AMS206B

Let X_i denote the number of fires in a town for week i, $i = 1, \ldots, n$. Suppose $x_1, \ldots, x_n$ be an iid sample from a Poisson(θ) distribution.

1. Find the Jeffreys prior for θ, $\pi_J(\theta)$. *[10 points]*

2. Use the prior in part (1) to find the resulting posterior distribution $\pi_J(\theta \mid \boldsymbol{x})$. *[20 points]*

3. Under the posterior distribution in part (2), find the Bayes estimate of θ with the following loss function:
$$L(\theta, a) = \theta(\theta - a)^2 \,.$$
[30 points]

4. Now assume instead that the prior on θ is Exponential, with density $\pi(\theta \mid \lambda) = \lambda \exp(-\lambda\, \theta)$ for known $\lambda > 0$. Suppose that, under the Poisson sampling model above, and with the same sample $\boldsymbol{x} = (x_1, \ldots, x_n)$ of data, it is desired to test $H_0 : \theta \le c$ versus $H_1 \colon \theta > c$ for some $c > 0$. Let $F(t \mid \alpha, \beta)$ be the CDF of the Gamma distribution, in the parameterization in which the mean is $\frac{\alpha}{\beta}$.

   (a) Using $F(t \mid \alpha, \beta)$ and having observed $\boldsymbol{x}$, give an explicit expression for the resulting Bayes factor in favor of H_0. *[35 points]*

   (b) If this Bayes factor worked out numerically to be 6.25, what would your interpretation be about the weight of data evidence for or against H_0? *[5 points]*

AMS256

Consider a two-way ANOVA model

$$y_{i,j} = \alpha_i + \tau_j + \epsilon_{i,j}, \quad i = 1, 2; j = 1, 2. \tag{1}$$

The parameters, α_i and τ_j are unknown. Suppose the errors, $\epsilon_{i,j} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$.

1. Express these observations into a general linear model form, $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\beta} = [\alpha_1, \alpha_2, \tau_1, \tau_2]^T$. That is, specify your $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{\epsilon}$. Identify the rank of $\boldsymbol{X}$. *[10 points]*

2. Show that for this design, if a function of $\boldsymbol{\beta}$, $c_1\alpha_1 + c_2\alpha_2 + d_1\tau_1 + d_2\tau_2$ is estimable, then $c_1 + c_2 = d_1 + d_2$. *[15 points]*

3. Consider a function of $\boldsymbol{\beta}$, $\alpha_1 - \alpha_2$. A g-inverse of the matrix $\boldsymbol{X}^T\boldsymbol{X}$ for the model in (1) is given below;

$$(\boldsymbol{X}^T\boldsymbol{X})^- = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & -2 & -2 \\ 0 & -2 & 3 & 1 \\ 0 & -2 & 1 & 3 \end{bmatrix}$$

   Find the best linear unbiased estimators of the function. Also, find its probability distribution (assume that σ^2 is known for this part). *[40 points]*

4. Is y_{11} the best linear unbiased estimator of $\alpha_1 + \tau_1$? Explain why or why not. *[15 points]*

5. Assume that σ^2 is unknown. Derive a test for $H0 : \alpha_1 = \alpha_2$ at significance level γ. That is, describe a test statistic, its distribution under the null and the alternative, and a rejection region. Expressing SSE in the form of vectors is enough for an answer. *[20 points]*

Solution 206B:

(1)

$$f(\boldsymbol{x} \mid \theta) = \prod f(x_i \mid \theta) \propto \theta^s \, \exp(-n\,\theta)\,,$$

in which $s = \sum_{i=1}^n x_i$. Thus the Fisher information is

$$I(\theta) = -\,\mathrm{E}\left(\frac{s}{\theta^2}\right) \propto \frac{1}{\theta}\,.$$

So the Jeffreys prior is

$$\pi_J(\theta) \propto \frac{1}{\sqrt{\theta}}\,.$$

(2)

$$\pi_J(\theta \mid \boldsymbol{x}) \propto \theta^{s-1/2} \exp(-n\,\theta)\,.$$

This is $\Gamma(s + 1/2, n)$, with the parameterization in which the kernel of the Gamma is $\theta^{\alpha-1} e^{-\beta\,\theta}$ and the mean is $\frac{\alpha}{\beta}$.

(3) As usual the optimal estimator minimizes posterior expected loss. Students will not be required to show that the continuity conditions are satisfied to allow differentiation under the integral sign. Having performed that operation, it now follows after some algebra that

$$\hat{\theta} = \frac{\mathrm{E}\,(\theta^2 \mid \boldsymbol{x})}{\mathrm{E}\,(\theta \mid \boldsymbol{x})}\,.$$

Students now need to remember that, in the Gamma parameterization above, the variance is $\frac{\alpha}{\beta^2}$, from which it follows that

$$\mathrm{E}\,(\theta^2) = \frac{\alpha}{\beta^2} + \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha(\alpha+1)}{\beta^2}$$

and

$$\frac{\mathrm{E}\,(\theta^2)}{\mathrm{E}\,(\theta)} = \frac{\alpha+1}{\beta}\,.$$

Plugging in the posterior values $\alpha = s + 1/2$ and $\beta = n$ then yields

$$\hat{\theta} = \frac{s + 3/2}{n}\,.$$

(4) (a) Writing Bayes's Theorem in odds form, with BF for the Bayes factor,

$$\frac{p(\theta \le c \mid \boldsymbol{x})}{p(\theta > c \mid \boldsymbol{x})} = \frac{p(\theta \le c)}{p(\theta > c)} \cdot BF\,,$$

from which

$$BF = \frac{p(\theta \le c \mid \boldsymbol{x})\, p(\theta > c)}{p(\theta > c \mid \boldsymbol{x})\, p(\theta \le c)}\,.$$

The Exponential(λ) prior is just $\Gamma(1, \lambda)$ in the parameterization we're using, and with this prior (by conjugate updating) the posterior is $\Gamma(s+1, n+\lambda)$, so immediately

$$BF = \frac{F(c \,|\, s+1, n+\lambda)\,[1 - F(c \,|\, 1, \lambda)]}{[1 - F(c \,|\, s+1, n+\lambda)]\,F(c \,|\, 1, \lambda)}\,.$$

Some students will notice that $F(c \,|\, 1, \lambda) = 1 - e^{-c\lambda}$, but this is not required for full credit.

