Answers to selected exercises

[]

Chapter 1

3 As hinted in the problem statement, level of urbanicity might well explain the poverty pattern evident in Figure 1.2. Other regional spatially oriented covariates to consider might include percent of minority residents, percent with high school diploma, unemployment rate, and average age of the housing stock. The point here is that spatial patterns can often be explained by patterns in existing covariate data. Accounting for such covariates in a statistical model may result in residuals that show little or no spatial pattern, thus obviating the need for formal spatial modeling.

8(a) The length of a position vector with coordinates $\mathbf{u} = (u_1, u_2)$ is the length of the hypotenuse of the right-angled triangle formed with the origin $(0,0)$, the point $\mathbf{u}$ and the perpendicular dropped from $\mathbf{u}$ to the x-axis.

(b) These are straightforward verifications. For example,

$$(\alpha\mathbf{u})\cdot\mathbf{v} = \sum_{i=1}^{d}(\alpha u_i)v_i = \sum_{i=1}^{d}\alpha u_i v_i = \alpha\sum_{i=1}^{d}u_i v_i = \alpha(\mathbf{u}\cdot\mathbf{v}) \ .$$

(c) Again straightforward verifications. For example,

$$\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \sum_{i=1}^{d}u_i(v_i+w_i) = \sum_{i=1}^{d}u_i v_i + \sum_{i=1}^{d}u_i w_i = \mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w} \ .$$

(d) Noting that $\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})$, we obtain the Pythagorean identity

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

when $\mathbf{u}\cdot\mathbf{v} = 0$. Similarly, $\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

(e) Define the residual $\mathbf{w} := \mathbf{u} - \beta\mathbf{v}$, where $\beta := (\mathbf{u}\cdot\mathbf{v})/\|\mathbf{v}\|^2$. Then,

$$\mathbf{w}\cdot\mathbf{v} = (\mathbf{u}-\beta\mathbf{v})\cdot\mathbf{v} = \mathbf{u}\cdot\mathbf{v} - \beta\|\mathbf{v}\|^2 = \mathbf{u}\cdot\mathbf{v} - \frac{(\mathbf{u}\cdot\mathbf{v})}{\|\mathbf{v}\|^2}\|\mathbf{v}\|^2 = \mathbf{u}\cdot\mathbf{v} - \mathbf{u}\cdot\mathbf{v} = 0 \ .$$

Therefore, $\mathbf{w}\cdot(\beta\mathbf{v}) = \beta(\mathbf{w}\cdot\mathbf{v}) = 0$ and the Pythagorean identity applies to $\|\mathbf{u}\|^2 = \|\mathbf{w}+\beta\mathbf{v}\|^2 = \|\mathbf{w}\|^2 + \beta^2\|\mathbf{v}\|^2$. Therefore,

$$\|\mathbf{u}\|^2 = \|\mathbf{w}\|^2 + \beta^2\|\mathbf{v}\|^2 = \|\mathbf{w}\|^2 + \frac{(\mathbf{u}\cdot\mathbf{v})^2}{\|\mathbf{v}\|^4}\|\mathbf{v}\|^2 = \|\mathbf{w}\|^2 + \frac{(\mathbf{u}\cdot\mathbf{v})^2}{\|\mathbf{v}\|^2} \geq \frac{(\mathbf{u}\cdot\mathbf{v})^2}{\|\mathbf{v}\|^2} \ .$$

Multiplying both sides by $\|\mathbf{v}\|^2$ and taking the square root gives the desired inequality.

(f) Set up a triangle with two position vectors **u** and **v** and the third side is the vector $\mathbf{u} - \mathbf{v}$. The angle between **u** and **v** is θ (opposite side $\mathbf{u} - \mathbf{v}$) and, hence,

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2(\mathbf{u} \cdot \mathbf{v}) \ .$$

The cosine law of a triangle with sides of length a, b and c and angle C is the angle between made by sides a and b (angle opposite the side c) is $c^2 = a^2 + b^2 - 2ab\cos C$. This corresponds to $a := \|\mathbf{u}\|$, $b := \|\mathbf{v}\|$, $c := \|\mathbf{u} - \mathbf{v}\|$ and $\theta := C$. Therefore, the cosine law implies that $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta$, which agrees with the definition of $\cos\theta$ using dot products.

(g) If $\mathbf{u} \cdot \mathbf{v} = 0$, then $\cos\theta = 0$, which means that the angle between the two vectors is $\theta = \pi/2$. Hence, this agrees the geometric notion of orthogonality.

10(a) The appropriate R code is as follows:

```
#  R program to compute geodesic distance
#  see also www.auslig.gov.au/geodesy/datums/distance.htm

#  input:   point1=(long,lat) and point2=(long,lat)
              in degrees
#  output:  distance in km between the two points
#  example:
point1 <- c(87.65,41.90)  # Chicago (downtown)
point2 <- c(87.90,41.98)  # Chicago (O'Hare airport)
point3 <- c(93.22,44.88)  # Minneapolis (airport)
#  geodesic(point1,point3) returns 558.6867

geodesic <- function(point1, point2){
  R <- 6371
    point1 <- point1 * pi/180
    point2 <- point2 * pi/180
      d <- sin(point1[2]) * sin(point2[2]) +
              cos(point1[2]) * cos(point2[2]) *
              cos(abs(point1[1] - point2[1]))
    R*acos(d)
}
```

(b) Chicago to Minneapolis, 562 km; New York to New Orleans, 1897.2 km.

11 Chicago to Minneapolis, 706 km; New York to New Orleans, 2172.4 km. This overestimation is expected since the approach stretches the meridians and parallels, or equivalently, presses the curved domain onto a plane, thereby stretching the domain (and hence the distances). As the geodesic distance increases, the quality of the naive estimates deteriorates.

12 Chicago to Minneapolis, 561.8 km; New York to New Orleans, 1890.2 km. Here, the slight underestimation is expected, since it finds the straight line by penetrating (burrowing through) the spatial domain. Still, this approximation seems quite good even for distances close to 2000 km (e.g., New York to New Orleans).

13(a) Chicago to Minneapolis, 562.2 km; New York to New Orleans, 1901.5 km.

(b) Whenever all of the points are located along a parallel or a meridian, this projection will not be defined.

15(a) These are straightforward verifications using the definition in (1.22). For example,

$$\mathbf{u}\times\mathbf{u}=\mathbf{U}_\times\mathbf{u}=\begin{bmatrix}0&-u_3&u_2\\u_3&0&-u_1\\-u_2&u_1&0\end{bmatrix}\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}=\begin{bmatrix}u_2u_3-u_3u_2\\u_3u_1-u_1u_3\\u_1u_2-u_2u_1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Similarly, $\mathbf{v}\times\mathbf{u}$ is calculated by interchanging the u_i's and v_i's in (1.22), which reveals

$$\mathbf{v}\times\mathbf{u}=\begin{bmatrix}v_2u_3-v_3u_2\\v_3u_1-v_1u_3\\v_1u_2-v_2u_1\end{bmatrix}=\begin{bmatrix}-(v_3u_2-v_2u_3)\\-(v_1u_3-v_3u_1)\\-(v_2u_1-v_1u_2)\end{bmatrix}=\begin{bmatrix}-(u_2v_3-u_3v_2)\\-(u_3v_1-u_1v_3)\\-(u_1v_2-u_2v_1)\end{bmatrix}=-(\mathbf{u}\times\mathbf{v})$$

Finally, for multiplication by a scalar c,

$$(c\mathbf{u})\times\mathbf{v}=(c\mathbf{U}_\times)\mathbf{v}=\underbrace{\mathbf{U}_\times(c\mathbf{v})}_{\mathbf{u}\times(c\mathbf{v})}=\underbrace{c(\mathbf{U}_\times\mathbf{v})}_{c(\mathbf{u}\times\mathbf{v})}.$$

(b) The distributive laws are a consequence of the linear transformation in (1.22):

$$\mathbf{u}\times(\mathbf{v}+\mathbf{w})=\mathbf{U}_\times(\mathbf{v}+\mathbf{w})=\mathbf{U}_\times\mathbf{v}+\mathbf{U}_\times\mathbf{w}=\mathbf{u}\times\mathbf{v}+\mathbf{u}\times\mathbf{w}$$

For $(\mathbf{v}+\mathbf{w})\times\mathbf{u}$, we can simply use one of the results in (a):

$$(\mathbf{v}+\mathbf{w})\times\mathbf{u}=-(\mathbf{u}\times(\mathbf{v}+\mathbf{w}))=-(\mathbf{u}\times\mathbf{v}+\mathbf{u}\times\mathbf{w})=-(-(\mathbf{v}\times\mathbf{u})-(\mathbf{w}\times\mathbf{u}))=\mathbf{v}\times\mathbf{u}+\mathbf{w}\times\mathbf{u}.$$

