STAT 576 SOLUTION FOR HOMEWORK 2

1 (Total Variation Distance)**.** Let F and G be two distributions over $\mathbb{R}$ with probability density functions f and g correspondingly. Define the total variation distance between F and G by

$$\|f-g\|_{TV} := \sup_{A\subset\mathbb{R}} |F(A)-G(A)|.$$

We ignore the measurability issues for this problem.

(a) Show that
$$\|f-g\|_{TV} = F(A)-G(A)$$
for the event $A=\{x : f(x)>g(x)\}$.

(b) Show that
$$\|f-g\|_{TV} = \int (f(x)-g(x))_+ d\mu(x),$$
where
$$(y)_+ = \begin{cases} y & \text{if } y\geq 0 \\ 0 & \text{if } y<0. \end{cases}$$

(c) Show that
$$\|f-g\|_{TV} = \frac{1}{2}\int |f(x)-g(x)|d\mu(x)$$

Solution.

(a) Let $S\subset\mathbb{R}$ be an arbitrary subset of $\mathbb{R}$. Notice that $S=(S\cap A)\cup(S\cap A^c)$. We have
$$\begin{aligned} F(S)-G(S) &= [F(S\cap A)+F(S\cap A^c)]-[G(S\cap A)+G(S\cap A^c)] \\ &= [F(S\cap A)-G(S\cap A)]+[F(S\cap A^c)-G(S\cap A^c)] \end{aligned}$$
Because $S\cap A^c\subseteq A^c$ and $f(x)\leq g(x)$ for any $x\in A^c$, we have $F(S\cap A^c)-G(S\cap A^c)\leq 0$. On the other hand,
$$\begin{aligned} F(S\cap A)-G(S\cap A) &= [F(A)-F(S^c\cap A)]-[G(A)-G(S^c\cap A)] \\ &= [F(A)-G(A)]-[F(S^c\cap A)-G(S^c\cap A)] \leq F(A)-G(A) \end{aligned}$$
Therefore, we have
$$F(S)-G(S)\leq F(A)-G(A)$$
for all $S\subseteq\mathbb{R}$. The result is now immediate.

(b) Using result (a), we have
$$\|f-g\|_{TV} = F(A)-G(A) = \int_A (f(x)-g(x))d\mu(x) = \int (f(x)-g(x))\mathbb{I}\{x\in A\}d\mu(x)$$
By observing $(f(x)-g(x))\mathbb{I}\{x\in A\}\equiv(f(x)-g(x))_+$, the result is immediate.

(c) Notice that
$$(f(x)-g(x))_+ = \frac{1}{2}\left[(f(x)-g(x))+|f(x)-g(x)|\right].$$
Therefore,
$$\|f-g\|_{TV} = \int (f(x)-g(x))_+ d\mu(x) = \frac{1}{2}\int (f(x)-g(x))d\mu(x) + \frac{1}{2}\int |f(x)-g(x)|d\mu(x).$$
The first term is 0 because $\int f(x)d\mu(x)=\int g(x)d\mu(x)=1$. The result is now immediate.

2 (Textbook Problems)**.** Finish Problem 1 in Chapter 4 of the textbook. (You may skip the ploting step in part (c).)

Solution.

(a) The posterior density is

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta) \propto \frac{1}{\prod_{i=1}^{t}\left[1+(y_i-\theta)^2\right]},$$

with the log posterior density

$$\log p(\theta \mid y) = -\sum_{i=1}^{5} \log\left[1+(y_i-\theta)^2\right] + C$$

for some constant C.
The first-order derivative is

$$\frac{d}{d\theta}\log p(\theta \mid y) = \sum_{i=1}^{5} \frac{2(y_i-\theta)}{1+(y_i-\theta)^2}.$$

The second-order derivative is

$$\frac{d^2}{d\theta^2}\log p(\theta \mid y) = -2\sum_{i=1}^{5} \frac{1-(\theta-y_i)^2}{\left[1+(\theta-y_i)^2\right]^2}$$

