STK9021 - Applied Bayesian Analysis and Numerical Methods

CHAPTER 3: EXERCISE 1 AND 2

MOHAMMED AHMED KEDIR

3.10 Exercises

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

   (a) Write the marginal posterior distribution for α.

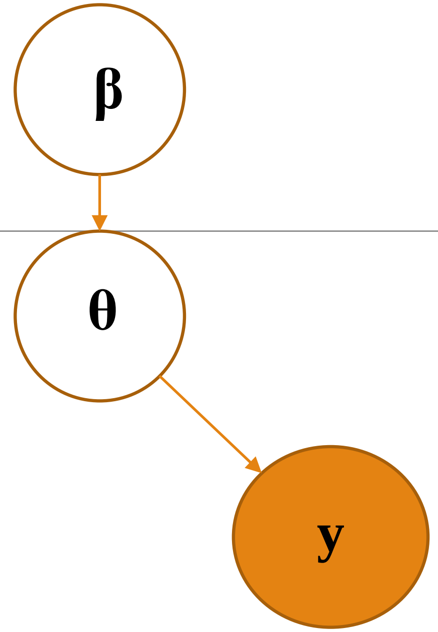
   (b) Show that this distribution is identical to the posterior distribution for α obtained by treating y_1 as an observation from the binomial distribution with probability α and sample size $y_1 + y_2$, ignoring the data $y_3, \ldots, y_J$.

   This result justifies the application of the binomial distribution to multinomial problems when we are only interested in two of the categories; for example, see the next problem.

Solution: (1-A)

- Multinomial distribution: is a convenient way to represent discrete variable that can take on of possible mutually exclusive states. e.g. 1 of K scheme

$$y = (0,0,0,1,0,0)^{T}$$



★ $y_1, \ldots\ldots, y_J \mid \theta$ Data which follows a multinomial distributions

★ $\theta_{1..J}$ Multinomial distribution parameters

$$p(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{j=1}^{J} \theta_j^{y_j}$$

★ $\theta \mid \beta$ $Dir(\theta \mid \beta) \propto \prod_{j=1}^{J} \theta_j^{\beta_j - 1}$ Prior distributions on multinomial distribution (Dirichlet)

We are interested in marginal posterior distribution for $\alpha = \dfrac{\theta_1}{\theta_1 + \theta_2}$

Cont'

Posterior distribution:

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$$
$$\propto \prod_{j=1}^{J} \theta_j^{y_j} \prod_{j=1}^{J} \theta_j^{\beta_j - 1}$$
$$= \prod_{j=1}^{J} \theta_j^{y_j + \beta_j - 1}$$

Which is Dirichlet $Dir(\boldsymbol{\theta}|y_j + \beta_j)$

The joint probability density can be written as

$$f(\theta_1, \theta_{2,} \ldots \theta_m) = f_1(\theta_1) f_2(\theta_2 \mid \theta_1) f_3(\theta_3 \mid \theta_1, \theta_2) \ldots f_{m-1}(\theta_{m-1} \mid \theta_1, \theta_2, \ldots, \theta_{m-2})$$

e.g. For three variables, $(\theta_1, \theta_2, \theta_3)$

$$p(\theta_1, \theta_2, 1 - \theta_1 - \theta_2) \propto \theta_1^{\beta_1 - 1} \theta_2^{\beta_2 - 1} (1 - \theta_1 - \theta_2)^{\beta_3 - 1}$$

Cont': Marginal distributions of θ_j

Marginal distribution for the earlier three variables

$$p(\theta_1,\theta_2,1-\theta_1-\theta_2)\propto \theta_1^{\beta_1-1}\theta_2^{\beta_2-1}(1-\theta_1-\theta_2)^{\beta_3-1}$$

$$p(\theta_1)=\int_{-\infty}^{+\infty} p(\theta_1,\theta_2)d\theta_2$$

$$p(\theta_1)\propto \int_{-\infty}^{+\infty} \theta_1^{\beta_1-1}\theta_2^{\beta_2-1}(1-\theta_1-\theta_2)^{\beta_3-1}d\theta_2=\theta_1^{\beta_1-1}\int_0^{1-\theta_1}\theta_2^{\beta_2-1}(1-\theta_1-\theta_2)^{\beta_3-1}d\theta_2$$

Let $\theta_2=(1-\theta_1)u$

$$=\theta_1^{\beta_1-1}\int_0^1(1-\theta_1)^{\beta_2}u^{\beta_2-1}(1-\theta_1-(1-\theta_1)u)^{\beta_3-1}du$$

$$=\theta_1^{\beta_1-1}(1-\theta_1)^{\beta_2}\int_0^1 u^{\beta_2-1}(1-\theta_1-(1-\theta_1)u)^{\beta_3-1}du=\theta_1^{\beta_1-1}(1-\theta_1)^{\beta_2}\int_0^1 u^{\beta_2-1}((1-\theta_1)(1-u))^{\beta_3-1}du$$

$$=\theta_1^{\beta_1-1}(1-\theta_1)^{\beta_2+\beta_3-1}\int_0^1 u^{\beta_2-1}(1-u)^{\beta_3-1}du$$ But $\int_0^1 u^{\beta_2-1}(1-u)^{\beta_3-1}du=Beta(\beta_2,\beta_3)$

$p(\theta_1)=Beta(\beta_1,\beta_2+\beta_3)$ The marginal distribution of single θ_j is Beta distribution

Cont'

Similarly it can be shown that the marginal distribution of a subvector of θ is Dirichlet.