(c) Some linear algebra produces the desired result:

$$\|\mathbf{u}\times\mathbf{v}\|^2=\|\mathbf{U}_\times\mathbf{v}\|^2=\mathbf{v}^{\mathrm{T}}\mathbf{U}_\times^{\mathrm{T}}(\mathbf{U}_\times\mathbf{v})=\mathbf{v}^{\mathrm{T}}\underbrace{\begin{bmatrix}0&u_3&-u_2\\-u_3&0&u_1\\u_2&-u_1&0\end{bmatrix}}_{\mathbf{U}_\times^{\mathrm{T}}}\underbrace{\begin{bmatrix}u_2v_3-u_3v_2\\u_3v_1-u_1v_3\\u_1v_2-u_2v_1\end{bmatrix}}_{\mathbf{U}_\times\mathbf{v}}$$

$$=\mathbf{v}^{\mathrm{T}}\begin{bmatrix}u_3(u_3v_1-u_1v_3)-u_2(u_1v_2-u_2v_1)\\u_1(u_1v_2-u_2v_1)-u_3(u_2v_3-u_3v_2)\\u_2(u_2v_3-u_3v_2)-u_1(u_3v_1-u_1v_3)\end{bmatrix}=\mathbf{v}^{\mathrm{T}}\begin{bmatrix}v_1(u_2^2+u_3^2)-u_1(u_2v_2+u_3v_3)\\v_2(u_3^2+u_1^2)-u_2(u_3v_3+u_1v_1)\\v_3(u_1^2+u_2^2)-u_3(u_1v_1+u_2v_2)\end{bmatrix}$$

$$=\mathbf{v}^{\mathrm{T}}\begin{bmatrix}v_1(\|\mathbf{u}\|^2-u_1^2)-u_1(u_2v_2+u_3v_3)\\v_2(\|\mathbf{u}\|^2-u_2^2)-u_2(u_3v_3+u_1v_1)\\v_3(\|\mathbf{u}\|^2-u_3^2)-u_3(u_1v_1+u_2v_2)\end{bmatrix}=\mathbf{v}^{\mathrm{T}}\begin{bmatrix}v_1\|\mathbf{u}\|^2-u_1(u_1v_1+u_2v_2+u_3v_3)\\v_2\|\mathbf{u}\|^2-u_2(u_2v_2+u_3v_3+u_1v_1)\\v_3\|\mathbf{u}\|^2-u_3(u_3v_3+u_1v_1+u_2v_2)\end{bmatrix}$$

$$=\mathbf{v}^{\mathrm{T}}\begin{bmatrix}v_1\|\mathbf{u}\|^2-u_1(\mathbf{u}\cdot\mathbf{v})\\v_2\|\mathbf{u}\|^2-u_2(\mathbf{u}\cdot\mathbf{v})\\v_3\|\mathbf{u}\|^2-u_3(\mathbf{u}\cdot\mathbf{v})\end{bmatrix}=\mathbf{v}^{\mathrm{T}}\left(\|\mathbf{u}\|^2\mathbf{v}-(\mathbf{u}\cdot\mathbf{v})\mathbf{u}\right)=\|\mathbf{u}\|^2\underbrace{\mathbf{v}^{\mathrm{T}}\mathbf{v}}_{\|\mathbf{v}\|^2}-(\mathbf{u}\cdot\mathbf{v})\underbrace{(\mathbf{v}^{\mathrm{T}}\mathbf{u})}_{(\mathbf{u}\cdot\mathbf{v})}$$

$$=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})^2=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-\|\mathbf{u}\|^2\|\mathbf{v}\|^2\cos^2\theta=\|\mathbf{u}\|^2\|\mathbf{v}\|^2\sin^2\theta.$$

(d) Observe that $\mathbf{U}_\times^{\mathrm{T}}\mathbf{u}=\mathbf{0}$ (easy direct calculation), which shows that

$$\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v})=\mathbf{u}^{\mathrm{T}}\mathbf{U}_\times\mathbf{v}=\underbrace{\left(\mathbf{U}_\times^{\mathrm{T}}\mathbf{u}\right)^{\mathrm{T}}}_{\mathbf{0}^{\mathrm{T}}}\mathbf{v}=\mathbf{0}^{\mathrm{T}}\mathbf{v}=0.$$

To evaluate $\mathbf{v}\cdot(\mathbf{u}\times\mathbf{v})$, we note from one of the properties in part (a) that $\mathbf{u}\times\mathbf{v}=-(\mathbf{v}\times\mathbf{u})$. Hence, $\mathbf{v}\cdot(\mathbf{u}\times\mathbf{v})=-\mathbf{v}\cdot(\mathbf{v}\times\mathbf{u})$, which is equal to 0 because of exactly how we showed $\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v})=0$.

Another easy way to show that $\mathbf{u}\times\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$ is to use the result in (1.25) and use the property of the determinant. Thus, $\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v})=\det(\begin{bmatrix}\mathbf{u}&\mathbf{u}&\mathbf{v}\end{bmatrix})=0$ because the determinant has 2 identical columns. Similarly, $\mathbf{v}\cdot(\mathbf{u}\times\mathbf{v})=\det(\begin{bmatrix}\mathbf{v}&\mathbf{u}&\mathbf{v}\end{bmatrix})=0$.

(e) Draw a parallelogram with the vectors $\mathbf{u}$ and $\mathbf{v}$ as sides. The altitude from point $\mathbf{v}$ to base $\mathbf{u}$ is $\|\mathbf{v}\| \sin\theta$. Therefore, the area is $\|\mathbf{u}\|\|\mathbf{v}\| \sin\theta = \|\mathbf{u} \times \mathbf{v}\|$.

(f) One can verify the following result using matrix multiplication and some straightforward (but a bit tedious) algebra:

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = \mathbf{u} \times (\mathbf{V}_\times \mathbf{w}) = \mathbf{U}_\times (\mathbf{V}_\times \mathbf{w}) = \mathbf{U}_\times \mathbf{V}_\times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w} \; .$$

But a more conceptually appealing way of deriving the above identity exploits some geometric intuition. The vector $\mathbf{v} \times \mathbf{w}$ is orthogonal to both $\mathbf{v}$ and $\mathbf{w}$, hence it is the normal vector to the plane spanned by $\mathbf{v}$ and $\mathbf{w}$. Therefore, the vector $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ is normal to the plane formed by $\mathbf{u}$ and the normal vector to the plane containing $\mathbf{v}$ and $\mathbf{w}$. This maps $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ back to the plane containing $\mathbf{v}$ and $\mathbf{w}$. This is perhaps best visualized by a diagram, which we leave the reader to try out. We can, then, start with an educated guess that

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = \alpha\mathbf{v} + \beta\mathbf{w} \tag{17.42}$$

and attempt to find the coefficients α and β in terms of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$.

Calculations are simplified using a convenient 3-dimensional frame of unit-length orthogonal vectors $\mathbf{I}$, $\mathbf{J}$ and $\mathbf{K}$ with $\mathbf{K} = \mathbf{I} \times \mathbf{J}$, $\mathbf{J} \times \mathbf{K} = \mathbf{I}$ and $\mathbf{I} \times \mathbf{K} = -\mathbf{J}$ so that:

$$\mathbf{v} = v_1\mathbf{I} \; ; \quad \mathbf{w} = w_1\mathbf{I} + w_2\mathbf{J} \quad \text{and} \quad \mathbf{u} = u_1\mathbf{I} + u_2\mathbf{J} + u_3\mathbf{K} \; . \tag{17.43}$$

It is easy to compute $\mathbf{v} \times \mathbf{w}$ in this coordinate frame as

$$\mathbf{v} \times \mathbf{w} = (v_1\mathbf{I}) \times (w_1\mathbf{I} + w_2\mathbf{J}) = v_1 w_1 \underbrace{(\mathbf{I} \times \mathbf{I})}_{\mathbf{0}} + v_1 w_2 \underbrace{(\mathbf{I} \times \mathbf{J})}_{\mathbf{K}} = v_1 w_2 \mathbf{K} \; .$$

Now we proceed to evaluate $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$. The idea will be to first express this vector in terms of $\mathbf{I}$, $\mathbf{J}$ and $\mathbf{K}$ and then do some reverse calculations to try to express this vector as a linear combination of $\mathbf{v}$ and $\mathbf{w}$. Here we go:

$$\begin{aligned}\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) &= (u_1\mathbf{I} + u_2\mathbf{J} + u_3\mathbf{K}) \times (\mathbf{v} \times \mathbf{w}) = (u_1\mathbf{I} + u_2\mathbf{J} + u_3\mathbf{K}) \times (v_1 w_2 \mathbf{K}) \\ &= u_1 v_1 w_2 \underbrace{(\mathbf{I} \times \mathbf{K})}_{-\mathbf{J}} + u_2 v_1 w_2 \underbrace{(\mathbf{J} \times \mathbf{K})}_{\mathbf{I}} + u_3 v_1 w_2 \underbrace{(\mathbf{K} \times \mathbf{K})}_{\mathbf{0}} \\ &= -u_1 v_1 w_2 \mathbf{J} + u_2 v_1 w_2 \mathbf{I} \; .\end{aligned} \tag{17.44}$$

We now express $\mathbf{J}$ and $\mathbf{I}$ in terms of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ using (17.43):

$$\mathbf{I} = (1/v_1)\mathbf{v} \quad \text{and} \quad w_2\mathbf{J} = \mathbf{w} - w_1\mathbf{I} = \mathbf{w} - (w_1/v_1)\mathbf{v} \; . \tag{17.45}$$

Proceeding with (17.44) and substituting the expressions in (17.45), we obtain

$$\begin{aligned}\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) &= -u_1 v_1 w_2 \mathbf{J} + u_2 v_1 w_2 \mathbf{I} = -u_1 v_1 (w_2\mathbf{J}) + u_2 w_2 (v_1\mathbf{I}) \\ &= -u_1 v_1 (\mathbf{w} - (w_1/v_1)\mathbf{v}) + u_2 w_2 \mathbf{v} = \underbrace{(u_2 w_2 + u_1 w_1)}_{\alpha} \mathbf{v} + \underbrace{(-u_1 v_1)}_{\beta} \mathbf{w} \; .\end{aligned}$$

This proves (17.42)with $\alpha = u_2 w_2 + u_1 w_1$ and $\beta = -u_1 v_1$.