(b) By setting the first-order derivative to zero, we have

$$\sum_{i=1}^{5} \frac{2(y_i-\theta)}{1+(y_i-\theta)^2} = 0$$

or equivalently,

$$\sum_{i=1}^{5} \frac{\theta}{1+(\theta-y_i)^2} = \sum_{i=1}^{5} \frac{y_i}{1+(\theta-y_i)^2}.$$

We can solve this equation by the following iteration:

$$\theta^{(t+1)} \leftarrow \left(\sum_{i=1}^{5} \frac{1}{1+(\theta^{(t)}-y_i)^2}\right)^{-1} \sum_{i=1}^{5} \frac{y_i}{1+(\theta^{(t)}-y_i)^2}$$

Or one can use Newton-Raphson iteration. This will give a solution

$$\hat{\theta} = -0.138.$$

Note that this is the estimation for the mode of the posterior density (extending to the whole real line). The true mode of the posterior is at $\theta = 0$.

(c) Pluging the value $\hat{\theta} = -0.138$ into the second-order derivative, we have

$$\frac{d^2}{d\theta^2}\widehat{\log p(\theta \mid y)}\big|_{\hat{\theta}} \approx -1.375$$

Therefore, one can estimate the posterior by the normal density of

$$\mathcal{N}(-0.138, 1.375^{-1}) \sim \mathcal{N}(-0.138, 0.727)$$

when restricted to $[0, 1]$.

3 (Hierarchical Poisson)**.** Suppose a datasets contains the numbers of traffic accidents in J districts for the last year. Denote the number of accidents in district j by y_j.

(a) The number of accidents y_j can be modeled by a Poisson distribution with intensity λ_j. Write down the probability mass function for y_j given λ_j.

(b) Build a hierarchical model for $(y_1, \ldots, y_J)$, where the λ_j follows a Gamma(α, β) distribution with the hyperprior $p(\alpha, \beta)$.

(c) Compute the marginal distribution of y_1.
(d) Compute the joint posterior distribution density (in proportional form) for $(\alpha, \beta, \lambda_1, \ldots, \lambda_J)$.
(e) Compute the exact conditional posterior distribution density for $\lambda_1, \ldots, \lambda_J \mid \alpha, \beta, y_1, \ldots, y_J$.
(f) Compute the marginal posterior distribution for (α, β) in proportional form.

Solution.

(a) The probability mass function is

$$p[y_i \mid \lambda_i] = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}$$

(b) The hierarchical model is

$$\begin{aligned} \alpha, \beta &\sim p(\alpha, \beta) \\ \lambda_1, \lambda_2, \ldots, \lambda_J \mid \alpha, \beta &\sim \text{Gamma}(\alpha, \beta) \\ y_j \mid \lambda_j &\sim \text{Poisson}(\lambda_j), \quad \text{for } j = 1, \ldots, J, \text{,independently} \end{aligned}$$

(c) We first give its conditional density given α, β:

$$p(y_1 \mid \alpha, \beta) = \int p(y_1 \mid \lambda_1) p(\lambda_1 \mid \alpha, \beta) d\mu(\lambda_1) = \int \frac{\lambda_1^{y_1} e^{-\lambda_1}}{y_1!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_1^{\alpha-1} e^{-\beta\lambda_1} d\mu(\lambda_1) = \frac{\beta^\alpha}{(\beta+1)^{\alpha+y_1}} \frac{\Gamma(\alpha+y_1)}{y_1!\Gamma(\alpha)}$$

The marginal of y_1 is the integral of above with respect to $p(\alpha, \beta)$.

(d) The joint prior is

$$p(\alpha, \beta, \lambda_1, \ldots, \lambda_J) \propto p(\alpha, \beta) \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^J \prod_{j=1}^J \lambda_j^{\alpha-1} \cdot e^{-\beta \sum_j \lambda_j}.$$

The likelihood is

$$p(y_1, \ldots, y_J \mid \lambda_1, \ldots, \lambda_J) = \prod_{j=1}^J \lambda_j^{y_j} e^{-\sum_j \lambda_j}$$

