

APPM 4/5720: Computational Bayesian Statistics, Spring 2018

Solutions to Problem Set Three

1. Let V_y denote the variance-covariance matrix for \vec{Y} . Note that V_y is the upper left $k \times k$ block of V .

Take any $k \times 1$ vector $\vec{y} \neq \vec{0}$. We want to show that

$$\vec{y}^t V_y \vec{y} > 0.$$

Let \vec{x} be the $n \times 1$ vector whose first k elements are the elements of \vec{y} and whose last $n - k$ elements are filled out with zeros. Then Note that

$$\vec{y}^t V_y \vec{y} = \vec{x}^t V \vec{x} > 0$$

with the last inequality coming from the fact that V is positive definite.

2. “Noninformative priors” are not well defined. I will use flat (and therefore in this case improper) priors. You may have used Jeffreys priors or something else! In this solution, I will call the unrounded measurements y_1, y_2, \dots, y_5 instead of z_1, z_2, \dots, z_5 since the latter look more like standard normal random variables.

(a) The likelihood is

$$f(\vec{x}|\mu, \sigma^2) = (2\pi\sigma^2)^{-5/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^5 (x_i - \mu)^2 \right].$$

I will take the priors on μ and σ^2 to be

$$f(\mu) = 1, \quad -\infty < \mu < \infty$$

$$f(\sigma^2) = 1 \quad \sigma^2 > 0.$$

The posterior is then

$$\begin{aligned} f(\mu, \sigma^2|\vec{x}) &\propto f(\vec{x}|\mu, \sigma^2) \cdot f(\mu) \cdot f(\sigma^2) \\ &\propto (\sigma^2)^{-5/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^5 (x_i - \mu)^2 \right] \\ &= (\sigma^2)^{-5/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^5 x_i^2 \right] \cdot \exp \left[-\frac{1}{2\sigma^2} \left(5\mu^2 - 2\mu \sum_{i=1}^5 x_i \right) \right] \\ &= (\sigma^2)^{-5/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^5 x_i^2 \right] \cdot \exp \left[-\frac{5}{2\sigma^2} (\mu^2 - 2\mu\bar{x}) \right] \\ &= (\sigma^2)^{-5/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^5 x_i^2 \right] \cdot \exp \left[-\frac{5}{2\sigma^2} (\mu - \bar{x})^2 + \frac{5}{2\sigma^2} \bar{x}^2 \right] \\ &= (\sigma^2)^{-5/2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^5 x_i^2 + \frac{5}{2\sigma^2} \bar{x}^2 \right] \cdot \exp \left[-\frac{5}{2\sigma^2} (\mu - \bar{x})^2 + \right] \\ &= (\sigma^2)^{-5/2} \exp \left[-\frac{1}{2\sigma^2} 4s^2 \right] \cdot \exp \left[-\frac{5}{2\sigma^2} (\mu - \bar{x})^2 + \right] \end{aligned}$$

where s^2 is the sample variance

$$s^2 = \frac{\sum x_i^2 - (\sum x_i)^2/5}{5-1} = \frac{\sum x_i^2 - 5\bar{x}^2}{4}.$$

Which gives us the posteriors

$$\sigma^2|\vec{x} \sim IG(3/2, (n-1)s^2/2) = IG(3/2, 2s^2)$$

and

$$\mu|\sigma^2, \vec{x} \sim N(\bar{x}, \sigma^2/5).$$

Note that we have actual data and have that $\bar{x} = 10.4$ and $s^2 = 1.3$. So,

$$\sigma^2|\vec{x} \sim IG(1.5, 2.6)$$

and

$$\mu|\sigma^2, \vec{x} \sim N(10.4, \sigma^2/5).$$

(b) Now, the likelihood is

$$\begin{aligned} f(\vec{\mu}, \sigma^2) &= P(X_1 = x_1, X_2 = x_2, \dots, X_5 = x_5 | \mu, \sigma^2) \\ &\stackrel{indep}{=} \prod_{i=1}^5 P(X_i = x_i | \mu, \sigma^2) \\ &= \prod_{i=1}^5 P(x_i - 0.5 < Y_i < x_i + 0.5 | \mu, \sigma^2) \end{aligned}$$

where Y_i is the unrounded measurement corresponding to X_i .

