Bayesian Statistics:

An Introduction

PETER M. LEE

Formerly Provost of Wentworth College,
University of York, England

Fourth Edition

John Wiley & Sons, Ltd

Appendix D

Answers to Exercises

D.1 Exercises on Chapter 1

1. Considering trumps and non-trumps separately, required probability is

$$2\binom{3}{3}\binom{23}{10}\Big/\binom{26}{13} = \frac{11}{50}.$$

Probability of a $2:1$ split is 39/50, and the conditional probability that the king is the odd one is 1/3, so probability one player has the king and the other the remaining two is 13/50.

2. (a) If $\mathsf{P}(A) = 0$ or $\mathsf{P}(A) = 1$.

(b) Take A as "odd red die", B as "odd blue die" and C as "odd sum".

(c) Take C as $\{HHH, THH, THT, TTH\}$.

3. (a) $\mathsf{P}(\text{homozygous}) = 1/3$; $\mathsf{P}(\text{heterozygous}) = 2/3$.

(b) By Bayes' Theorem

$$\mathsf{P}(BB \,|\, 7 \text{ black}) = \frac{(1/3)(1)}{(1/3)(1) + (2/3)(1/2^7)} = \frac{64}{65}.$$

(c) Similarly we see by induction (wih case $n = 0$ given by part (a)) that $\mathsf{P}(BB \,|\, \text{first } n \text{ black}) = 2^{n-1}/(2^{n-1}+1)$ since

$$\begin{aligned}\mathsf{P}(BB \,|\, \text{first } n+1 \text{ black}) &= \frac{\{2^{n-1}/(2^{n-1}+1)\}(1)}{\{2^{n-1}/(2^{n-1}+1)\}(1) + \{1/(2^{n-1}+1)\}(1/2)} \\ &= \frac{2^n}{2^n+1}.\end{aligned}$$

4. Use $\mathsf{P}(GG) = \mathsf{P}(M)\mathsf{P}(GG|M) + \mathsf{P}(D)\mathsf{P}(GG|D)$ to get

$$\mathsf{P}(M) = \frac{\mathsf{P}(GG) - p^2}{p(1-p)}$$

5. $p(3) = p(4) = 1/36$, $p(5) = p(6) = 2/36$, $p(7) = \cdots = p(14) = 3/36$, $p(15) = p(16) = 2/36$, $p(17) = p(18) = 1/36$. As for the distribution function,

$$F(x) = \begin{cases} 0 & \text{for } x < 3; \\ 1/36 & \text{for } 3 \leqslant x < 4; \\ 2/36 & \text{for } 4 \leqslant x < 5; \\ 4/36 & \text{for } 5 \leqslant x < 6; \\ 6/36 & \text{for } 6 \leqslant x < 7; \\ (3[x]-12)/36 & \text{for } n \leqslant x < n+1 \text{ where } 7 \leqslant n < 15; \\ 32/36 & \text{for } 15 \leqslant x < 16; \\ 34/36 & \text{for } 16 \leqslant x < 17; \\ 35/36 & \text{for } 17 \leqslant x < 18; \\ 1 & \text{for } x \geqslant 18 \end{cases}$$

where $[x]$ denotes the integer part of x.

6. $\mathsf{P}(k=0) = (1-\pi)^n = (1-\lambda/n)^n \to \mathrm{e}^{-\lambda}$. More generally

$$\begin{aligned} p(k) &= \binom{n}{k}\pi^k(1-\pi)^{n-k} \\ &= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n \frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}\left(1-\frac{\lambda}{n}\right)^{-k} \\ &\to \frac{\lambda^k}{k!}\exp(-\lambda). \end{aligned}$$

7. (a) $\mathsf{P}\big(\widetilde{k}=0\big) = \mathsf{P}(\widetilde{m}=\widetilde{n}=0) = \mathsf{P}(\widetilde{m}=0)\,\mathsf{P}(\widetilde{n}=0) = \mathrm{e}^{-\lambda}\mathrm{e}^{-\mu} = \mathrm{e}^{-(\lambda+\mu)}$,

$$\begin{aligned} \mathsf{P}\big(\widetilde{k}=1\big) &= \mathsf{P}\big(\{\widetilde{m}=1,\,\widetilde{n}=0\} \text{ or } \{\widetilde{m}=0,\,\widetilde{n}=1\}\big) \\ &= \lambda\mathrm{e}^{-(\lambda+\mu)} + \mu\mathrm{e}^{-(\lambda+\mu)} = (\lambda+\mu)\mathrm{e}^{-(\lambda+\mu)}. \end{aligned}$$