The marginal posterior distribution for three variables.

Let find the posterior marginal distribution of $\theta_1, \theta_2, \theta_{\text{others}}$, $\theta_{others} = 1 - \theta_1 - \theta_2$

The posterior marginal distribution will be Dirichlet.

$$= \prod_{j=1}^{J} \theta_j^{y_j + \beta_j - 1}$$

Hence,

$$p(\theta_1, \theta_2 \mid y) \propto \theta_1^{y_1 + \beta_1 - 1} \theta_2^{y_2 + \beta_2 - 1} (1 - \theta_1 - \theta_2)^{y_{others} + \beta_{others} - 1}$$

But we are interested in

$$p\left(\alpha = \frac{\theta_1}{\theta_1 + \theta_2}, \tau = \theta_1 + \theta_2 \mid y\right)$$

By change of variables: The transformation is one-to-one because we can solve for θ_1, θ_2, in terms of α, τ by $\theta_1 = \alpha\tau$

$$\theta_2 = (\alpha - 1)\tau$$

Cont.'

$$p\left(\alpha,\tau \mid y\right)=\frac{p\left(\theta_1,\theta_2 \mid y\right)}{\left|\det(J)\right|}\Bigg|_{\substack{\theta_1=\alpha\tau \\ \theta_2=(\alpha-1)\tau}}$$

$$p\left(\alpha=\frac{\theta_1}{\theta_1+\theta_2},\tau=\theta_1+\theta_2 \mid y\right) \quad \begin{matrix}\theta_1=\alpha\tau \\ \theta_2=(\alpha-1)\tau\end{matrix}$$

The Jacobean

$$\frac{\partial(\alpha,\tau)}{\partial(\theta_1,\theta_2)}=\begin{vmatrix} \partial\alpha/\partial\theta_1 & \partial\alpha/\partial\theta_2 \\ \partial\tau/\partial\theta_1 & \partial\tau/\partial\theta_2 \end{vmatrix}=\begin{vmatrix} \frac{\theta_2}{(\theta_1+\theta_2)^2} & \frac{\theta_1}{(\theta_1+\theta_2)^2} \\ 1 & 1 \end{vmatrix}=\left|\frac{\theta_2-\theta_1}{(\theta_1+\theta_2)^2}\right|=\left|\frac{(\alpha-1)\tau-\alpha\tau}{\tau^2}\right|=\left|\frac{1}{\tau}\right|=\frac{1}{\tau}$$

$$=\tau\left(\alpha\tau\right)^{y_1+\beta_1-1}\left(\left(\alpha-1\right)\tau\right)^{y_2+\beta_2-1}\left(1-\alpha\tau-\left(\alpha-1\right)\tau\right)^{y_{others}+\beta_{others}-1}$$

$$=\left(\alpha\right)^{y_1+\beta_1-1}\left(\alpha-1\right)^{y_2+\beta_2-1}\tau^{y_1+y_2+\beta_1+\beta_2-1}\left(1-\tau\right)^{y_{others}+\beta_{others}-1}$$

$$p\left(\alpha,\tau \mid y\right)\propto Beta\left(\alpha \mid y_1+\beta_1,y_2+\beta_2\right)Beta\left(\tau \mid y_1+y_2+\beta_1+\beta_2,y_{others}+\beta_{others}\right)$$

$$p\left(\alpha \mid y\right)=\int Beta\left(\alpha \mid y_1+\beta_1,y_2+\beta_2\right)Beta\left(\tau \mid y_1+y_2+\beta_1+\beta_2,y_{others}+\beta_{others}\right)d\tau$$

$$p\left(\alpha \mid y\right)=Beta\left(\alpha \mid y_1+\beta_1,y_2+\beta_2\right)\int Beta\left(\tau \mid y_1+y_2+\beta_1+\beta_2,y_{others}+\beta_{others}\right)d\tau$$

Hence,

$$p\left(\alpha \mid y\right)=Beta\left(\alpha \mid y_1+\beta_1,y_2+\beta_2\right)$$

(1-B)

(b) Show that this distribution 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 $y_3, \ldots, y_J$.