Standardizing, we get

$$\begin{aligned} f(\vec{x}|\mu, \sigma^2) &= \prod_{i=1}^5 P\left(\frac{x_i - 0.5 - \mu}{\sigma} < \frac{Y_i - \mu}{\sigma} < \frac{x_i + 0.5 - \mu}{\sigma} \middle| \mu, \sigma^2\right) \\ &= \prod_{i=1}^5 P\left(\frac{x_i - 0.5 - \mu}{\sigma} < Z_i < \frac{x_i + 0.5 - \mu}{\sigma}\right) \\ &= \prod_{i=1}^5 \left[\Phi\left(\frac{x_i + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{x_i - 0.5 - \mu}{\sigma}\right) \right] \end{aligned}$$

Using the same flat priors we used in part (a), we have that the posterior distribution pdf for μ and σ^2 is

$$f(\mu, \sigma^2|\vec{x}) \propto \prod_{i=1}^5 \left[\Phi\left(\frac{x_i + 0.5 - \mu}{\sqrt{\sigma^2}}\right) - \Phi\left(\frac{x_i - 0.5 - \mu}{\sqrt{\sigma^2}}\right) \right]$$

for $-\infty < \mu < \infty$, and $\sigma^2 > 0$.

(c) I'll graph them in R. We can definitely see similarities! The truncated one looks like it is "leaning" and we had to go out a lot farther in σ^2 to see it tail off.

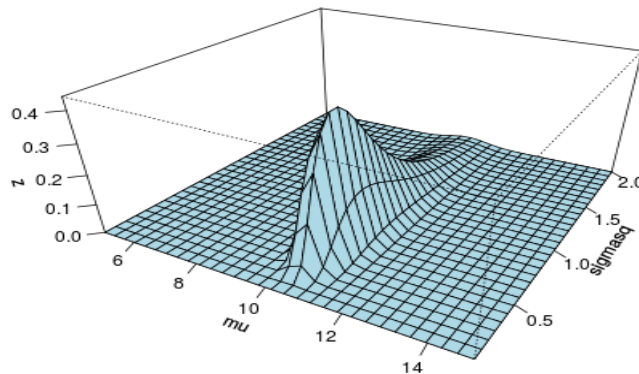
```
> mu <- seq(5, 15, length= 30)
> sigmasq <- seq(0.1, 2, length=30)
> alpha <- 1.5
> beta <- 2*1.3
> xbar <- 10.4
> f <- function(mu, sigmasq, alpha, beta, xbar) {
>   hold1 <- (1/gamma(alpha))*(beta^alpha)*(sigmasq^(-alpha-1))*
>     exp(-beta/sigmasq)
>   hold2 <- ((2*pi*sigmasq/5)^(-5/2))*exp(-1/(2*sigmasq/5))*(mu-xbar)^2
```

```

> hold1*hold2
> }
>
> z <- outer(mu, sigmasq, f,alpha,beta,xbar)
> z[is.na(z)] <- 1
> persp(mu, sigmasq, z, theta = 30, phi = 30, expand = 0.5,
> ticktype="detailed", col = "lightblue")

```

Figure 1: Posterior of 2a



```

> mu <- seq(5, 15, length= 30)
> sigmasq <-seq(0.1,2,length=30)
> x<-c(10,10,12,11,9)
> f<-function(mu,sigmasq,x){
> # Note: Using "prod" would have been slicker but
> # caused problems when combined with "outer"
> hold1<-pnorm((x[1]+0.5-mu)/sqrt(sigmasq))-pnorm((x[1]-0.5-mu)/
> sqrt(sigmasq))
> hold2<-pnorm((x[2]+0.5-mu)/sqrt(sigmasq))-pnorm((x[2]-0.5-mu)/
> sqrt(sigmasq))
> hold3<-pnorm((x[3]+0.5-mu)/sqrt(sigmasq))-pnorm((x[3]-0.5-mu)/
> sqrt(sigmasq))
> hold4<-pnorm((x[4]+0.5-mu)/sqrt(sigmasq))-pnorm((x[4]-0.5-mu)/
> sqrt(sigmasq))
> hold5<-pnorm((x[5]+0.5-mu)/sqrt(sigmasq))-pnorm((x[5]-0.5-mu)/
> sqrt(sigmasq))
> hold1*hold2*hold3*hold4*hold5
> }
> z <- outer(mu, sigmasq, f,x)
> z[is.na(z)] <- 1
> persp(mu, sigmasq, z, theta = 30, phi = 30, expand = 0.5,
> ticktype="detailed", col = "lightblue")

```

- (d) My clarifying parenthetical comment was anything but clarifying. What I was trying to ask (and what I think Gelman was trying to ask) is that we draw values from $f(\vec{y}|\vec{x})$.