(b) More generally

$$\begin{aligned} \mathsf{P}\big(\widetilde{k}=k\big) &= \sum_{m=0}^{k}\mathsf{P}(\widetilde{m}=m,\,\widetilde{n}=k-m) = \sum_{m=0}^{k}\mathsf{P}(\widetilde{m}=m)\,\mathsf{P}(\widetilde{n}=k-m) \\ &= \sum_{m=0}^{k}\frac{\lambda^k}{k!}\exp(-\lambda)\frac{\mu^{k-m}}{(k-m)!}\exp(-\mu) \end{aligned}$$

$$= \frac{1}{k!}\exp(-(\lambda+\mu))\sum_{m=0}^{k}\binom{k}{m}\lambda^m\mu^{k-m} = \frac{(\lambda+\mu)^k}{k!}\exp(-(\lambda+\mu)).$$

(c) By definition of conditional probability

$$\begin{aligned}
\mathsf{P}\big(\widetilde{m}=m\,|\,\widetilde{k}=k\big) &= \frac{\mathsf{P}\big(\widetilde{m}=m,\,\widetilde{k}=k\big)}{\mathsf{P}\big(\widetilde{k}=k\big)} = \frac{\mathsf{P}\big(\widetilde{m}=m,\;\;\widetilde{n}=k-m\big)}{\mathsf{P}\big(\widetilde{k}=k\big)} \\
&= \frac{\frac{\lambda^m}{m!}\exp(-\lambda))\,\frac{\mu^{k-m}}{(k-m)!}\exp(-\mu)}{\frac{(\lambda+\mu)^k}{k!}\exp(-(\lambda+\mu))} \\
&= \binom{k}{m}\left(\frac{\lambda}{\lambda+\mu}\right)^m\left(1-\frac{\lambda}{\lambda+\mu}\right)^{k-m}.
\end{aligned}$$

8. Let $y = x^2$ where $x \sim \mathrm{N}(0,1)$. Then

$$\begin{aligned}
\mathsf{P}\left(\widetilde{y}\leqslant y\right) &= \mathsf{P}\left(\widetilde{x}^2\leqslant y\right) = \mathsf{P}\left(-\sqrt{y}\leqslant\widetilde{x}\leqslant\sqrt{y}\,\right) \\
&= \mathsf{P}\left(\widetilde{x}\leqslant\sqrt{y}\,\right) - \mathsf{P}(\widetilde{x}<-\sqrt{x}\,)
\end{aligned}$$

so that (because $\mathsf{P}(\widetilde{x}=-\sqrt{y})=0$)

$$F_{\widetilde{y}}(y) = F_{\widetilde{x}}\left(\sqrt{y}\,\right) - F_{\widetilde{x}}\left(-\sqrt{y}\,\right)$$

and on differentiation

$$p_{\widetilde{y}}(y) = \tfrac{1}{2}y^{-\frac{1}{2}}p_{\widetilde{x}}\left(\sqrt{y}\,\right) + \tfrac{1}{2}y^{-\frac{1}{2}}p_{\widetilde{x}}\left(-\sqrt{y}\,\right).$$

Alternatively, you could argue that

$$\mathsf{P}(y<\widetilde{y}\leqslant y+\,\mathrm{d}y) = \mathsf{P}(x<\widetilde{x}\leqslant x+\,\mathrm{d}x) + \mathsf{P}(-x-\,\mathrm{d}x\leqslant\widetilde{x}<-x)$$

implying

$$p_{\widetilde{y}}(y)\,\mathrm{d}y = p_{\widetilde{x}}(x)\,\mathrm{d}x + p_{\widetilde{x}}(-x)\,\mathrm{d}x$$

which as $\mathrm{d}y/\,\mathrm{d}x = 2x = 2\sqrt{y}$ leads to the same conclusion.

In the case where $x \sim \mathrm{N}(0,1)$ this gives

$$p_{\widetilde{y}}(y) = (2\pi y)^{-\frac{1}{2}}\exp\left(-\tfrac{1}{2}y\right)$$

which is the density of χ_1^2.