It remains to verify that $\alpha = \mathbf{u} \cdot \mathbf{w}$ and $\beta = -\mathbf{u} \cdot \mathbf{v}$. Since $\mathbf{I}$, $\mathbf{J}$ and $\mathbf{K}$ are orthonormal,

$$\begin{aligned}\mathbf{u} \cdot \mathbf{w} &= (u_1\mathbf{I} + u_2\mathbf{J} + u_3\mathbf{K}) \cdot (w_1\mathbf{I} + w_2\mathbf{J}) = u_1 w_1 (\mathbf{I} \cdot \mathbf{I}) + u_2 w_2 (\mathbf{J} \cdot \mathbf{J}) = u_1 w_1 + u_2 w_2 = \alpha \; ; \\ \mathbf{u} \cdot \mathbf{v} &= (u_1\mathbf{I} + u_2\mathbf{J} + u_3\mathbf{K}) \cdot (v_1\mathbf{I}) = u_1 v_1 = -\beta \; .\end{aligned}$$

Therefore, $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = \alpha\mathbf{v} + \beta\mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}$.

16(a) This follows from simple rearrangements and the definition of a 3×3 determinant:

$$\begin{aligned}\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) &= \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} v_2 w_3 - v_3 w_2 \\ v_3 w_1 - v_1 w_3 \\ v_1 w_2 - v_2 w_1 \end{bmatrix} \\ &= u_1 \det\left(\begin{bmatrix} v_2 & w_2 \\ v_3 & w_3 \end{bmatrix}\right) - u_2 \det\left(\begin{bmatrix} v_1 & w_1 \\ v_3 & w_3 \end{bmatrix}\right) + u_3 \det\left(\begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}\right) = \det\left(\begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{bmatrix}\right) .\end{aligned}$$

(b) Even permutations of columns do not alter the value of the determinant. Therefore,

$$\underbrace{\det\left(\begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{bmatrix}\right)}_{\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w})} = \underbrace{\det\left(\begin{bmatrix} \mathbf{v} & \mathbf{w} & \mathbf{u} \end{bmatrix}\right)}_{\mathbf{v}\cdot(\mathbf{w}\times\mathbf{u})} = \underbrace{\det\left(\begin{bmatrix} \mathbf{w} & \mathbf{u} & \mathbf{v} \end{bmatrix}\right)}_{\mathbf{w}\cdot(\mathbf{u}\times\mathbf{v})}$$

Circular shifts in 3×3 determinants correspond to even permutations of columns.

(c) Identical columns result in the determinant equaling zero.

(d) Consider the parallelepiped with sides $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$. Treating the plane spanned by $\mathbf{v}$ and $\mathbf{w}$ as the base, $\|\mathbf{v} \times \mathbf{w}\|$ gives the area of the base. The altitude from the position vector $\mathbf{u}$ to the plane spanned by $\mathbf{v}$ and $\mathbf{w}$ is given by $\|\mathbf{u}\| \cos\theta$, where θ is the angle between $\mathbf{u}$ and the normal vector to the base (which is precisely the vector $\mathbf{v} \times \mathbf{w}$). A conceptual diagram helps here, which we encourage the reader to sketch. Therefore,

$$\text{Volume} = \underbrace{\|\mathbf{v} \times \mathbf{w}\|}_{\text{base area}} \underbrace{\|\mathbf{u}\| \cos\theta}_{\text{altitude}} = \|\mathbf{u}\| \|\mathbf{v} \times \mathbf{w}\| \cos\theta = |\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})| .$$

17 The hard work has already been done in the previous two exercises. Here, we apply those results to the spherical triangle.

(a) Let $\overrightarrow{u} = \overrightarrow{OB} \times \overrightarrow{OA}$. We apply (1.24) to compute

$$\begin{aligned}\overrightarrow{u} \times (\overrightarrow{OA} \times \overrightarrow{OC}) &= \left(\overrightarrow{u} \cdot \overrightarrow{OC}\right) \overrightarrow{OA} - \underbrace{\left(\overrightarrow{u} \cdot \overrightarrow{OA}\right)}_{0} \overrightarrow{OC} \\ &= \left(\overrightarrow{OC} \cdot \overrightarrow{u}\right) \overrightarrow{OA} = \left(\overrightarrow{OC} \cdot \left(\overrightarrow{OB} \times \overrightarrow{OA}\right)\right) \overrightarrow{OA} \\ &= \det\left(\begin{bmatrix} \overrightarrow{OC} & \overrightarrow{OB} & \overrightarrow{OA} \end{bmatrix}\right) \overrightarrow{OA} ,\end{aligned}$$

where we have used the fact that $\overrightarrow{u} \cdot \overrightarrow{OA} = \overrightarrow{OA} \cdot \overrightarrow{u} = \det\left(\begin{bmatrix} \overrightarrow{OA} & \overrightarrow{OB} & \overrightarrow{OA} \end{bmatrix}\right) = 0$.

(b) Nothing to do besides noting that $\|\overrightarrow{OB} \times \overrightarrow{OA}\|^2 = \sin^2 c$ and $\|\overrightarrow{OA} \times \overrightarrow{OC}\|^2 = \sin^2 b$.

(c) Analogous steps in parts (a) and (b) for the definition of $\sin^2 B$ in (1.27).

(d) The two determinants are obtained from each other by permuting columns. Hence, either they are equal or are negative of each other. Their squared values are equal and we do not need to check if the permutation is even or odd. The rest is immediate.

18 The Cauchy-Binet formula is a well-known result on determinants. We prove this for vector cross-products without using the more general formula. Let $\vec{a} \times \vec{b} = \vec{x}$. Then,

$$\begin{aligned}(\vec{a} \times \vec{b}) \cdot (\vec{c} \times \vec{d}) &= \vec{x} \cdot (\vec{c} \times \vec{d}) = \vec{c} \cdot (\vec{d} \times \vec{x}) = \vec{c} \cdot (\vec{d} \times (\vec{a} \times \vec{b})) \\ &= \vec{c} \cdot \left((\vec{d} \cdot \vec{b})\vec{a} - (\vec{d} \cdot \vec{a})\vec{b}\right) = (\vec{a} \cdot \vec{c})(\vec{b} \cdot \vec{d}) - (\vec{a} \cdot \vec{d})(\vec{b} \cdot \vec{c}) ,\end{aligned}$$

where the second equality follows from (1.26) and the fourth equality follows from (1.24).

Chapter 2

8(a) Conditional on $\mathbf{u}$, finite realizations of Y are clearly Gaussian since $(Y(\mathbf{s}_i))_{i=1}^n = (W(x_i))_{i=1}^n$, where $x_i = \mathbf{s}_i^{\mathrm{T}}\mathbf{u}$, and W is Gaussian. The covariance function is given by $\mathrm{Cov}(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) = c(\mathbf{h}^{\mathrm{T}}\mathbf{u})$, where c is the (stationary) covariance function of W.

(b) For the marginal process, we need to take expectation over the distribution of $\mathbf{u}$, which is uniform over the n-dimensional sphere. Note that

$$\begin{aligned}\mathrm{Cov}(Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) &= \mathrm{E}_{\mathbf{u}}\left[\mathrm{Cov}\left(W(\mathbf{s}^{\mathrm{T}}\mathbf{u}), W\left((\mathbf{s}+\mathbf{h})^{\mathrm{T}}\mathbf{u}\right)\right)\right] \\ &= \mathrm{E}_{\mathbf{u}}\left[c(\mathbf{h}^{\mathrm{T}}\mathbf{u})\right] .\end{aligned}$$

Then, we need to show that $\mathrm{E}_{\mathbf{u}}[c(\mathbf{h}^{\mathrm{T}}\mathbf{u})]$ is a function of $||\mathbf{h}||$. Now, $\mathbf{h}^{\mathrm{T}}\mathbf{u} = ||\mathbf{h}||\cos\theta$, so $\mathrm{E}_{\mathbf{u}}[c(\mathbf{h}^{\mathrm{T}}\mathbf{u})] = \mathrm{E}_{\theta}[c(||\mathbf{h}||\cos\theta)]$. But θ, being the angle made by a uniformly distributed random vector $\mathbf{u}$, has a distribution that is invariant over the choice of $\mathbf{h}$. Thus, the marginal process $Y(\mathbf{s})$ has isotropic covariance function $K(r) = \mathrm{E}_{\theta}[c(r\cos\theta)]$.

Note: The above covariance function (in $\Re^n$) can be computed using spherical integrals as

$$K(r) = \frac{2\Gamma(n/2)}{\sqrt{\pi}\Gamma((n-1)/2)}\int_0^1 c(r\nu)\left(1-\nu^2\right)^{(n-3)/2} d\nu.$$

9 If $\tau^2 = 0$, then $\Sigma = \sigma^2 H(\phi)$. If $\mathbf{s}_0 = \mathbf{s}_k$, where $\mathbf{s}_k$ is a monitored site, we have $\boldsymbol{\gamma}^{\mathrm{T}} = \sigma^2 [H(\phi)]_{k*}$, the kth row of $\sigma^2 H(\phi)$. Thus, $\mathbf{e}_k^{\mathrm{T}} H(\phi) = \left(1/\sigma^2\right)\boldsymbol{\gamma}^{\mathrm{T}}$, where $\mathbf{e}_k = (0,\ldots,1,\ldots,0)^{\mathrm{T}}$ is the kth coordinate vector. So $\mathbf{e}_k^{\mathrm{T}} = \left(1/\sigma^2\right)\boldsymbol{\gamma}^{\mathrm{T}} H^{-1}(\phi)$. Substituting this into equation (2.22), we get

$$\mathrm{E}[Y(\mathbf{s}_k)|\mathbf{y}] = \mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{e}_k^{\mathrm{T}}(\mathbf{y} - X\boldsymbol{\beta}) = \mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta} + y(\mathbf{s}_k) - \mathbf{x}_k^{\mathrm{T}}\boldsymbol{\beta} = y(\mathbf{s}_k) .$$

When $\tau^2 > 0$, $\Sigma = \sigma^2 H(\phi) + \tau^2 I$, so the Σ^{-1} in equation (2.22) does not simplify, and we do not have the above result.