Figure 2: Posterior of 2b

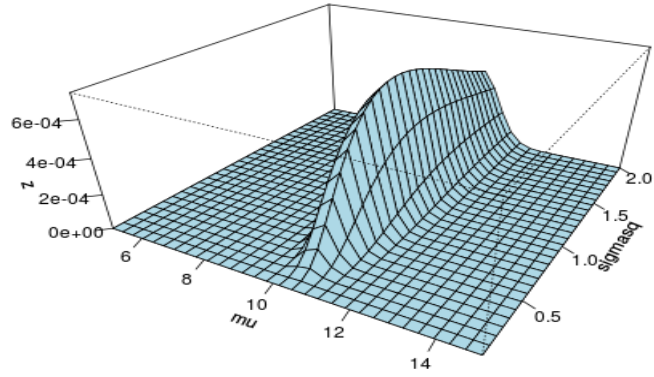
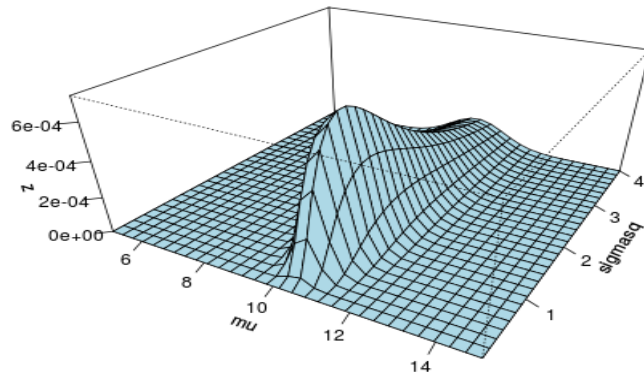


Figure 3: Posterior of 2b: Extended σ^2 Axis



Let's discuss this first in one dimension. That, is let us try to draw values y_1 from $f(y_1|x_1)$. For simplicity, I will refer to x_1 and y_1 as x and y . We have $x = 10$ and y is the unrounded version of x . Therefore, y is a normal random variable that takes values between 9.5 and 10.5.

Note that

$$f(y|x) = \int \int f(y, \mu, \sigma^2|x) d\mu d\sigma^2.$$

Since I don't want to integrate μ and σ^2 out, I will draw y 's from $f(y, \mu, \sigma^2|x)$ and then get draws from the marginal distribution in y by ignoring the output for μ and σ^2 .

Note that

$$f(y, \mu, \sigma^2|x) = f(y|\mu, \sigma^2, x) \cdot f(\mu, \sigma^2|x).$$

To draw from $f(y, \mu, \sigma^2|x)$, we must

- First draw (μ, σ^2) from $f(\mu, \sigma^2|x)$.
- Then draw y from $f(y|\mu, \sigma^2, x)$.

We're going to have to pick a prior here. I'm going to use the one from part (a). That is, I'm going to simulate σ^2 from the $IG(1.5, 2.6)$ and then simulate μ from the $N(10.4, \sigma^2/5)$.

In R, we can draw from this normal inverse gamma distribution as follows. (We will draw 100,000 pairs all at once.)

```
> n<-100000
> alpha<-1.5
> beta<-2.6
> mumean<-10.4
> sigmasq<-1/rgamma(n,alpha,beta)
> mu<-sqrt(sigmasq/5)*rnorm(n,0,1)+mumean
```

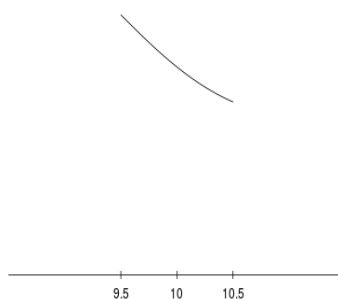
For each (μ, σ^2) pair, we must draw a y from the $N(\mu, \sigma^2)$ distribution, restricted to the domain $(x - 0.5, x + 0.5) = (9.5, 10.5)$. The pdf is

$$f(y|\mu, \sigma^2, x) \propto \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right] I_{(9.5, 10.5)}(y). \quad (1)$$

(The x came in in defining the distribution of μ and σ^2 and also in defining that indicator. Once they are determined, we are done with x .)

