STK 4021 - Applied Bayesian Analysis and Numerical Methods
Thordis L. Thorarinsdottir
Fall 2014

Problem Set 12

Problem 1 (binomial and multinomial models). Suppose data $(y_1, \dots, y_J)$ follow a multinomial distribution with parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_J)$. Also suppose that $\boldsymbol{\theta}$ has a Dirichlet prior distribution. Let $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$.

(a) Write the marginal posterior distribution for α.

Solution: The sampling model is

$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{(y_1 + \cdots + y_J)!}{y_1! \cdots y_J!} \theta_1^{y_1} \cdots \theta_J^{y_J}$$

and the prior is of the form

$$p(\boldsymbol{\theta}|\mathbf{a}) \propto \theta_1^{a_1-1} \cdots \theta_J^{a_J-1}.$$

The posterior distribution thus fulfils

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \theta_1^{y_1+a_1-1} \cdots \theta_J^{y_J+a_J-1},$$

which makes it a Dirichlet$(y_1+a_1, \dots, y_J+a_J)$ distribution. It follows from the properties of the Dirichlet distribution that the marginal distribution of $(\theta_1, \theta_2, 1-\theta_1-\theta_2)$ is also a Dirichlet distribution

$$p(\theta_1, \theta_2|\mathbf{y}) \propto \theta_1^{y_1+a_1-1}\theta_2^{y_2+a_2-1}(1-\theta_1-\theta_2)^{y_s+a_s-1},$$

where $y_s = y_3 + \cdots y_J$ and $a_s = a_3 + \cdots + a_J$. This can also be proved directly using induction, by first integrating out θ_n, then θ_{n-1} and so on.

Now do a change of variables to $(\alpha, \beta) = (\frac{\theta_1}{\theta_1+\theta_2}, \theta_1 + \theta_2)$. The Jacobian of this transformation is $\frac{1}{\beta}$, so the transformed density is

$$\begin{aligned} p(\alpha, \beta|y) &\propto \beta(\alpha\beta)^{y_1+a_1-1}((1-\alpha)\beta)^{y_2+a_2-1}(1-\beta)^{y_s+a_s-1} \\ &= \alpha^{y_1+a_1-1}(1-\alpha)^{y_2+a_2-1}\beta^{y_1+y_2+a_1+a_2-1}(1-\beta)^{y_s+a_s-1} \\ &\propto \text{Beta}(\alpha|y_1+a_2, y_2+a_2)\text{Beta}(\beta|y_1+y_2+a_1+a_2, y_s+a_s). \end{aligned}$$

Since the posterior density divides into separate factors for α and β, they are independent, and $\alpha|y \sim \text{Beta}(y_1+a_2, y_2+a_2)$.

(b) Show that the distribution in (a) is identical to the posterior distribution for α obtained by treating y_1 as an observation from the binomial distribution with probability α and sample size $y_1 + y_2$, ignoring the data $y_3, \dots, y_J$.

Solution: The $\text{Beta}(y_1+a_2, y_2+a_2)$ posterior distribution can also be derived from a $\text{Beta}(a_1, a_2)$ prior and a binomial observation y_1 with sample size $y_1 + y_2$.

Problem 2 (**Poisson models**).

(a) Suppose $y|\theta \sim \text{Po}(\theta)$. Find Jeffreys' prior density for θ, and then find α and β for which the $\Gamma(\alpha, \beta)$ density is a close match to Jeffreys' density.

Solution: The Poisson sampling density is $p(y|\theta) = \theta^y e^{-\theta}/y!$, and so $I(\theta) = -\mathbb{E}(\,\mathrm{d}^2 \log p(Y|\theta)/\,\mathrm{d}\theta^2) = \mathbb{E}(Y/\theta^2)$. Thus, $p_J(\theta) \propto \sqrt{I(\theta)} = \theta^{-1/2}$ which is an (improper) $\Gamma(1/2, 0)$ density.

(b) Suppose $y|\theta \sim \text{Po}(\theta)$ and $\theta \sim \Gamma(\alpha, \beta)$. Then the marginal (prior predictive) distribution of y is negative binomial with parameters α and β. Use the formulas

$$\begin{aligned}\mathbb{E}(\theta) &= \mathbb{E}(\mathbb{E}(\theta|y))\\ \text{var}(\theta) &= \mathbb{E}(\text{var}(\theta|y)) + \text{var}(\mathbb{E}(\theta|y))\end{aligned}$$

to derive the mean and the variance of this marginal distribution.