$y_1, y_2 \mid \alpha$	Data which follows a binomial distributions
α	Binomial distribution parameters
$\alpha \mid \beta_1, \beta_2$	Beta Prior

From earlier class on conjugacy we know that the posterior distribution will be Beta distribution

$$p(\alpha \mid y_1, y_2) \propto \alpha^{y_1} (1-\alpha)^{y_2} \alpha^{\beta_1 - 1} (1-\alpha)^{\beta_2 - 1} = \alpha^{y_1 + \beta_1 - 1} (1-\alpha)^{y_2 + \beta_2 - 1} = Beta(\alpha \mid y_1 + \beta_1, y_2 + \beta_2)$$

Which is A Beta distribution similar to the result 1-A

Exercise 2:

2. Comparison of two multinomial observations: on September 25, 1988, the evening of a presidential campaign debate, ABC News conducted a survey of registered voters in the United States; 639 persons were polled before the debate, and 639 different persons were polled after. The results are displayed in Table 3.2. Assume the surveys are independent simple random samples from the population of registered voters. Model the data with two different multinomial distributions. For $j = 1, 2$, let α_j be the proportion of voters who preferred Bush, out of those who had a preference for either Bush or Dukakis at the time of survey j. Plot a histogram of the posterior density for $\alpha_2 - \alpha_1$. What is the posterior probability that there was a shift toward Bush?

Soln:

ABC news survey of registered voters before and after debate

Survey	Bush	Dukakis	No opinion/other	Total
pre-debate	294	307	38	639
post-debate	288	332	19	639

Table 3.2 *Number of respondents in each preference category from ABC News pre- and post-debate surveys in 1988.*

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

$$\propto \prod_{j=1}^{J} \theta_j^{y_j}$$

$$= \prod_{j=1}^{J} \theta_j^{y_j}$$

Which is a multinomial distribution

$$= \prod_{j=1}^{J} \theta_j^{y_j} = \prod_{j=1}^{J} \theta_j^{y_j+1-1}$$

Which is a Dirichlet distribution with $Dir(\boldsymbol{\theta} | y_j + 1)$

Since they are independent:

Pre-debate: $p\left(\theta_1^{pre}, \theta_2^{pre}, \theta_3^{pre} \mid y\right) = Direchlet(295, 308, 39)$

Post-debate: $p\left(\theta_1^{post}, \theta_2^{post}, \theta_3^{post} \mid y\right) = Direchlet(289, 333, 20)$

Cont'

For j=1,2, let α_j be the proportion of voters who prefer Bush out of Total of who preferred Bush and Dukakis

People who prefer bush **Pre debate** $\alpha_1 = \dfrac{\theta_1^{pre}}{\theta_1^{pre} + \theta_2^{pre}}$, θ_1^{pre} = pre debate preference Bush, θ_2^{pre} = pre debate preference Dukakis

People who prefer bush **Post debate** $\alpha_2 = \dfrac{\theta_1^{post}}{\theta_1^{post} + \theta_2^{post}}$, θ_1^{post} = post debate preference Bush, θ_2^{post} = post debate preference Dukakis

But we are interested in $z = \alpha_2 - \alpha_1$

From the earlier exercise we know that $p(\alpha_j \mid y) = Beta(\alpha_j \mid y_1 + \beta_1, y_2 + \beta_2)$ →

$$p(\alpha_1 \mid y^{pre}) = Beta(\alpha_1 \mid 295, 308)$$
$$p(\alpha_2 \mid y^{post}) = Beta(\alpha_2 \mid 289, 333)$$

Analytic solution:

$$p(z = (\alpha_2 - \alpha_1), w = \alpha_2) == \left.\frac{p(\alpha_1, \alpha_2 \mid y)}{|\det(J)|}\right|_{\substack{\alpha_1 = z+w \\ \alpha_2 = w}}$$

$$p(z) = \int \left.\frac{p(\alpha_1, \alpha_2 \mid y)}{|\det(J)|}\right|_{\substack{\alpha_1 = z+w \\ \alpha_2 = w}} dw = \int_0^1 (z+w)^{294} (1 - z + w)^{307} w^{288} (1-w)^{332} \, dw$$

Numerical Solution: with R

Step 1: Simulate N number of samples from $p\left(\alpha_1 \mid y^{pre}\right) = Beta\left(295, 308\right)$

```
alpha1<-rbeta(10000,295,308)
```

Step2 : Simulate N number of samples from $p\left(\alpha_2 \mid y^{post}\right) = Beta\left(289, 333\right)$

```
alpha2<-rbeta(10000,289,333)
```

Step 3: Obtain the posterior difference $p(\alpha_2 - \alpha_1 \mid y)$ by $p\left(\alpha_1 \mid y^{post}\right) - p\left(\alpha_2 \mid y^{pre}\right)$

```
diff<-alpha2-alpha1
```

Step 4: Find the mean of the difference greater than zero

```
mean(diff>0)
```

Result

```
[1] "posterior probability that there was a
shift toward Bush is 0.1982000000"
```