We will be drawing y from a little piece of the normal distribution which will look like this assuming $\mu < 9.5$. (If $\mu > 10.5$ it will instead be increasing, and if $9.5 < \mu < 10.5$, it will be like the top of a normal curve with a maximum at μ .)

Figure 4: One Example of $f(y|\mu, \sigma^2, 10)$ (Assuming $\mu < 9.5$.)



Properly normalized so that it integrates to 1, (1) becomes

$$f(y|\mu, \sigma^2, x) = \left[\Phi\left(\frac{10.5 - \mu}{\sigma}\right) - \Phi\left(\frac{9.5 - \mu}{\sigma}\right) \right]^{-1} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right] I_{(9.5, 10.5)}(y). \quad (2)$$

We can simulate values from (2) using the accept/reject algorithm with proposal density being uniform over (9.5, 10.5). That is, we will take g (notation from the accept/reject algorithm) to be the flat line at the high point:

$$g(y) = \left[\Phi\left(\frac{10.5 - \mu}{\sigma}\right) - \Phi\left(\frac{9.5 - \mu}{\sigma}\right) \right]^{-1} \exp\left[-\frac{1}{2\sigma^2}(9.5 - \mu)^2\right] I_{(9.5, 10.5)}(y).$$

Then

$$\begin{aligned} c &= \int_{9.5}^{10.5} g(y) dy = \left[\Phi\left(\frac{10.5 - \mu}{\sigma}\right) - \Phi\left(\frac{9.5 - \mu}{\sigma}\right) \right]^{-1} \exp\left[-\frac{1}{2\sigma^2}(9.5 - \mu)^2\right] \cdot (10.5 - 9.5) \\ &= \int_{9.5}^{10.5} g(y) dy = \left[\Phi\left(\frac{10.5 - \mu}{\sigma}\right) - \Phi\left(\frac{9.5 - \mu}{\sigma}\right) \right]^{-1} \exp\left[-\frac{1}{2\sigma^2}(9.5 - \mu)^2\right]. \end{aligned}$$

The normalized version of g is $h(y) = I_{(9.5, 10.5)}(y)$, which is the pdf for the uniform distribution over (9.5, 10.5).

The accept/reject algorithm says

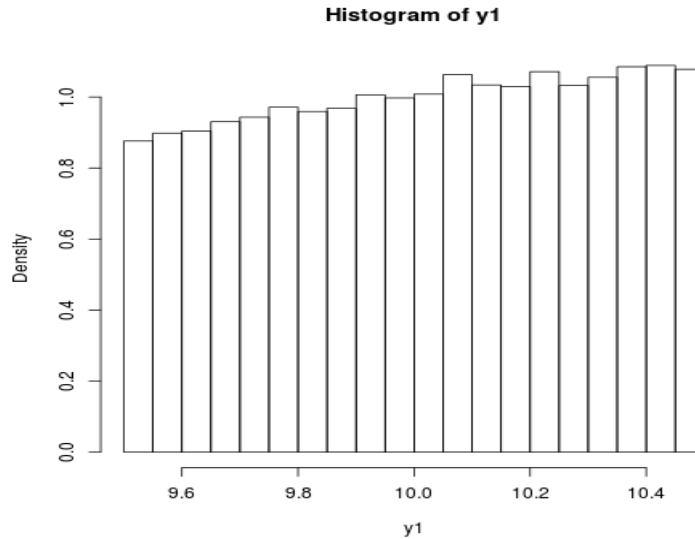
- Draw $W \sim \text{unif}(9.5, 10.5)$.
- Draw $U \sim \text{unif}(0, 1)$.
- If $U \leq \frac{f(W)}{ch(W)} = \frac{f(W)}{g(W)}$, accept W as a draw from $f(y|\mu, \sigma^2, 10)$. Set $Y = W$. Otherwise, reject and return to first bullet.