Solution: We have $\mathbb{E}(Y) = \theta$, $\text{Var}(Y) = \theta$, $\mathbb{E}(\theta) = \alpha/\beta$, $\text{Var}(\theta) = \alpha/\beta^2$. Thus,

$$\mathbb{E}Y = \mathbb{E}(\mathbb{E}(Y|\theta)) = \mathbb{E}(\theta) = \alpha/\beta,$$

and

$$\text{Var}(Y) = \mathbb{E}(\text{Var}(Y|\theta)) + \text{Var}(\mathbb{E}(Y|\theta)) = \mathbb{E}(\theta) + \text{Var}(\theta) = \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2} = \frac{\alpha}{\beta(\beta+1)}.$$

Problem 3 (**Poisson and binomial distributions**). A student sits on a street corner for an hour and records the number of bicycles b and the number of other vehicles v that go by. Two models are considered:

- The outcomes b and v have independent Poisson distributions, with unknown means θ_b and θ_v.
- The outcome b has a binomial distribution, with unknown probability p and sample size $b + v$.

Show that the two models have the same likelihood if we define $p = \frac{\theta_b}{\theta_v + \theta_b}$.

Solution: The likelihood under the binomial model is

$$p\left(b \middle| \frac{\theta_b}{\theta_v + \theta_b}, b+v\right) = \binom{b+v}{b}\left(\frac{\theta_b}{\theta_v+\theta_b}\right)^b\left(1 - \frac{\theta_b}{\theta_v + \theta_b}\right)^v = \binom{b+v}{b}\frac{\theta_b^b\theta_v^v}{(\theta_v+\theta_b)^{b+v}}.$$

As b and v and independent, we have $b + v \sim \text{Po}(\theta_v + \theta_b)$ under the Poisson model. Now,

$$\begin{aligned}p(b = k|\theta_v, \theta_b, b+v = n) &= \frac{P(b = k, b+v = n|\theta_v, \theta_b)}{P(b+v = n|\theta_v,\theta_b)} = \frac{P(b = k|\theta_v, \theta_b)P(v = n-b|\theta_v, \theta_b)}{P(b+v = n|\theta_v, \theta_b)}\\ &= \frac{e^{-\theta_b}\theta_b^k/k!e^{-\theta_v}\theta_v^{n-k}/(n-k)!}{e^{-(\theta_v+\theta_b)}(\theta_v+\theta_b)^n/n!} = \binom{n}{k}\frac{\theta_b^k\theta_v^{n-k}}{(\theta_v+\theta_b)^n}.\end{aligned}$$

Problem 4 (discrete mixture models).

(a) If $p_m(\theta)$, for $m = 1, \ldots, M$ are conjugate prior densities for the sampling model $p(y|\theta)$, show that the class of finite mixture prior densities given by

$$p(\theta) = \sum_{m=1}^{M} \lambda_m p_m(\theta)$$

is also a conjugate class, where the λ_m's are nonnegative weights that sum to 1.

Solution: For each m, denote by $p_m(\theta|y)$ the posterior density corresponding to the prior $p_m(\theta)$, that is, $p_m(\theta|y) = p(y|\theta)p_m(\theta)/p_m(y)$, where $p_m(y) = \int p(y|\theta)p_m(\theta)\,\mathrm{d}\theta$ is the marginal density of y. If $p(\theta) = \sum_m \lambda_m p_m(\theta)$, then the posterior is proportional to $\sum_m \lambda_m p_m(\theta)p(y|\theta) = \sum_m \lambda_m p_m(y)p_m(\theta|y)$. This is a mixture of the posterior densities $p_m(\theta|y)$ with weigths proportional to $\lambda_m p_m(y)$. Since each $p_m(\theta)$ is a conjugate prior, each posterior $p_m(\theta|y)$ is from the same family as $p_m(\theta)$. The posterior mixture is thus from the same family of finite mixtures as the prior mixture.

(b) Use the mixture form to create a bimodal prior density for a normal mean, that is thought to be near 1, with a standard deviation of 0.5, but has a small probability of being near -1, with the same standard deviation. If the variance of each observation $y_1, \ldots, y_{10}$ is known to be 1, and their observed mean is $\bar{y} = -0.25$, derive your posterior distribution for the mean.

Solution: Consider e.g. $p_1(\theta) \sim \mathcal{N}(1, 0.5^2)$, $p_2(\theta) \sim \mathcal{N}(-1, 0.5^2)$ with $\lambda_1 = 0.9$ and $\lambda_2 = 0.1$. We then obtain $p_1(\theta|y) \sim \mathcal{N}(1.5/14, 1/14)$ and $p_2(\theta|y) \sim \mathcal{N}(-6.5/14, 1/14)$. The marginal distributions are $p_1(y) \sim \mathcal{N}(1, 0.5^2{+}1/10)$ and $p_2(y) \sim \mathcal{N}(-1, 0.5^2{+}1/10)$ such that $p_1(-0.25) = 0.072$ and $p_2(-0.25) = 0.302$. The posterior weights are thus

$$\frac{0.9 \cdot 0.072}{0.9 \cdot 0.072 + 0.1 \cdot 0.302} = 0.68 \text{ and } \frac{0.1 \cdot 0.302}{0.9 \cdot 0.072 + 0.1 \cdot 0.302} = 0.32.$$