Chapter 4

1 Brook's Lemma, equation (4.7), is easily verified as follows: Starting with the extreme right-hand side, observe that

$$\frac{p(y_{10},\ldots,y_{n0})}{p(y_{n0}|y_{10},\ldots,y_{n-1,0})} = p(y_{10},\ldots,y_{n-1,0}) .$$

Now observe that

$$p(y_n|y_{10},\ldots,y_{n-1,0})p(y_{10},\ldots,y_{n-1,0}) = p(y_{10},\ldots,y_{n-1,0},y_n) .$$

The result follows by simply repeating these two steps, steadily moving leftward through (4.7).

3 We provide two different approaches to solving the problem. The first approach is a direct manipulative approach, relying upon elementary algebraic simplifications, and might seem a bit tedious. The second approach relies upon some relatively advanced concepts in matrix analysis, yet does away with most of the manipulations of the first approach.

Method 1: In the first method we derive the following identity:

$$\mathbf{u}^{\mathrm{T}} D^{-1}(I - B)\mathbf{u} = \sum_{i=1}^{n}\frac{u_i^2}{\tau_i^2}\left(1 - \sum_{j=1}^{n} b_{ij}\right) + \sum_{i<j}\frac{b_{ij}}{\tau_i^2}(u_i - u_j)^2 , \qquad (17.46)$$

where $\mathbf{u} = (u_1, ..., u_n)^{\mathrm{T}}$. Note that if this identity is indeed true, the right-hand side must be strictly positive; all the terms in the r.h.s. are strictly positive by virtue of the conditions on the elements of the B matrix unless $\mathbf{u} = \mathbf{0}$. This would imply the required positive definiteness.
We may derive the above identity either by starting with the l.h.s. and eventually obtaining the r.h.s., or vice versa. We adopt the former. So,

$$\begin{aligned}
\mathbf{u}^{\mathrm{T}} D^{-1}(I-B)\mathbf{u} &= \sum_i \frac{u_i^2}{\tau_i^2} - \sum_i \sum_j \frac{b_{ij}}{\tau_i^2} u_i u_j \\
&= \sum_i \frac{u_i^2}{\tau_i^2} - \sum_i \frac{b_{ii}}{\tau_i^2} u_i^2 - \sum_i \sum_{j \neq i} \frac{b_{ij}}{\tau_i^2} u_i u_j \\
&= \sum_i \frac{u_i^2}{\tau_i^2} (1 - b_{ii}) - \sum_i \sum_{j \neq i} \frac{b_{ij}}{\tau_i^2} u_i u_j \,.
\end{aligned}$$

Adding and subtracting $\sum_i \sum_{j \neq i} \left(u_i^2/\tau_i^2\right) b_{ij}$ to the last line of the r.h.s., we write

$$\begin{aligned}
\mathbf{u}^{\mathrm{T}} D^{-1}(I-B)\mathbf{u} &= \sum_i \frac{u_i^2}{\tau_i^2} \left(1 - \sum_j b_{ij}\right) + \sum_i \sum_{j \neq i} \frac{u_i^2}{\tau_i^2} b_{ij} \\
&\quad - \sum_i \sum_{j \neq i} \frac{b_{ij}}{\tau_i^2} u_i u_j \\
&= \sum_i \frac{u_i^2}{\tau_i^2} \left(1 - \sum_j b_{ij}\right) + \sum_i \sum_{j \neq i} \frac{b_{ij}}{\tau_i^2} \left(u_i^2 - u_i u_j\right) \\
&= \sum_i \frac{u_i^2}{\tau_i^2} \left(1 - \sum_j b_{ij}\right) + \sum_{i<j} \frac{b_{ij}}{\tau_i^2} \left(u_i - u_j\right)^2 \,.
\end{aligned}$$

To explain the last manipulation,

$$\sum_i \sum_{j \neq i} \frac{b_{ij}}{\tau_i^2} \left(u_i^2 - u_i u_j\right) = \sum_{i<j} \frac{b_{ij}}{\tau_i^2} \left(u_i - u_j\right)^2 , \tag{17.47}$$

note that the sum on the l.h.s. of (17.47) extends over the $2 \times \binom{n}{2}$ (unordered) pairs of (i, j). Consider any particular pair, say, (k, l) with $k < l$, and its "reflection" (l, k). Using the symmetry condition, $b_{kl}/\tau_k^2 = b_{lk}/\tau_l^2$, we may combine the two terms from this pair as

$$\frac{b_{kl}}{\tau_k^2} \left(u_k^2 - u_k u_l\right) + \frac{b_{lk}}{\tau_l^2} \left(u_l^2 - u_l u_k\right) = \frac{b_{kl}}{\tau_k^2} \left(u_k - u_l\right)^2 .$$

Performing the above trick for each of the $\binom{n}{2}$ pairs, immediately results in (17.47).

Method 2: The algebra above may be skipped using the following argument, based on eigenanalysis. First, note that, with the given conditions on B, the matrix $D^{-1}(I - B)$ is (weakly) diagonally dominant. This means that, if $A = D^{-1}(I - B)$, and $R_i(A) = \sum_{j \neq i} |a_{ij}|$ (the sum of the absolute values of the ith row less that of the diagonal element), then $|a_{ii}| \geq R_i(A)$, for all i, with strict inequality for at least one i. Now, using the Gershgorin Circle Theorem (see, e.g., Theorem 7.2.1 Golub and Van Loan, 2013, p. 320), we immediately see that 0 cannot be an interior point of Gershgorin circle. Therefore, all the eigenvalues of A must be nonnegative. But note that all the elements of B are strictly positive. This means that all the elements of A are nonzero, which means that

0 cannot be a boundary point of the Gershgorin circle. Therefore, 0 must be an exterior point of the circle, proving that all the eigenvalues of A must be strictly positive. So A, being symmetric, must be positive definite.

Note: It is important that the matrix D be chosen so as to ensure $D^{-1}(I-B)$ is symmetric. To see that this condition cannot be relaxed, consider the following example. Let us take $B = \begin{pmatrix} 0.3 & 0.5 \\ 0.1 & 0.9 \end{pmatrix}$. Clearly the matrix satisfies the conditions laid down in the problem statement. If we are allowed to choose an arbitrary D, we may take $D = I_2$, the 2×2 identity matrix, and so $D^{-1}(I-B) = \begin{pmatrix} 0.7 & -0.5 \\ -0.1 & 0.1 \end{pmatrix}$. But this is not positive definite, as is easily seen by noting that with $\mathbf{u}^{\mathrm{T}} = (1,2)$, we obtain $\mathbf{u}^{\mathrm{T}} D^{-1}(I-B)\,\mathbf{u} = -0.1 < 0$.

4 Using the identity in (17.46), it is immediately seen that, taking B to be the scaled proximity matrix (as in the text just above equation (4.15)), we have $\sum_{j=1}^{n} b_{ij} = 1$, for each i. This shows that the first term on the r.h.s. of (17.46) vanishes, leading to the second term, which is a pairwise difference prior.

Chapter 5

1 The complete BUGS code to fit this model is given below. Recall "#" is a comment in BUGS, so this version actually corresponds model for part (c).

```
model
{
    for (i in 1:N) {
        y[i] ~ dbern(p[i])
#        logit(p[i]) <- b0 + b1*kieger[i] + b2*team[i]
#        logit(p[i]) <- b0 + b2*(team[i]-mean(team[]))
        logit(p[i]) <- b0 + b1*(pct[i]-mean(pct[]))
        pct[i] <- kieger[i]/(kieger[i]+team[i])
}
   b0 ~ dnorm(0, 1.E-3)
   b1 ~ dnorm(0, 1.E-3)
   b2 ~ dnorm(0, 1.E-3)
}

HERE ARE INITS:
list(b0=0, b1=0, b2=0)


HERE ARE THE DATA:
list(N = 9,                          # number of observations
  y = c(1,1,1,1,0,1,1,1,0),                   # team win/loss
  kieger = c(31,31,36,30,32,33,31,33,32),  # Kieger points
  team = c(31,16,35,42,19,37,29,23,15))    # team points
```

Running a single Gibbs sampling chain for 20,000 iterations after a 1,000-iteration burn-in period, Table 17.3 gives the resulting 95% equal tail posterior credible intervals for β_1 and β_2 for each model, as well as the corresponding DIC and p_D scores.

(a) Running this model produces MCMC chains with slowly moving sample traces and very high autocorrelations and cross-correlations (especially between β_0 and β_1, since Kieger's uncentered scores are nearly identical). The 95% equal-tail confidence interval

	95% Credible intervals			
Model	β_1	β_2	DIC	p_D
(a)	(–3.68, 1.21)	(.152, 2.61)	8.82	1.69
(b)	—	(.108, 1.93)	9.08	1.61
(c)	(–70.8, –3.65)	—	8.07	1.59

Table 17.3 *Posterior summaries, Carolyn Kieger prep basketball logit model.*

for β_1 includes 0, suggesting Kieger's score is not a significant predictor of game outcome; the p_D score of just 1.69 also suggests there are not 3 "effective" parameters in the model (although none of these posterior summaries are very trustworthy due to the high autocorrelations, hence low effective sample MCMC sample size). Thus, the model is not acceptable either numerically (poor convergence; unstable estimates due to low effective sample size) or statistically (model is overparametrized).