Full code for drawing from $f(y|\mu, \sigma^2, x)$ taking into account the different shapes we can get for it depending on μ is given below in R.

```
>y1<-rep(0,n)
>for(i in 1:n){
> accept<-F
> while(accept == F){
>   w<-runif(1)+9.5
>   u<-runif(1)
>
>   if(mu[i]< 9.5){
>     ratio<-exp(-(1/(2*sigmasq[i]))*((w-mu[i])^2-(9.5-mu[i])^2))
>   } else if (mu[i] > 10.5){
>     ratio<-exp(-(1/(2*sigmasq[i]))*((w-mu[i])^2-(10.5-mu[i])^2))
>   } else{
>     ratio<-exp(-(1/(2*sigmasq[i]))*((w-mu[i])^2))
>   }
>
>   if(u<=ratio){
>     y1[i]<-w
>     accept<-T
>   }
> }
> }
```

A histogram of $y = y_1$ is shown below. The sample mean is 10.02, which is close to x_1 . However, the posterior mean of 10.4 is pulling values of y_1 up more towards 10.4!

Figure 5: 100,000 Draws from $f(y_1|x_1)$



3. The target pdf is

$$f(x) = 12x^2(1-x)I_{(0,1)}(x).$$

It is easy to check that the maximum occurs at $x = 2/3$ and that the maximum is $16/9$.

(a) Here, we can take $g(x) = \frac{16}{9}I_{(0,1)}(c)$. Then $c = 16/9$ and $h(x) = I_{(0,1)}(x)$.

The f/g ratio for the accept reject algorithm will be $12x^2(1-x)/(16/9) = \frac{27}{4}x^2(1-x)$

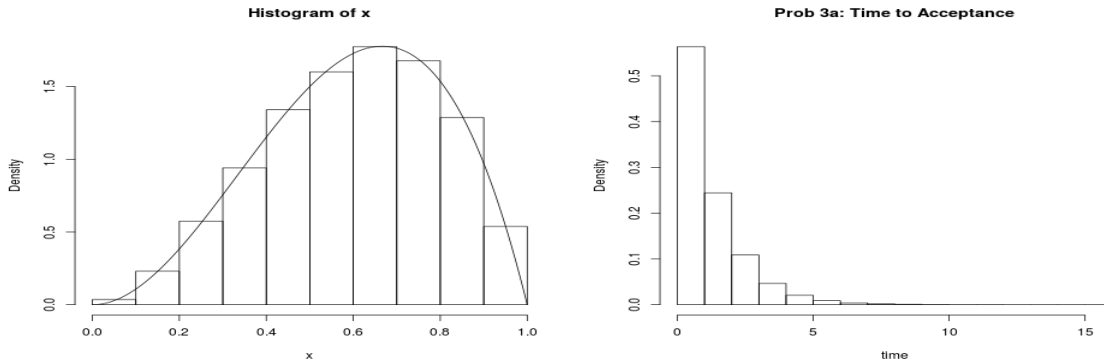
Here is the R-Code:

```
> n<-100000
> x<-rep(0,n)
> for(i in 1:n){
+   accept<-F
+   while(accept == F){
+     y<-runif(1)
+     u<-runif(1)
+     ratio<-(27/4)*(y^2)*(1-y)
+     if(u <= ratio){
+       x[i]<-y
+       accept <-T
+     }
+   }
+ }
```

Here is code for the histogram which is pictured below. The mean time to acceptance was 1.78 with a maximum time of 16. A histogram of time to acceptance is also given below.

```
> br<-seq(0,1,0.1)
> hist(x,prob=T,breaks=br)
> y<-seq(0,1,0.001)
> f<-12*(y^2)*(1-y)
> lines(y,f)
```

Figure 6: Problem 3a

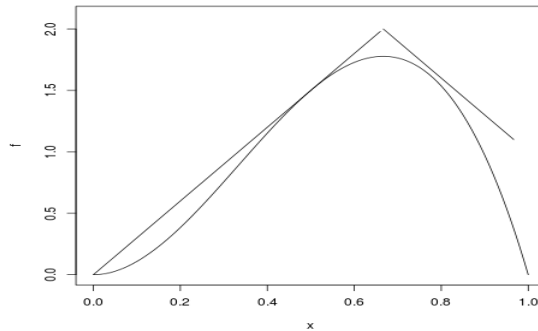


(b) One (non-optimal) piecewise linear bounding function is given by

$$g(x) = \begin{cases} 3x & , 0 \leq x \leq 2/3 \\ -3x + 4 & , 2/3 \leq x \leq 1. \end{cases}$$

This is depicted below.

