}\, I_{\{0,1\}}(x) \\ &= [\tfrac{1}{2}(\theta_1+\theta_2)]^x \cdot [1-\tfrac{1}{2}(\theta_1+\theta_2)]^{1-x}\, I_{\{0,1\}}(x) \end{aligned}$$

   This is the pdf for the *Bernoulli*(θ) distribution with $p = \frac{1}{2}(\theta_1+\theta_2)$. (Note: If you want to use θ here, then you should not use it as the moving variable in the deFinetti integral!)
   For any integers $1 \leq k \leq n$, we have

$$P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) \stackrel{iid}{=} p^k(1-p)^{n-k}.$$

   We wish to find a cdf F such that

$$\int_0^1 \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}\, dF(\theta) = p^k(1-p)^{n-k}.$$

   It is easy to verify that we can get this to hold for the step function cdf that puts all of its mass on p:

$$F(\theta) = \begin{cases} 0 & , \quad \theta < p \\ 1 & , \quad \theta \geq p \end{cases}$$

---

3. Disregarding the masses on the endpoints 0 and 1 for a moment, the exponential rate λ, restricted to $(0,1)$ has pdf

$$f(x) = \frac{1}{1-e^{-\lambda}}\,\lambda e^{-\lambda x}\, I_{(0,1)}(x)$$

That first part comes from renormalizing the pdf by

$$\int_0^1 \lambda e^{-\lambda x}\,dx = 1 - e^{-\lambda}.$$

However, because of the total mass of 1/2 on the two endpoints, we should have the continuous middle part integrate to 1/2 only and not 1, using the pdf

$$f(x) = \frac{1}{2}\,\frac{1}{1-e^{-\lambda}}\,\lambda e^{-\lambda x}\, I_{(0,1)}(x).$$

The cdf for any point x in $(0,1)$ will include the probability 1/4 obtained at $x=0$ plus the integral of this pdf from 0 to x.

(a) The cdf is

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{1}{4} + \frac{1}{2}\,\frac{1}{1-e^{-\lambda}}(1-e^{-\lambda x}) & , \quad 0 \le x < 1 \\ 1 & , \quad x \ge 1. \end{cases}$$

(b)

$$\begin{aligned} \mathsf{E}[X] &= \int_{-\infty}^{\infty} x\,dF(x) \\ &= 0 \cdot \tfrac{1}{4} + \tfrac{1}{2}\,\tfrac{1}{1-e^{-\lambda}}\,\lambda \int_0^1 x e^{-\lambda x}\,dx + 1 \cdot \tfrac{1}{4} \\ &= \tfrac{1}{2}\,\tfrac{1}{1-e^{-\lambda}}\,\tfrac{1}{\lambda}[1 - (\lambda+1)e^{-\lambda}] + \tfrac{1}{4} \end{aligned}$$

4.

$$Var\left[\sum_{i=1}^n X_i\right] = Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j)$$

So, we have to add up n^2 terms of the variance-covariance matrix.

On the diagonals, we have $Var[X_1]$, $Var[X_2]$, ..., $Var[X_n]$. We have seen that exchangeable random variables are always identically distributed. So, these terms are all the same and sum up to $n \cdot Var[X_1]$.

Also, by exchageability, all bivariate marginal distributions and hence all terms like $Cov(X_i, X_j)$, for $i \neq j$ are the same. There are $n^2 - n = n(n-1)$ of these off-diagnal terms. Thus, we have that

$$Var\left[\sum_{i=1}^n X_i\right] = nVar[X_1] + n(n-1)Cov(X_1, X_2),$$

as desired.

---

5. The prior for (α, β) is

$$\begin{aligned} f(\alpha,\beta) &= \tfrac{1}{\mathcal{B}(a,b)}\alpha^{a-1}(1-\alpha)^{b-1}\, I_{(0,1)}(\alpha)\cdot \tfrac{1}{\mathcal{B}(c,d)}\beta^{c-1}(1-\beta)^{d-1}\, I_{(0,1)}(\beta) \\ &\propto \alpha^{a-1}(1-\alpha)^{b-1}\, I_{(0,1)}(\alpha)\cdot \beta^{c-1}(1-\beta)^{d-1}\, I_{(0,1)}(\beta) \end{aligned}$$

(a)-(c) (I will make sure it is clear on the exam whether or not you should consider an entire random sample $X_1, X_2, \ldots, X_n$ or not.)

The joint pdf for a random sample of size x from the distribution with the given density is

$$f(\vec{x}|\alpha,\beta) = \prod_{i=1}^{n}\left[\alpha\, I_{\{0\}}(x) + (1-\alpha)\beta(1-\beta)^{x_i-1}\, I_{\{1,2,\ldots,\}}(x_i)\right]$$

Let n_0 be the number of x_i in the sample that are equal to zero. Then $n - n_0$ of thr x_i are in $\{1, 2, \ldots\}$.