(b) Since "kieger" was not a significant predictor in part (a), we delete it, and center the remaining covariate ("team") around its own mean. This helps matters immensely: numerically, convergence is much better and parameter and other estimates are much more stable. Statistically, the DIC score is not improved (slightly higher), but the p_D is virtually unchanged at 1.6 (so both of the remaining parameters in the model are needed), and β_2 is more precisely estimated.

(c) Again convergence is improved, and now the DIC score is also better. β_1 is significant and negative, since the higher the proportion of points scored by Kieger (i.e., the lower the output by the rest of the team), the less likely a victory becomes.

(d) The p_i themselves have posteriors implied by the β_j posteriors and reveal that the team was virtually certain to win Games 1, 3, 4, 6, and 7 (where Kieger scored a lower percentage of the points), but could well have lost the others, especially Games 2 and 9 (the former of which the team was fortunate to win anyway). This implies that the only thing that might still be missing from our model is some measure of how *few* points the opponent scores, which is of course governed by how well Kieger and the other team members play on *defense*. But fitting such a model would obviously require defensive statistics (blocked shots, etc.) that we currently lack.

5(a) In our implementation of `BUGS`, we obtained a slightly better DIC score with Model XI (7548.3, versus 7625.2 for Model XII), suggesting that the full time-varying complexity is not required in the survival model. The fact that the 95% posterior credible interval for γ_3, (–0.43, .26), includes 0 supports this conclusion.

(b) We obtained point and 95% interval estimates of –0.20 and (–0.25, –0.14) for γ_1, and –1.61 and (–2.13, –1.08) for γ_2.

(c) Figure 17.9 plots the estimated posteriors (smoothed histograms of `BUGS` output). In both the separate (panel a) and joint (panel b) analyses, this patient's survival is clearly better if he receives ddC instead of ddI. However, the joint analysis increases the estimated median survival times by roughly 50% in both groups.

(d) Estimation of the random effects in `NLMIXED` is via empirical Bayes, with associated standard errors obtained by the delta method. Approximate 95% prediction intervals can then be obtained by assuming asymptotic normality. We obtained point and interval estimates in rough agreement with the above `BUGS/JAGS` results, and for broadly comparable computer runtimes (if anything, our `NLMIXED` code ran slower). However, the asymmetry of some of the posteriors in Figure 17.9 (recall they are truncated at 0) suggests traditional confidence intervals based on asymptotic normality and approximate standard errors will not be very accurate. Only the fully Bayesian-MCMC

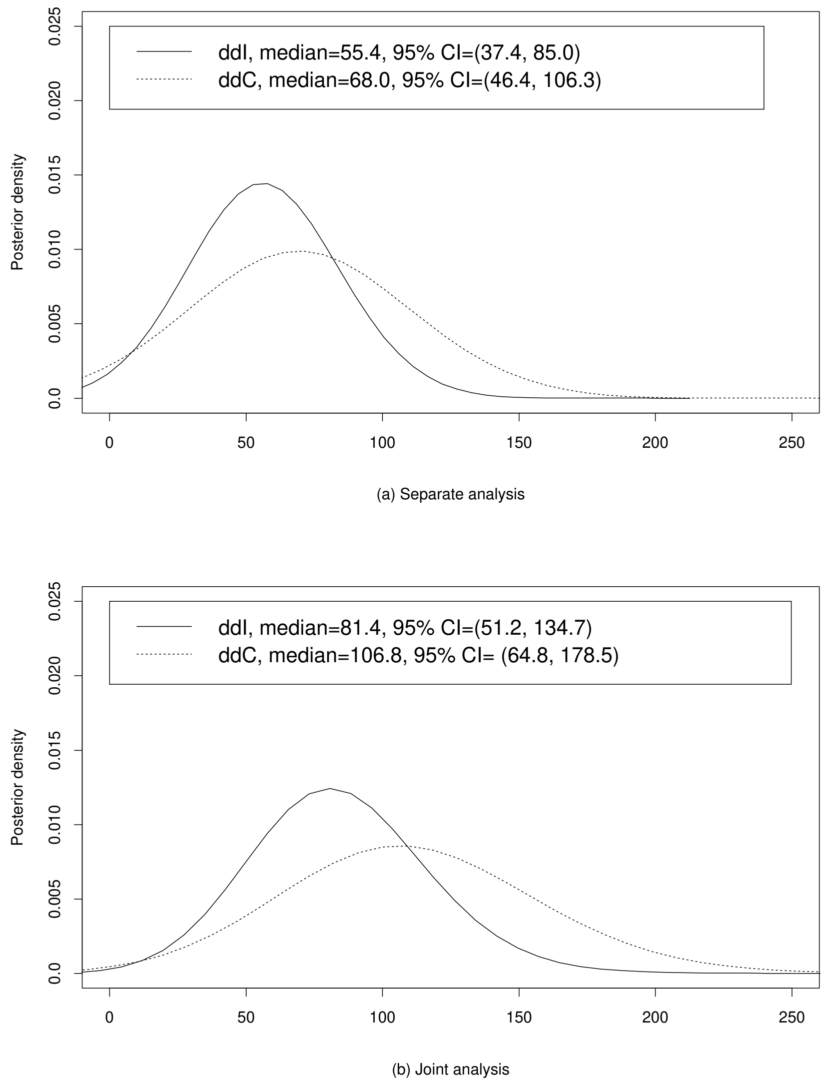

Figure 17.9 Median survival time for a hypothetical patient (male, negative AIDS diagnosis at study entry, intolerant of AZT): (a) estimated posterior density of median survival time of the patient from separate analysis; (b) estimated posterior density of median survival time of the patient from joint analysis.

(`BUGS/JAGS`) approach can produce exact results and corresponding full posterior inference.

Chapter 6

1 The calculations for the full conditionals for $\boldsymbol{\beta}$ and $\mathbf{W}$ follow from the results of the general linear model given in Example 4.2. Thus, with a $N(A\boldsymbol{\alpha}, V)$ prior on $\boldsymbol{\beta}$, (i.e.,

$p(\boldsymbol{\beta}) = N(A\boldsymbol{\alpha}, V))$ the full conditional for $\boldsymbol{\beta}$ is $N(D\mathbf{d}, D)$, where

$$\begin{aligned} D^{-1} &= \left(\frac{1}{\tau^2}X^{\mathrm{T}}X + V^{-1}\right)^{-1} \\ \text{and } \mathbf{d} &= \frac{1}{\tau^2}X^{\mathrm{T}}(\mathbf{Y} - \mathbf{W}) + V^{-1}A\boldsymbol{\alpha}. \end{aligned}$$

Note that with a flat prior on $\boldsymbol{\beta}$, we set $V^{-1} = 0$ to get

$$\boldsymbol{\beta}|\mathbf{Y}, \mathbf{W}, X, \tau^2 \sim N\left(\left(X^{\mathrm{T}}X\right)^{-1}X^{\mathrm{T}}(\mathbf{Y} - \mathbf{W}), \tau^2\left(X^{\mathrm{T}}X\right)^{-1}\right).$$

Similarly for $\mathbf{W}$, since $p(\mathbf{W}) = N\left(\mathbf{0}, \sigma^2 H(\phi)\right)$, the full conditional distribution is again of the form $N(D\mathbf{d}, D)$, but where this time

$$\begin{aligned} D^{-1} &= \left(\frac{1}{\tau^2}I + \frac{1}{\sigma^2}H^{-1}(\phi)\right)^{-1} \\ \text{and } \mathbf{d} &= \frac{1}{\tau^2}(\mathbf{Y} - X\boldsymbol{\beta}). \end{aligned}$$

Next, with $p\left(\tau^2\right) = IG(a_\tau, b_\tau)$, we compute the full conditional distribution for τ^2, $p\left(\tau^2|\mathbf{Y}, X, \boldsymbol{\beta}, \mathbf{W}\right)$, as proportional to

$$\begin{aligned} &\tfrac{1}{(\tau^2)^{a_\tau+1}}\exp\left(-b_\tau/\tau^2\right) \\ &\quad\times \tfrac{1}{(\tau^2)^{n/2}}\exp\left(-\tfrac{1}{2\tau^2}(\mathbf{Y} - X\boldsymbol{\beta} - \mathbf{W})^{\mathrm{T}}(\mathbf{Y} - X\boldsymbol{\beta} - \mathbf{W})\right) \\ &\propto \tfrac{1}{(\tau^2)^{a_\tau+n/2}}\exp\left(-\tfrac{1}{\tau^2}\left(b_\tau + \tfrac{1}{2}(\mathbf{Y} - X\boldsymbol{\beta} - \mathbf{W})^{\mathrm{T}}(\mathbf{Y} - X\boldsymbol{\beta} - \mathbf{W})\right)\right), \end{aligned}$$

where n is the number of sites. Thus we have the conjugate distribution

$$IG\left(a_\tau + \frac{n}{2}\,,\, b_\tau + \frac{1}{2}(\mathbf{Y} - X\boldsymbol{\beta} - \mathbf{W})^{\mathrm{T}}(\mathbf{Y} - X\boldsymbol{\beta} - \mathbf{W})\right).$$

Similar calculations for the spatial variance parameter, σ^2, yield a conjugate full conditional when $p\left(\sigma^2\right) = IG(a_\sigma, b_\sigma)$, namely

$$\sigma^2 \mid \mathbf{W}, \phi \sim IG\left(a_\sigma + \frac{n}{2}\,,\, b_\sigma + \frac{1}{2}\mathbf{W}^{\mathrm{T}}H^{-1}(\phi)\,\mathbf{W}\right).$$

Finally, for the spatial correlation function parameter ϕ, no closed form solution is available, and one must resort to Metropolis-Hastings or slice sampling for updating. Here we would need to compute

$$p\left(\phi|\mathbf{W}, \sigma^2\right) \propto p(\phi) \times \exp\left(-\frac{1}{2\sigma^2}\mathbf{W}^{\mathrm{T}}H^{-1}(\phi)\,\mathbf{W}\right).$$

Typically the prior $p(\phi)$ is taken to be uniform or gamma.