(b) In the qualitative scale proposed by Jeffreys, a Bayes factor of 6.25 in favor of H_0 is *substantial* data evidence supporting the null hypothesis; in the words of Kass and Raftery it's *positive* evidence. But these interpretations are just conventional; the real meaning of 6.25 would depend on the consequences of any real-world decision that depended on whether $\theta \le c$. (The previous sentence would be nice to see in a student answer, but is not required for full credit.)

Solution 256:

1. $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where

$$\boldsymbol{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \tau_1 \\ \tau_2 \end{bmatrix}, \ \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \end{bmatrix}$$

2. $\boldsymbol{\lambda}^T\boldsymbol{\beta} = [c_1, c_2, d_1, d_2]\boldsymbol{\beta}$ is estimable iff $\exists \boldsymbol{a}$ s.t $\boldsymbol{\lambda} = \boldsymbol{X}^T\boldsymbol{a}$. That is,

$$\begin{aligned} a_1 + a_2 &= c_1, \\ a_3 + a_4 &= c_2, \\ a_1 + a_3 &= d_1, \\ a_2 + a_4 &= d_2. \end{aligned}$$

Thus, $a_1 + a_2 + a_3 + a_4 = c_1 + c_2 = d_1 + d_2$ for any estimable $\boldsymbol{\lambda}^T\boldsymbol{\beta}$.

3.

$$\boldsymbol{\lambda}^T\boldsymbol{\beta} = \begin{bmatrix} 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 - \alpha_2 \end{bmatrix}$$

LSE of $\boldsymbol{\beta}$ is

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^T\boldsymbol{X})^- \boldsymbol{X}^T\boldsymbol{y} \\ &= \frac{1}{4}\begin{bmatrix} 0 \\ 4(y_{21} + y_{22}) - 2(y_{11} + y_{21}) - 2(y_{12} + y_{22}) \\ -2(y_{21} + y_{22}) + 3(y_{11} + y_{21}) + (y_{12} + y_{22}) \\ -2(y_{21} + y_{22}) + (y_{11} + y_{21}) + 3(y_{12} + y_{22}) \end{bmatrix} \\ &= \frac{1}{4}\begin{bmatrix} 0 \\ 2(y_{21} + y_{22}) - 2(y_{11} + y_{12}) \\ 3y_{11} + y_{12} + y_{21} - y_{22} \\ y_{11} + 3y_{12} - y_{21} + y_{22} \end{bmatrix} \end{aligned}$$

$\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}}$ is the BLUE of $\boldsymbol{\lambda}^T\boldsymbol{\beta}$ where

$$\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} = \frac{1}{2}(y_{21} + y_{22}) - 2(y_{11} + y_{12})$$

From lecture, $\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} \sim \mathrm{N}(\boldsymbol{\lambda}^T\boldsymbol{\beta}, \sigma^2\boldsymbol{\lambda}^T\boldsymbol{\lambda})$. We can easily verify $\boldsymbol{\lambda}^T\boldsymbol{\lambda} = 1$.

4. $\alpha_1 + \tau_1 = \boldsymbol{\lambda}^T \boldsymbol{\beta}$ with $\boldsymbol{\lambda} = [1, 0, 1, 0]^T$. From part 2, we know $\alpha_1 + \tau_1$ is estimable. The BLUE of $\boldsymbol{\lambda}^T \boldsymbol{\beta}$ is $\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}}$ and unique. From part 3, the BLUE of $\boldsymbol{\lambda}^T \boldsymbol{\beta}$ is
$$\frac{3y_{11} + y_{12} + y_{21} - y_{22}}{4} \neq \bar{y}_{11\cdot\cdot}$$

5. Reject $H0$ if
$$F = \frac{(\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} - 0)^T H^{-1} (\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} - 0)}{SSE/(N-r)} > F_{s,N-r,\alpha},$$
where
$$\boldsymbol{\lambda}^T \hat{\boldsymbol{\beta}} = [1, -1, 0, 0] \hat{\boldsymbol{\beta}} = \frac{(y_{11} + y_{12}) - (y_{21} + y_{22})}{2} \text{ from part(3)},$$
$$H^{-1} = 1 \text{ from part(3)},$$
$$SSE = (\boldsymbol{y} - \hat{\boldsymbol{y}})^T (\boldsymbol{y} - \hat{\boldsymbol{y}}) = \sum_{i,j} (y_{i,j} - \bar{y}_{i,j})^2,$$
$s = 1$, $N = 4$ and $r = \text{rank}(\boldsymbol{X}) = 3$. The test statistic F follows the central F distribution under $H0$ with d.f. 1 and 1. If $H0$ is false, F follows the noncentral F distribution with the same degrees of freedom and the noncentrality parameter,
$$\delta = \frac{(\boldsymbol{\lambda}^T \boldsymbol{\beta})^2}{2\sigma^2} = \frac{(\alpha_1 - \alpha_2)^2}{2\sigma^2}.$$