The joint pdf can be written as

$$f(\vec{x}|\alpha,\beta) = \alpha^{n_0}\cdot[(1-\alpha)\beta]^{n-n_0}\,(1-\beta)^{\left[\left(\sum_{\{i:x_{i\neq 0}\}} x_i\right)-(n-n_0)\right]}$$

The posterior distribution for (α, β) is then given by

$$f(\alpha,\beta|\vec{x}) \quad\propto\quad f(\vec{x}|\alpha,\beta)\cdot f(\alpha,\beta)$$

which will turn out to be another another product of independent Beta distributions. So, yes, given the data $\vec{x}$, α and β are a prioriori independent with

$$\alpha \sim Beta(a^*, b^*) \qquad \text{and} \qquad \beta \sim Beta(c^*, d^*)$$

where

$$\begin{aligned} a^* &= a + n_0 \\ b^* &= b + n - n_0 \\ c^* &= c + n - n_0 \\ d^* &= d + \left(\sum_{\{i:x_{i\neq 0}\}} x_i\right) - (n - n_0) \end{aligned}$$

and the prior is a conjugate prior.

---

6. The joint pdf for the X_i is

$$f(\vec{x}|\theta) \overset{iid}{=} \prod_{i=1}^{n}\frac{1}{\theta}\, I_{(0,\theta)}(x_i) = \frac{1}{\theta^n}\, I_{(0,x_{(n)})}(x_{(1)})\, I_{(0,\theta)}(x_{(n)})$$

where $x_{(1)}$ and $x_{(2)}$ are the minimum and maximum, respectively of $x_1, x_2, \ldots, x_n$.

The posterior is

$$\begin{aligned}
f(\alpha,\beta|\vec{x}) &\propto f(\vec{x}|\alpha,\beta)\cdot f(\alpha,\beta) \\
&\propto \tfrac{1}{\theta^n}\, I_{(0,\theta)}(x_{(n)}) \cdot \tfrac{\alpha\beta^\alpha}{\theta^{\alpha+1}}\, I_{(\beta,\infty)}(\theta) \\
&\propto \tfrac{1}{\theta^{n+\alpha+1}}\, I_{(0,\theta)}(x_{(n)})\, I_{(\beta,\infty)}(\theta) \\
&= \tfrac{1}{\theta^{n+\alpha+1}}\, I_{(x_{(n)},\infty)}(\theta)\, I_{(\beta,\infty)}(\theta) \\
&= \tfrac{1}{\theta^{n+\alpha+1}}\, I_{(\max(x_{(n)},\beta),\infty)}(\theta)
\end{aligned}$$

Thus, $\theta|\vec{x}$ has a $Pareto(\alpha^*,\beta^*)$ distribution with

$$\begin{aligned}
\alpha^* &= n+\alpha \\
\beta^* &= \max(x_{(n)},\beta)
\end{aligned}$$

So, yes, this is a conjugate prior for the model!

---

7. We will need to compute the Fisher information $I_n(\beta)$. Since the X_i are iid, we have that $I_n(\beta) \stackrel{iid}{=} n\cdot I_1(\beta)$. Be careful, however, about using any computational simplifications for the Fisher information that depended on us being able to interchange $\frac{\partial}{\partial\beta}$. (For example, we do not know that the acore statistic has zero expectation nor do we necessarily have the second derivative simplification for the Fisher information.)

   Before we proceed, note that it doesn not make sense to take the log or the derivative of the indicator so we'll just keep it in mind on the side. For example, if one has a function

$$g(x) = \begin{cases} x^2 & , \quad 0<x<1 \\ x & , \quad x>1, \end{cases}$$

   the derivative is

$$g'(x) = \begin{cases} 2x & , \quad 0<x<1 \\ 1 & , \quad x\ge 1, \end{cases}$$

   but we do not take derivatives of the parts "$0<x<1$" and "$x\ge 1$"!

$$\ln f(x|\beta) = \ln\theta + \theta\ln\beta + (\theta-1)\ln x$$
$$\Downarrow$$
$$\frac{\partial}{\partial\beta}\ln f(x|\beta) = \frac{\theta}{\beta}.$$

   Now,

$$I_1(\beta) = \mathsf{E}\left[\left(\frac{\partial}{\partial\beta}\ln f(X_1|\beta)\right)^2\right] = \mathsf{E}\left[\left(\frac{\theta}{\beta}\right)^2\right] = \frac{\theta^2}{\beta^2}$$

   and

$$I_n(\beta) = \frac{n\theta^2}{\beta^2}.$$

The Jeffreys prior is

$$f_J(\beta) \propto \sqrt{I_n(\beta)} \propto \frac{1}{\beta}.$$

Since $\beta > 0$, this is an improper prior.