12(a) These relationships follow directly from the definition of $w(\mathbf{s})$ in equation (3.9):

$$\begin{aligned} &\mathrm{Cov}\left(w(\mathbf{s}), w(\mathbf{s}')\right) \\ &\quad= \mathrm{Cov}\left(\int_{\Re^2} k(\mathbf{s} - \mathbf{t})\,z(\mathbf{t})\,d\mathbf{t}\,,\, \int_{\Re^2} k(\mathbf{s}' - \mathbf{t})\,z(\mathbf{t})\,d\mathbf{t}\right) \\ &\quad= \sigma^2\int_{\Re^2} k(\mathbf{s} - \mathbf{t})\,k(\mathbf{s}' - \mathbf{t})\,d\mathbf{t}\,. \end{aligned}$$

and

$$\text{Var}\left(w\left(\mathbf{s}\right)\right) = \sigma^2 \int_{R^2} k^2\left(\mathbf{s}-\mathbf{t}\right) d\mathbf{t},$$

obtained by setting $\mathbf{s}=\mathbf{s}'$ above.

(b) This follows exactly as above, except that we adjust for the covariance in the stationary $z(\mathbf{t})$ process:

$$\begin{aligned}
\text{Cov}\left(w\left(\mathbf{s}\right), w\left(\mathbf{s}'\right)\right) &= \int_{\Re^2}\int_{\Re^2} k\left(\mathbf{s}-\mathbf{t}\right) k\left(\mathbf{s}'-\mathbf{t}\right) \\
&\qquad\qquad \times \text{Cov}\left(z\left(\mathbf{t}\right), z\left(\mathbf{t}'\right)\right) d\mathbf{t}d\mathbf{t}' \\
&= \sigma^2 \int_{\Re^2}\int_{\Re^2} k\left(\mathbf{s}-\mathbf{t}\right) k\left(\mathbf{s}'-\mathbf{t}\right) \rho\left(\mathbf{t}-\mathbf{t}'\right) d\mathbf{t}d\mathbf{t}' \\
\text{and } \text{Var}\left(w\left(\mathbf{s}\right)\right) &= \sigma^2 \int_{\Re^2}\int_{\Re^2} k\left(\mathbf{s}-\mathbf{t}\right) k\left(\mathbf{s}-\mathbf{t}'\right) \rho\left(\mathbf{t}-\mathbf{t}'\right) d\mathbf{t}d\mathbf{t}',
\end{aligned}$$

obtained by setting $\mathbf{s}=\mathbf{s}'$ above.

10 From (6.38), the full conditional $p(\phi_i|\boldsymbol{\phi}_{j\neq i}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{y})$ is proportional to the product of a Poisson and a normal density. On the log scale we have

$$\log p(\phi_i|\boldsymbol{\phi}_{j\neq i}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{y}) \propto -E_i e^{\mathbf{x}_i'\boldsymbol{\beta}+\theta_i+\phi_i} + \phi_i y_i - \frac{\tau_c m_i}{2}(\phi_i - \bar{\phi}_i)^2 .$$

Taking two derivatives of this expression, it is easy to show that in fact $(\partial^2/\partial\phi_i^2)\log p(\phi_i|\boldsymbol{\phi}_{j\neq i}, \boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{y}) < 0$, meaning that the log of the full conditional is a concave function, as required for ARS sampling.

Chapter 7

6(a) Denoting the likelihood by L, the prior by p, and writing $\mathbf{y} = (y_1, y_2)$, the joint posterior distribution of m_1 and m_2 is given as

$$\begin{aligned}
p(m_1, m_2 \mid \mathbf{y}) &\propto L(m_1, m_2; \mathbf{y}) p(m_1, m_2) \\
&\propto (7m_1 + 5m_2)^{y_1} e^{-(7m_1+5m_2)} \\
&\quad \times (6m_1 + 2m_2)^{y_2} e^{-(6m_1+2m_2)} \\
&\quad \times m_1^{a-1} e^{-m_1/b} m_2^{a-1} e^{-m_2/b} ,
\end{aligned}$$

so that the resulting full conditional distributions for m_1 and m_2 are

$$\begin{aligned}
p(m_1|m_2, \mathbf{y}) &\propto (7m_1 + 5m_2)^{y_1} (6m_1 + 2m_2)^{y_2} m_1^{a-1} e^{-m_1(13+b^{-1})}; \\
p(m_2|m_1, \mathbf{y}) &\propto (7m_1 + 5m_2)^{y_1} (6m_1 + 2m_2)^{y_2} m_2^{a-1} e^{-m_2(7+b^{-1})} .
\end{aligned}$$

We see immediately that conjugacy is absent; these two expressions are not proportional to any standard distributional form. As such, one might think of univariate Metropolis updating to obtain samples from the joint posterior distribution $p(m_1, m_2|\mathbf{y})$, though since this is a very low-dimensional problem, the use of MCMC methods here probably constitutes overkill!

Drawing our Metropolis candidates from Gaussian distributions with means equal to the current chain value and variances $(0.3)^2$ and $(0.1)^2$ for δ_1 and δ_2, respectively, for each parameter we ran five independent sampling chains with starting points overdispersed with respect to the suspected target distribution for 2000 iterations. The observed Metropolis acceptance rates were 45.4% and 46.4%, respectively, near the 50%

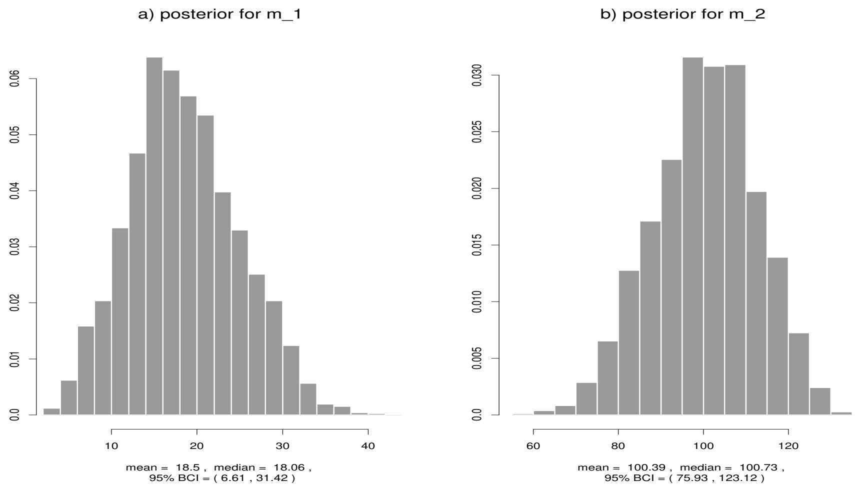

Figure 17.10 *Posterior histograms of sampled* $\mathbf{m}$ *values, motivating example.*

rate suggested by Gelman et al. (1996a) as well as years of Metropolis "folklore." The vagueness of the prior distributions coupled with the paucity of the data in this simple example (in which we are estimating two parameters from just two data points, y_1 and y_2) leads to substantial autocorrelation in the observed chains. However, plots of the observed chains as well as the convergence diagnostic of Gelman and Rubin (1992) suggested that a suitable degree of algorithm convergence obtains after 500 iterations. The histograms of the remaining $5 \times 1500 = 7500$ iterations shown in Figures 17.10(a) and (b) provide estimates of the marginal posterior distributions $p(m_1|\mathbf{y})$ and $p(m_2|\mathbf{y})$. We see that point estimates for m_1 and m_2 are 18.5 and 100.4, respectively, implying best guesses for $7m_1 + 5m_2$ and $6m_1 + 2m_2$ of 631.5 and 311.8, respectively, quite consistent with the observed data values $y_1 = 632$ and $y_2 = 311$. Also shown are 95% Bayesian credible intervals (denoted "95% BCI" in the figure legends), available simply as the 2.5 and 97.5 empirical percentiles in the ordered samples.