9. By independence, since $\widetilde{M}\leqslant M$ iff every one of the indvidual X_i are less than or equal to M

$$\begin{aligned}
&F_M(M) = \mathsf{P}(X_i\leqslant M\quad\forall i) = (F(M))^n; && F_m(m) = 1-(1-F(m))^n; \\
&p_M(M) = nf(M)(F(M))^{n-1}; && p_m(m) = nf(m)(1-F(m))^{n-1}
\end{aligned}$$

10. Let m be the minimum of u and v and c be the length of the centre section. Then $p(m,c) = p_{\widetilde{u},\widetilde{v}}(m,c) + p_{\widetilde{u},\widetilde{v}}(c,m) = 2$ if $m + c \leqslant 1$ and 0 otherwise.

11. If $F(x,y) = F(x)F(y)$ then by differentiation with respect to x and with respect to y we have $p(x,y) = p(x)p(y)$. The converse follows by integration.

In discrete cases if $F(x,y) = F(x)F(y)$ then $p(x,y) = F(x,y) - F(x-1,y) - F(x,y-1) + F(x-1,y-1) = (F_{\widetilde{x}}(x) - F_{\widetilde{x}}(x-1))(F_{\widetilde{y}}(y) - F_{\widetilde{y}}(y-1)) = p(x)p(y)$. Coversely if $p(x,y) = p(x)p(y)$ then $F(x,y) = \sum_{\xi \leqslant x} \sum_{\eta \leqslant y} p_{\widetilde{x},\widetilde{y}}(\xi,\eta)$ and so $F(x,y) = \sum_{\xi \leqslant x} p_{\widetilde{x}}(\xi) \sum_{\eta \leqslant y} p_{\widetilde{y}}(\eta) = F(x)F(y)$.

12. $\mathsf{E}X = n(1-\pi)/\pi$; $\mathcal{V}X = n(1-\pi)\pi^2$.

13. $\mathsf{E}X = \sum \mathsf{E}Z_i^2 = \sum 1 = \nu$, while $\mathsf{E}X^2 = \sum \mathsf{E}Z_i^4 + \sum_{i \neq j} \mathsf{E}X_i^2 X_j^2 = 3\nu + \nu(\nu - 1)$ so that $\mathcal{V}Z = \mathsf{E}X^2 - (\mathsf{E}X)^2 = 2\nu$. Similarly by integration.

14. $\mathsf{E}X = n\pi$, $\mathsf{E}X(X-1) = n(n-1)\pi^2$ and $\mathsf{E}X(X-1)(X-2) = n(n-1)(n-2)\pi^3$, so $\mathsf{E}X = n\pi$, $\mathsf{E}X^2 = \mathsf{E}X(X-1) + \mathsf{E}X = n(n-1)\pi + n\pi$ and

$$\begin{aligned}
\mathsf{E}(X - \mathsf{E}X)^3 &= \mathsf{E}X^3 - 3(\mathsf{E}X^2)(\mathsf{E}X) + 2(\mathsf{E}X)^3 \\
&= \mathsf{E}X(X-1)(X-2) - 3(\mathsf{E}X^2)(\mathsf{E}X) + 3\mathsf{E}X^2 + 2(\mathsf{E}X)^3 - 2\mathsf{E}X \\
&= n\pi(1-\pi)(1-2\pi),
\end{aligned}$$

and thus $\gamma_1 = (1-2\pi)/\sqrt{[n\pi(1-\pi)]}$.

15. $\phi \geqslant \int_{\{x;\,|x-\mu| \geqslant c\}} c^2 p(x)\,\mathrm{d}x = c^2 \mathsf{P}(|x-\mu| \geqslant c)$.

16. By symmetry $\mathsf{E}x = \mathsf{E}y = \mathsf{E}xy = 0$ so $\mathcal{C}(x,y) = \mathsf{E}xy - \mathsf{E}x\mathsf{E}y = 0$. However $0 = \mathsf{P}(x=0,\, y=0) \neq \mathsf{P}(x=0)\,\mathsf{P}(y=0) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