Therefore, the joint posterior gives

$$\begin{aligned} p(\alpha, \beta, \lambda_1, \ldots, \lambda_J \mid y_1, \ldots, y_J) &\propto p(\alpha, \beta, \lambda_1, \ldots, \lambda_J) p(y_1, \ldots, y_J \mid \lambda_1, \ldots, \lambda_J) \\ &\propto p(\alpha, \beta) \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^J \prod_{j=1}^J \lambda_j^{\alpha+y_j-1} \cdot e^{-(\beta+1)\sum_j \lambda_j} \end{aligned}$$

(e) From the joint posterior density, we can figure out

$$p(\lambda_1, \ldots, \lambda_J \mid \alpha, \beta, y_1, \ldots, y_J) \propto \prod_{j=1}^J \lambda_j^{\alpha+y_j-1} \cdot e^{-(\beta+1)\sum_j \lambda_j}$$

It corresponds to $\lambda_j \sim \text{Gamma}(\alpha + y_j, \beta + 1)$ independently. The exact density is, therefore,

$$p(\lambda_1, \ldots, \lambda_J \mid \alpha, \beta, y_1, \ldots, y_J) = \prod_{j=1}^J \frac{(\beta+1)^{\alpha+y_j}}{\Gamma(\alpha+y_j)} \lambda_j^{\alpha+y_j-1} e^{-(\beta+1)\lambda_j}$$

(f) The marginal posterior is given by

$$\begin{aligned} p(\alpha, \beta \mid y_1, \ldots, y_J) &= \frac{p(\alpha, \beta, \lambda_1, \ldots, \lambda_J \mid y_1, \ldots, y_J)}{p(\lambda_1, \ldots, \lambda_J \mid \alpha, \beta, y_1, \ldots, y_J)} \\ &\propto p(\alpha, \beta) \frac{\beta^{J\alpha}}{(\beta+1)^{J\alpha + \sum_j y_j}} \frac{\prod_{j=1}^J \Gamma(\alpha+y_j)}{(\Gamma(\alpha))^J} \end{aligned}$$

4 (Textbook Problems)**.** Finish Problems 4 and 5 in Chapter 5 of the textbook.

Solution.
Problem 4:

(a) Yes. The joint distribution can be written as

$$p(\theta_1, \ldots, \theta_{2J}) = \binom{2J}{J}^{-1} \sum_{\pi} \left(\prod_{j=1}^{J} \mathcal{N}(\theta_{\pi(j)} \mid 1, 1) \prod_{j=J+1}^{2J} \mathcal{N}(\theta_{\pi(j)} \mid -1, 1) \right),$$

where the summation is over all permutations π.

(b) Consider the covariance between θ_1 and θ_2. Let π be that the permutation in part (a) that indicates memberships. Then

$$\mathrm{Cov}(\theta_1, \theta_2) = \mathbb{E}[\mathrm{Var}(\theta_1, \theta_2 \mid \pi)] + \mathrm{Cov}[\mathbb{E}(\theta_1 \mid \pi), \mathbb{E}(\theta_2 \mid \pi)].$$

The first term is zero because the variables are independent given the memberships. The second term is

$$\mathrm{Cov}[\mathbb{E}(\theta_1 \mid \pi), \mathbb{E}(\theta_2 \mid \pi)] = \mathbb{E}[\mathbb{E}(\theta_1 \mid \pi)\mathbb{E}(\theta_2 \mid \pi)] - \mathbb{E}[\theta_1]\mathbb{E}[\theta_2] = \frac{2\binom{2J-2}{J-2}}{\binom{2J}{J}} - \frac{2\binom{2J-2}{J-1}}{\binom{2J}{J}} = -\frac{1}{2J-1} < 0$$

Therefore, $\mathrm{Cov}(\theta_1, \theta_2) < 0$, which contradicts the result in Problem 5.

(c) The covariance in part (b) converges to 0 when $J \to \infty$.

Problem 5:
Using the law of total covariance, we get

$$\mathrm{Cov}(\theta_1, \theta_2) = \mathbb{E}[\mathrm{Cov}(\theta_1, \theta_2 \mid \phi)] + \mathrm{Cov}[\mathbb{E}(\theta_1 \mid \phi), \mathbb{E}(\theta_2 \mid \phi)] = 0 + \mathrm{Var}[\mathbb{E}(\theta \mid \phi)]$$

The second term is always nonnegative.