(b) By the Law of Iterated Expectation, $\mathrm{E}(Y_{3a}|\mathbf{y}) = \mathrm{E}[\mathrm{E}(Y_{3a}|\mathbf{m}, \mathbf{y})]$. Now we need the following well-known result from distribution theory:

Lemma: If $X_1 \sim Po(\lambda_1)$, $X_2 \sim Po(\lambda_2)$, and X_1 and X_2 are independent, then

$$X_1 \mid (X_1 + X_2 = n) \;\sim\; Bin\left(n \, , \, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right) . \quad \blacksquare$$

We apply this lemma in our setting with Y_{3a} playing the role of X_1, y_1 playing the role of n, and the calculation conditional on $\mathbf{m}$. The result is

$$\begin{aligned} \mathrm{E}(Y_{3a}|\mathbf{y}) &= \mathrm{E}[\mathrm{E}(Y_{3a}|\mathbf{m}, \mathbf{y})] \;=\; \mathrm{E}[\mathrm{E}(Y_{3a}|m_1, y_1)] \\ &= \mathrm{E}\left[y_1 \left(\frac{2m_1 + 2m_2}{7m_1 + 5m_2} \right) \Big| \, y_1 \right] \\ &\approx \frac{y_1}{G} \sum_{g=1}^{G} \frac{2m_1^{(g)} + 2m_2^{(g)}}{7m_1^{(g)} + 5m_2^{(g)}} \;\equiv\; \hat{E}(Y_{3a}|\mathbf{y}) \, , \end{aligned} \tag{17.48}$$

where $\{(m_1^{(g)}, m_2^{(g)}), \, g = 1, \ldots, G\}$ are the Metropolis samples drawn above. A similar calculation produces a Monte Carlo estimate of $\mathrm{E}(Y_{3b}|\mathbf{y})$, so that our final estimate of $\mathrm{E}(Y_3|\mathbf{y})$ is the sum of these two quantities. In our problem this turns out to be $\hat{E}(Y_3|\mathbf{y}) = 357.0$.

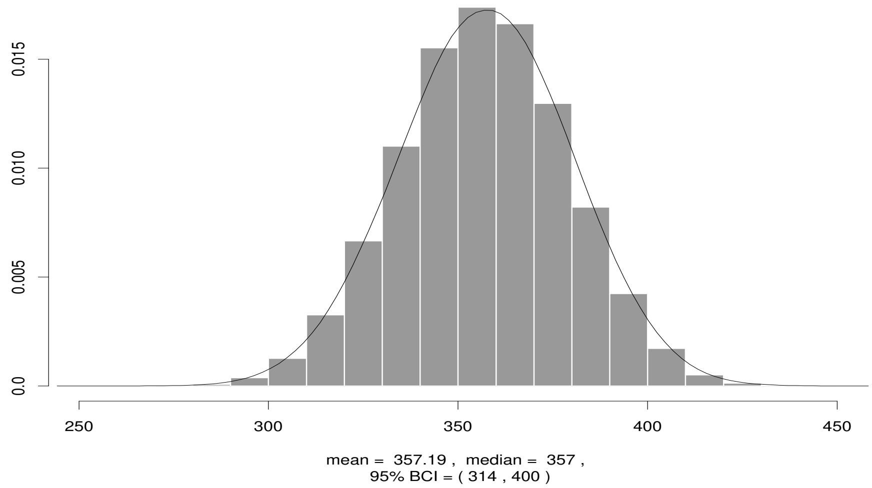

Figure 17.11 *Posterior histogram and kernel density estimate, sampled Y_3 values, motivating example.*

(c) Again using Monte Carlo integration, we write

$$p(y_3|\mathbf{y}) \;=\; \int p(y_3|\mathbf{m},\mathbf{y})p(\mathbf{m}|\mathbf{y})d\mathbf{m} \;\approx\; \frac{1}{G}\sum_{g=1}^{G} p(y_3|\mathbf{m}^{(g)},\mathbf{y}) \,.$$

Using the lemma again, $p(y_3|\mathbf{m},\mathbf{y})$ is the convolution of two independent binomials,

$$Y_{3a}|\mathbf{m},\mathbf{y} \sim Bin\left(y_1\,,\; \frac{2m_1+2m_2}{7m_1+5m_2}\right) \,, \tag{17.49}$$

$$\text{and} \quad Y_{3b}|\mathbf{m},\mathbf{y} \sim Bin\left(y_2\,,\; \frac{m_1+m_2}{6m_1+2m_2}\right) \,. \tag{17.50}$$

Since these two binomials do not have equal success probabilities, this convolution is a complicated (though straightforward) calculation that unfortunately will not emerge as another binomial distribution. However, we may perform the sampling analog of this calculation simply by drawing $Y_{3a}^{(g)}$ from $p(y_{3a}|\mathbf{m}^{(g)},y_1)$ in (17.49), $Y_{3b}^{(g)}$ from $p(y_{3b}|\mathbf{m}^{(g)},y_2)$ in (17.50), and defining $Y_3^{(g)} = Y_{3a}^{(g)} + Y_{3b}^{(g)}$. The resulting pairs $\{(Y_3^{(g)},\mathbf{m}^{(g)}),\ g=1,\ldots,G\}$ are distributed according to the joint posterior distribution $p(y_3,\mathbf{m}|\mathbf{y})$, so that marginally, the $\{Y_3^{(g)},\ g=1,\ldots,G\}$ values have the desired distribution, $p(y_3|\mathbf{y})$.

In our setting, we actually drew 25 $Y_{3a}^{(g)}$ and $Y_{3b}^{(g)}$ samples for each $\mathbf{m}^{(g)}$ value, resulting in 25(7500) = 187,500 $Y_3^{(g)}$ draws from the convolution distribution. A histogram of these values (and a corresponding kernel density estimate) is shown in Figure 17.11. The mean of these samples is 357.2, which agrees quite well with our earlier mean estimate of 357.0 calculated just below equation (17.48).

Chapter 10

4(a) This setup closely follows that below equation (2.21), so we imitate this argument in the case of a bivariate process, where now $\mathbf{Y}_1 = Y_1(\mathbf{s}_0)$ and $\mathbf{Y}_2 = \mathbf{y}$. Then, as in equation (2.22),

$$\mathrm{E}\left[Y_1(\mathbf{s}_0)\,|\mathbf{y}\right] = \mathbf{x}^{\mathrm{T}}(\mathbf{s}_0)\,\boldsymbol{\beta} + \boldsymbol{\gamma}^{\mathrm{T}}\Sigma^{-1}\left(\mathbf{y} - X\boldsymbol{\beta}\right) \;,$$

parameter	2.5%	50%	97.5%
θ_1	-0.437	-0.326	-0.216
β_1	3.851	5.394	6.406
β_2	-2.169	2.641	7.518
σ_1	0.449	0.593	2.553
σ_2	0.101	1.530	6.545
ϕ_1	0.167	0.651	0.980
ϕ_2	0.008	0.087	0.276
τ	4.135	5.640	7.176

Table 17.4 *Posterior quantiles for the conditional LMC model.*

where $\boldsymbol{\gamma}^{\mathrm{T}} = (\boldsymbol{\gamma}_1^{\mathrm{T}}, \boldsymbol{\gamma}_2^{\mathrm{T}})$, where $\boldsymbol{\gamma}_1^{\mathrm{T}} = (c_{11}(\mathbf{s}_0 - \mathbf{s}_1), \ldots, c_{11}(\mathbf{s}_0 - \mathbf{s}_n))$ and $\boldsymbol{\gamma}_2^{\mathrm{T}} = (c_{12}(\mathbf{s}_0 - \mathbf{s}_1), \ldots, c_{12}(\mathbf{s}_0 - \mathbf{s}_n))$. Also,

$$\Sigma_{2n\times 2n} = \begin{pmatrix} C_{11} & C_{12} \\ C_{12}^{\mathrm{T}} & C_{22} \end{pmatrix} + \begin{pmatrix} \tau_1^2 I_n & 0 \\ 0 & \tau_2^2 I_n \end{pmatrix},$$

with $C_{lm} = (c_{lm}(\mathbf{s}_i - \mathbf{s}_j))_{i,j=1,\ldots,n}$ with $l, m = 1, 2$.

(b) The approach in this part is analogous to that of Chapter 2, Exercise 9 Observe that with $\mathbf{s}_0 = \mathbf{s}_k$, $(\mathbf{e}_k^{\mathrm{T}} : \mathbf{0})\,\Sigma = \boldsymbol{\gamma}^{\mathrm{T}}$ if and only if $\tau_1^2 = 0$, where $\mathbf{e}_k^{\mathrm{T}} = (0, \ldots, 1, \ldots, 0)$ is the n-dimensional kth coordinate vector. This immediately leads to $\mathrm{E}[Y_1(\mathbf{s}_k)\,|\mathbf{y}] = y_1(\mathbf{s}_k)$; $\mathrm{E}[Y_2(\mathbf{s}_k)\,|\mathbf{y}] = y_2(\mathbf{s}_k)$ is shown analogously.

8(a) Let $Y_1(\mathbf{s})$ be the temperature at location $\mathbf{s}$, $Y_2(\mathbf{s})$ be the precipitation at location $\mathbf{s}$, and $X(\mathbf{s})$ be the elevation at location $\mathbf{s}$. We then fit the following conditional LMC, as in equation (10.78):

$$\begin{aligned} Y_1(\mathbf{s}) &= \theta_1 X(\mathbf{s}) + \sigma_1 w_1(\mathbf{s}) \\ Y_2(\mathbf{s})\,|Y_1(\mathbf{s}) &= \beta_1 X(\mathbf{s}) + \beta_2 Y_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s}) + \epsilon(\mathbf{s}) \,, \end{aligned}$$

where $\epsilon(\mathbf{s}) \sim N(0, \tau^2)$, $w_i(\mathbf{s}) \sim GP(0, \rho(\cdot, \phi_i))$, for $i = 1, 2$.

The file `https://github.com/sudiptobanerjee/BGC_2023/ColoradoLMCa.bug` contains the `BUGS` code for this problem. Table 17.4 gives a brief summary of the results. The results are more or less as expected: temperature is negatively associated with elevation, while precipitation is positively associated. Temperature and precipitation do not seem to be significantly associated with each other. The spatial smoothing parameters ϕ_1 and ϕ_2 were both assigned $U(0,1)$ priors for this analysis, but it would likely be worth investigating alternate choices in order to check prior robustness.