Figure 7: Problem 3a



Since g integrates to $7/6$, we can take

$$h(x) = \frac{6}{7}g(x) = \begin{cases} \frac{18}{7}x & , 0 \leq x \leq 2/3 \\ -\frac{18}{7}x + \frac{24}{7} & , 2/3 \leq x \leq 1. \end{cases}$$

The associated cdf is

$$H(x) = \begin{cases} \frac{9}{7}x^2 & , 0 \leq x \leq 2/3 \\ \frac{4}{7} - \frac{9}{7}x^2 + \frac{24}{7}x - \frac{12}{7} & , 2/3 \leq x \leq 1 \end{cases} = \begin{cases} \frac{9}{7}x^2 & , 0 \leq x \leq 2/3 \\ -\frac{9}{7}x^2 + \frac{24}{7}x - \frac{8}{7} & , 2/3 \leq x \leq 1 \end{cases}$$

This can be inverted in pieces to give so that we may draw from the distribution with cdf H as follows. We draw $U \sim \text{unif}(0, 1)$.

- If $U < \frac{9}{7} \left(\frac{2}{3}\right)^2 = \frac{4}{7}$, plug the uniform into $x = \sqrt{\frac{7}{9}U}$.
- Otherwise, plug U into $x = \frac{-24 + \sqrt{24^2 - 36(8 + 7U)}}{-18}$.

The acceptance ratio for the accept-reject algorithm is

$$f(x)/g(x) = \begin{cases} 4x(1-x) & , 0 \leq x \leq 2/3 \\ \frac{12x^2(1-x)}{-3x+4} & , 2/3 \leq x \leq 1 \end{cases}$$

Overall, the R code is as follows. (This time I have included keeping track of the time to acceptance as well.)


```

n<-100000
x<-rep(0,n)
time<-rep(0,n)
for(i in 1:n){
  accept<-F
  hold<-0
  while(accept == F){
    hold<-hold+1
    w<-runif(1)
    if(w <= 2/3){
      y<-sqrt(7*w/9)
    } else{
      y<- (-24+sqrt(24^2-36*(8+7*w)))/(-18)
    }
    u<-runif(1)
    if(y <=2/3){
      ratio<-4*y*(1-y)
    } else{
      ratio<-12*(y^2)*(1-y)/(-3*y+4)
    }
    if(u <= ratio){
      x[i]<-y
      time[i]<-hold
      accept <-T
    }
  }
}

```

The plots are uninteresting since the results look exactly the same and the histogram for time is boring either way. However, the mean time dropped from 1.78 to 1.16 and the max dropped from 16 to 6! Yay!

4. First we will show that marginals from the Dirichlet distribution are again Dirichlet. I won't bother using the posterior parameters yet.

Suppose that $(\theta_1, \theta_2, \dots, \theta_k) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \dots, \alpha_k)$. Then the joint pdf is

$$\begin{aligned}
 f(\theta_1, \theta_2, \dots, \theta_k) &= \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \cdot \theta_2^{\alpha_2-1} \dots \theta_k^{\alpha_k-1} \\
 &= \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \cdot \theta_2^{\alpha_2-1} \dots \theta_{k-1}^{\alpha_{k-1}-1} (1 - \theta_1 - \dots - \theta_{k-1})^{\alpha_k-1}
 \end{aligned}$$

Really, we can think of this as a joint pdf for $(\theta_1, \theta_2, \dots, \theta_{k-1})$ since θ_k is “locked in”. So, we will try to marginalize out θ_{k-1} :

$$\begin{aligned}
 &\frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \int \theta_1^{\alpha_1-1} \cdot \theta_2^{\alpha_2-1} \dots \theta_{k-1}^{\alpha_{k-1}-1} (1 - \theta_1 - \dots - \theta_{k-1})^{\alpha_k-1} d\theta_{k-1} \\
 &\frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \cdot \theta_2^{\alpha_2-1} \dots \theta_{k-2}^{\alpha_{k-2}-1} \int \theta_{k-1}^{\alpha_{k-1}-1} (1 - \theta_1 - \dots - \theta_{k-1})^{\alpha_k-1} d\theta_{k-1}
 \end{aligned}$$

The integrals go from 0 to $1 - \theta_1 - \dots - \theta_{k-2}$. (It looked horrible written in there!)