---

8. The joint pdf is

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

The prior is

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

The posterior is

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

So, the posterior distribution for θ, given $\vec{x}$ is $\Gamma(\alpha^*, \beta^*)$ where

$$\begin{aligned} \alpha^* &= \alpha + n \\ \beta^* &= \sum x_i + \beta. \end{aligned}$$

(a) The posterior Bayes estimator for θ is

$$\widehat{\theta}_{PBE} = \mathsf{E}[\Theta|\vec{x}] = \frac{\alpha^*}{\beta^*} = \frac{\alpha + n}{\sum X_i + \beta}.$$

(b) The posterior predictive density is

$$\begin{aligned} f(x_{n+1}|\vec{x}) &= \int_{-\infty}^{\infty} f(x_{n+1}|\theta) \cdot f(\theta|\vec{x}) \, d\theta \\ &= \int_0^\infty \theta e^{-\theta x_{n+1}} \, I_{(0,\infty)}(x_{n+1}) \cdot \tfrac{1}{\Gamma(\alpha^*)} (\beta^*)^{\alpha^*} \theta^{\alpha^*-1} e^{-\beta^*\theta} \, d\theta \\ &= \tfrac{1}{\Gamma(\alpha^*)} (\beta^*)^{\alpha^*} I_{(0,\infty)}(x_{n+1}) \int_0^\infty \underbrace{\theta^{\alpha^*} e^{-(x_{n+1}+\beta^*)\theta}}_{\substack{\text{looks like} \\ \Gamma(\alpha^*+1,\, x_{n+1}+\beta^*)}} \, d\theta \end{aligned}$$

$$= \tfrac{1}{\Gamma(\alpha^*)} (\beta^*)^{\alpha^*} \Gamma(\alpha^*+1) \tfrac{1}{(x_{n+1}+\beta^*)^{\alpha^*+1}} \, I_{(0,\infty)}(x_{n+1}) \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha^*+1)} (x_{n+1}+\beta^*)^{\alpha^*+1} \theta^{\alpha^*} e^{-(x_{n+1}+\beta^*)\theta} \, d\theta}_{1}$$

$$= \alpha^* (\beta^*)^{\alpha^*} \left(\tfrac{1}{x_{n+1}+\beta^*} \right)^{\alpha^*+1} \, I_{(0,\infty)}(x_{n+1})$$

(c)

$$
\begin{aligned}
P(X_{n+1} > 75|\vec{x}) &= \int_{75}^{\infty} \alpha^*(\beta^*)^{\alpha^*}(x_{n+1}+\beta^*)^{-\alpha^*-1}\,dx_{n+1} \\
&= -(\beta^*)^{\alpha^*}(x_{n+1}+\beta^*)^{-\alpha^*}\Big|_{75}^{\infty} \\
&= \frac{(\beta^*)^{\alpha^*}}{(75+\beta^*)^{\alpha^*}}.
\end{aligned}
$$

---

9. (a) $f(\mu) \propto 1$, $-\infty < \mu < \infty$

$$
\begin{aligned}
f(\mu|\vec{x}) &\propto f(\vec{x}|\mu)\cdot f(\mu) \\
&= \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]\cdot 1 \\
&= \exp\left[-\frac{n}{2}(\mu-\overline{x})^2\right]
\end{aligned}
$$

So

$$\mu|vecx \sim N(\overline{x}, 1/n).$$

The posterior Bayes estimator is

$$\widehat{\mu}_{PBE} = \mathsf{E}[\mu|\vec{X}] = \overline{X}.$$

(b) The natural conjugate prior is $N(\mu_0, \sigma_0^2)$ for some hyperparameters μ_0 and σ_0^2.

---

10. The joint pdf is

$$f(\vec{x}|\theta) = \prod_{i=1}^{n} e^{-(x_i-\theta)}\,I_{(\theta,\infty)}(x_i) = e^{-\sum x_i + n\theta}\,I_{(\theta,\infty)}(x_{(1)})$$

where $x_{(1)} = \min(x_1, x_2, \ldots, x_n)$.
The prior is

$$f(\theta) \propto 1, \quad -\infty < \theta < \infty.$$

The posterior is then

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

Let's find the constant that will make this a normalized pdf:

$$\int_{-\infty}^{x_{(1)}} e^{n\theta}\,d\theta = \frac{1}{n}e^{n\theta}\Big|_{-\infty}^{x_{(1)}} = \frac{1}{n}e^{nx_{(1)}} - 0 = \frac{1}{n}e^{nx_{(1)}}.$$

So, the posterior density is

$$f(\theta|\vec{x}) \;=\; ne^{-nx_{(1)}}e^{n\theta}\,I_{(-\infty,x_{(1)})}(\theta) = ne^{n(\theta-x_{(1)})}\,I_{(-\infty,x_{(1)})}(\theta)$$

We wish to find constants a and b such that

$$P(a < \theta < b|\vec{x}) = 0.95.$$

Upon inspection of the graph of the pdf for θ, we see that we will get the shortest (not a requirement for this solution) credible interval for θ if we take $b = x_{(1)}$. Then, we have

$$0.95 = \int_a^{x_{(1)}} ne^{n(\theta-x_{(1)})}\,d\theta = 1 - e^{n(a-x_{(1)})}$$

which implies that

$$a = x_{(1)} + \frac{1}{n}\ln(0.05).$$

The 95% credible interval for θ is

$$(x_{(1)} + \frac{1}{n}\ln(0.05), x_{(1)})$$