(b) These results can be obtained simply by switching Y_1 and Y_2 in the data labels for the model and computer code of part (a).

Chapter 11

1(a) This follows directly by noting $\mathrm{Var}(\mathbf{v}_1) = \lambda_{11} I_n$, $\mathrm{Var}(\mathbf{v}_2) = \lambda_{22} I_n$, and $cov(\mathbf{v}_1, \mathbf{v}_2) = \lambda_{12} I$.

(b) Note that $\mathrm{Var}(\boldsymbol{\phi}_1) = \lambda_{11} A_1 A_1^{\mathrm{T}}$, $\mathrm{Var}(\boldsymbol{\phi}_2) = \lambda_{22} A_2 A_2^{\mathrm{T}}$, and also that $cov(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2) = \lambda_{12} A_1 A_2^{\mathrm{T}}$. So with $A_1 = A_2$, the dispersion of $\boldsymbol{\phi}$ is given by

$$\Sigma_{(\boldsymbol{\phi})} = \begin{pmatrix} \lambda_{11} A A^{\mathrm{T}} & \lambda_{12} A A^{\mathrm{T}} \\ \lambda_{12} A^{\mathrm{T}} A & \lambda_{12} A A^{\mathrm{T}} \end{pmatrix} = \Lambda \otimes A A^{\mathrm{T}}.$$

Taking A as the square root of $(D_W - \rho W)^{-1}$ yields $\Sigma_{(\boldsymbol{\phi})} = \Lambda \otimes (D_W - \rho W)^{-1}$.

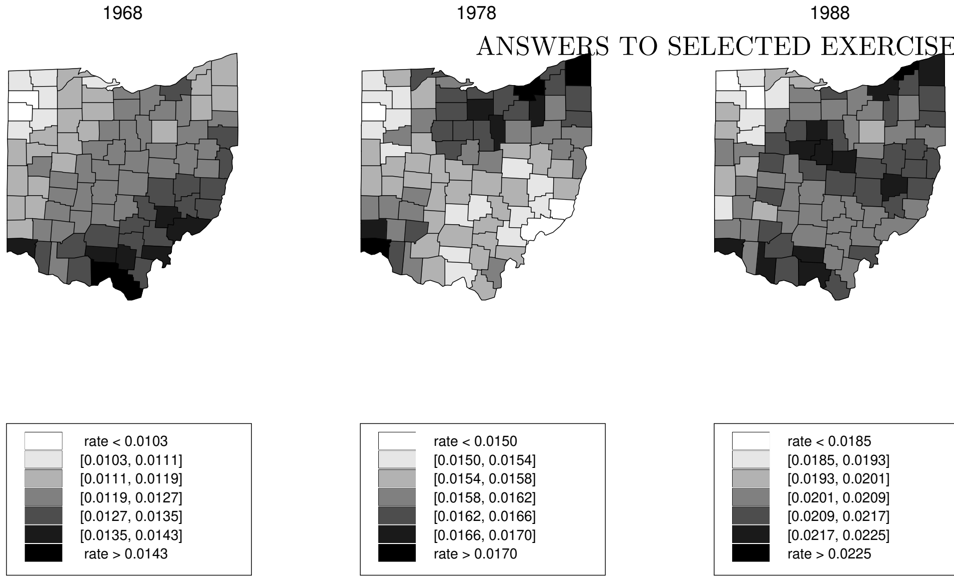

Figure 17.12 *Fitted median lung cancer death rates per 1000 population, nonwhite females.*

Note that the order of the Kronecker product is different from equation (11.4), since we have blocked the $\boldsymbol{\phi}$ vector by components rather than by areal units.

(c) In general, with $A_1 \neq A_2$, we have

$$\Sigma_{(\boldsymbol{\phi})} = \begin{pmatrix} \lambda_{11} A_1 A_1^{\mathrm{T}} & \lambda_{12} A_1 A_2^{\mathrm{T}} \\ \lambda_{12} A_2 A_1^{\mathrm{T}} & \lambda_{12} A_2 A_2^{\mathrm{T}} \end{pmatrix} = \mathcal{A} \left(\Lambda \otimes I \right) \mathcal{A}^{\mathrm{T}},$$

where $\mathcal{A} = BlockDiag\,(A_1, A_2)$. For the generalized MCAR, with different spatial smoothness parameters ρ_1 and ρ_2 for the different components, take A_i as the Cholesky square root of $(D_W - \rho_i W)^{-1}$ for $i = 1, 2$.

Chapter 12

2 The code in `https://github.com/sudiptobanerjee/BGC_2023/Chapter12/ColoradoS-T1.bug` fits model (12.6), the additive space-time model. This is a "direct" solution, where we explicitly construct the temporal process. By contrast, the file `https://github.com/sudiptobanerjee/BGC_2023/Chapter11/ColoradoS-T1.bug` uses the `spatial.exp` function, tricking it to handle temporal correlations by setting the y-coordinates to 0.

4(a) Running five chains of an MCMC algorithm, we obtained point and 95% interval estimates of –0.01 and [–0.20, 0.18] for β; using the same reparametrization under the chosen model (10) in Waller et al. (1997), the point and interval estimates instead are –0.20 and [–0.26, –0.15]. Thus, using this reparametrization shows that age adjusting has eliminated the statistical significance of the difference between the two female groups.

(b) Figure 17.12 shows the fitted age-adjusted lung cancer death rates per 1000 population for nonwhite females for the years 1968, 1978, and 1988. The scales of the three figures show that lung cancer death rates are increasing over time. For 1968, we see a strong spatial pattern of increasing rates as we move from northwest to southeast, perhaps the result of an unmeasured occupational covariate (farming versus mining). Except for persistent low rates in the northwest corner, however, this trend largely disappears over time, perhaps due to increased mixing of the population or improved access to quality health care and health education.

node (unit)	Mean	sd	MC error	2.5%	Median	97.5%
W_1 (A)	–0.0491	0.835	0.0210	–1.775	–0.0460	1.639
W_3 (C)	–0.183	0.9173	0.0178	–2.2	–0.136	1.52
W_5 (E)	–0.0320	0.8107	0.0319	–1.682	–0.0265	1.572
W_6 (F)	0.417	0.8277	0.0407	–1.066	0.359	2.227
W_9 (I)	0.255	0.7969	0.0369	–1.241	0.216	1.968
W_{11} (K)	–0.195	0.9093	0.0209	–2.139	–0.164	1.502
ρ_1 (A)	1.086	0.1922	0.0072	0.7044	1.083	1.474
ρ_3 (C)	0.901	0.2487	0.0063	0.4663	0.882	1.431
ρ_5 (E)	1.14	0.1887	0.0096	0.7904	1.139	1.521
ρ_6 (F)	0.935	0.1597	0.0084	0.6321	0.931	1.265
ρ_9 (I)	0.979	0.1683	0.0087	0.6652	0.971	1.339
ρ_{11} (K)	0.881	0.2392	0.0103	0.4558	0.861	1.394
τ	1.73	1.181	0.0372	0.3042	1.468	4.819
β_0	–7.11	0.689	0.0447	–8.552	–7.073	–5.874
β_1	0.596	0.2964	0.0105	0.0610	0.578	1.245
RR	3.98	2.951	0.1122	1.13	3.179	12.05

Table 17.5 *Posterior summaries, MAC survival model (10,000 samples, after a burn-in of 1,000).*

Chapter 15

1(a) Table 17.5 summarizes the results from the nonspatial model, which are based on 10,000 posterior samples obtained from a single MCMC chain after a burn-in of 1,000 iterations. Looking at this table and the raw data in Table 15.14, basic conclusions are as follows:

- Units A and E have moderate overall risk ($W_i \approx 0$) but increasing hazards ($\rho > 1$): few deaths, but they occur late.
- Units F and I have high overall risk ($W_i > 0$) but decreasing hazards ($\rho < 1$): several early deaths, many long-term survivors.
- Units C and K have low overall risk ($W_i < 0$) and decreasing hazards ($\rho < 1$): no deaths at all; a few survivors.
- The two drugs differ significantly: CI for β_1 (RR) excludes 0 (1).

2(b) The appropriate interval censored BUGS/JAGS code is as follows:

```
model
{
for (i in 1:N) {
  TimeSmoking[i] <- Age[i] - AgeStart[i]
  RelapseT[i] ~ dweib(rho[i],mu[i])I(censored.time1[i],
    censored.time2[i])
  log(mu[i]) <- beta0 + beta[1]*TimeSmoking[i]
    + beta[2]*SexF[i] + beta[3]*SIUC[i]
    + beta[4]*F10Cigs[i] + W[County[i]]
  rho[i] <- exp(lrho[County[i]])
  }

# for (i in 1:regions) {W[i] ~ dnorm(0.0, tau_W)}
# for (i in 1:regions) {lrho[i] ~ dnorm(0.0, tau_rho)}
```

```
for (i in 1:sumnum) {weights[i] <- 1}

W[1:regions] ~ car.normal(adj[], weights[], num[], tau_W)
lrho[1:regions] ~ car.normal(adj[], weights[], num[],
  tau_rho)

for (i in 1:4) { beta[i] ~ dnorm(0.0, 0.0001)}
beta0 ~ dnorm(0.0,0.0001)
tau_W ~ dgamma(0.1,0.1)
tau_rho ~ dgamma(0.1,0.1)
}
```