Let $c = \theta_1 + \theta_2 + \dots + \theta_{k-2}$. Then we want to integrate

$$\int \theta_{k-1}^{\alpha_{k-1}-1} (1 - c - \theta_{k-1})^{\alpha_k-1} d\theta_{k-1}$$

Consider the substitution $\theta_{k-1} = (1-c)u$. Note that u goes from 0 to 1. Also, $d\theta_{k-1} = (1-c)du$. The integral becomes

$$\begin{aligned} & \int_0^1 [(1-c)u]^{\alpha_{k-1}-1} [1-c-(1-c)u]^{\alpha_k-1} (1-c) du \\ &= (1-c)^{\alpha_{k-1}+\alpha_k-1} \int_0^1 u^{\alpha_{k-1}-1} (1-u)^{\alpha_k-1} du \\ & \quad (1-c)^{\alpha_{k-1}+\alpha_k-1} \mathcal{B}(\alpha_{k-1}, \alpha_k) \\ & \quad (1-c)^{\alpha_{k-1}+\alpha_k-1} \frac{\Gamma(\alpha_{k-1})\Gamma(\alpha_k)}{\Gamma(\alpha_{k-1}+\alpha_k)}. \end{aligned}$$

Putting it all together, the marginal for $(\theta_1, \theta_2, \dots, \theta_{k-2})$ is

$$\begin{aligned} f(\theta_1, \theta_2, \dots, \theta_{k-2}) &= \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \theta_1^{\alpha_1-1} \cdot \theta_2^{\alpha_2-1} \dots \theta_{k-2}^{\alpha_{k-2}-1} \cdot (1-\theta_1-\dots-\theta_{k-2})^{\alpha_{k-1}+\alpha_k-1} \frac{\Gamma(\alpha_{k-1})\Gamma(\alpha_k)}{\Gamma(\alpha_{k-1}+\alpha_k)} \\ &= \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^{k-2} \Gamma(\alpha_i)\Gamma(\alpha_{k-1}+\alpha_k)} \theta_1^{\alpha_1-1} \cdot \theta_2^{\alpha_2-1} \dots \theta_{k-2}^{\alpha_{k-2}-1} \cdot (1-\theta_1-\dots-\theta_{k-2})^{\alpha_{k-1}+\alpha_k-1} \end{aligned}$$

So, we see that (recall that θ_{k-1} is in there but locked in and represented as $1-\theta_1-\dots-\theta_{k-2}$)

$$(\theta_1, \theta_2, \dots, \theta_{k-1}) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \dots, \alpha_{k-1}, \alpha_{k-1} + \alpha_k).$$

Continuing down (formal induction argument anyone?), we get that

$$(\theta_1, \theta_2) \sim \text{Dirichlet}(\alpha_1, \sum_{i=2}^k \alpha_i).$$

Going back to “posterior land”, (I just worked so far with a generic Dirichlet and not the posterior Dirichlet) we have that

$$(\theta_1, \theta_2) \sim \text{Dirichlet}(\alpha_1 + x_1, \sum_{i=2}^k (\alpha_i + x_i)).$$

However, since $\theta_2 = 1 - \theta_1$, we have that $\theta = \theta_1 \sim \text{Beta}(\alpha_1 + x_1, \sum_{i=2}^k (\alpha_i + x_i))$.

This is not going to give us the desired result in part (b). So, i think Gelman meant for us to really find that

$$(\theta_1, \theta_2, \theta_3) \sim \text{Dirichlet}(\alpha_1 + x_1, \alpha_2 + x_2, \sum_{i=3}^k (\alpha_i + x_i)),$$

in which case we can find the marginal posterior for $\theta = \theta_1/(\theta_1 + \theta_2)$ but using the Jacobian method to find the joint distribution of this and the denominator. You can then marginalize out the denominator and should find that

$$\theta|\vec{x} \sim \text{Beta}(\alpha_1 + x_1, \alpha_2 + x_2).$$

Note that we can get this by multiplying thr suggested binomial distribution and a $\text{Beta}(\alpha_1, \alpha_2)$ prior.