17. $p(y|x) = \{2\pi(1-\rho^2)\}^{-\frac{1}{2}} \exp\{-\frac{1}{2}(y-\rho x)^2/(1-\rho^2)\}$ so conditional on $\widetilde{x} = x$ we have $y \sim \mathrm{N}(\rho x, 1-\rho^2)$, so $\mathsf{E}(y|x) = \rho x$. $\mathsf{E}(y^2|x) = \mathcal{V}(y|x) + \{\mathsf{E}(y|x)\}^2 = 1 - \rho^2 + \rho^2 x^2$, hence $\mathsf{E}(xy|x) = \rho x^2$, $\mathsf{E}(x^2y^2|x) = x^2 - \rho^2 x^2 + \rho^2 x^4$. Therefore $\mathsf{E}xy = \rho$ and (as $\mathsf{E}x^4 = 3$) $\mathsf{E}x^2y^2 = 1 + 2\rho^2$, so $\mathsf{E}xy - (\mathsf{E}x)(\mathsf{E}y) = \rho$ and $\mathsf{E}x^2y^2 - (\mathsf{E}x^2)(\mathsf{E}y^2) = 2\rho^2$. As $\mathcal{V}x = 1$ and $\mathcal{V}x^2 = \mathsf{E}x^4 - (\mathsf{E}x^2)^2 = 3 - 1 = 2 = \mathcal{V}y$, the result follows.

18. (a) $p(x,y) = (\lambda^x \mathrm{e}^{-x}/x!)\binom{x}{y}\pi^y(1-\pi)^{x-y}$ so adding over $x = y, y+1, \dots$ and using $\sum \lambda^{x-y}(1-\pi)^{x-y} = \mathrm{e}^{\lambda(1-\pi)}$ we get $p(y) = (\lambda\pi)^y \mathrm{e}^{-\lambda\pi}/y!$ so that $\widetilde{y} \sim \mathrm{P}(\lambda\pi)$. Now note that $\mathsf{E}_{\widetilde{y}|\widetilde{x}}(\widetilde{y}|\widetilde{x}) = x\pi$ and this has expectation $\lambda\pi$.

(b) Note that $\mathcal{V}_{\widetilde{y}|\widetilde{x}}(\widetilde{y}|\widetilde{x}) = x\pi(1-\pi)$ which has expectation $\lambda\pi(1-\pi)$ and that

$\mathsf{E}_{\widetilde{y}|\widetilde{x}}(\widetilde{y}|\widetilde{x}) = x\pi$ which has variance $\lambda\pi^2$ so that the right hand side adds to $\lambda\pi$, the variance of $\widetilde{y}$.

19. We note that

$$I = \int_0^\infty \exp(-\tfrac{1}{2}z^2)\,dz = \int_0^\infty \exp(-\tfrac{1}{2}(xy)^2)\,y\,dx$$

for any y (on setting $z = xy$). Putting z in place of y, it follows that for any z

$$I = \int_0^\infty \exp(-\tfrac{1}{2}(zx)^2)\,z\,dx$$

so that

$$I^2 = \left(\int_0^\infty \exp(-\tfrac{1}{2}z^2)\,dz\right)\left(\int_0^\infty \exp(-\tfrac{1}{2}(zx)^2)\,dx\right) = \int_0^\infty\int_0^\infty \exp\{-\tfrac{1}{2}(x^2{+}1)z^2\}\,z\,dz\,dx.$$

Now set $(1+x^2)z^2 = 2t$ so that $z\,dz = dt/(1+x^2)$ to get

$$\begin{aligned}
I^2 &= \int_0^\infty\int_0^\infty \exp(-t)\,\frac{dt}{(1+x^2)}\,dx = \left(\int_0^\infty \exp(-t)\,dt\right)\left(\int_0^\infty \frac{dx}{(1+x^2)}\right) \\
&= [-\exp(-t)]_0^\infty\,\left[\tan^{-1}x\right]_0^\infty = \left[1\right]\left[\tfrac{1}{2}\pi\right] \\
&= \frac{\pi}{2}
\end{aligned}$$

and hence $I = \sqrt{\pi/2}$ so that the integral of ϕ from $-\infty$ to ∞ is 1, and hence ϕ *is* a probability density function. This method is apparently due to Laplace (1812, Section 24, pages 94–95 in the first edition).

D.2 Exercises on Chapter 2

1. $p(\pi) = \big(\mathsf{B}(k+1, n-k+1)\big)^{-1}\pi^k(1-\pi)^{n-k}$ or $\pi \sim \mathrm{Be}(k+1, n-k+1)$.

2. $\overline{x} = 16.35525$, so assuming uniform prior, posterior is $\mathrm{N}(16.35525,\ 1/12)$. A 90% HDR is $16.35525 \pm 1.6449/\sqrt{12}$ or 16.35525 ± 0.47484, that is, the interval $(15.88041, 16.83009)$.

3. $x - \theta \sim \mathrm{N}(0,1)$ and $\theta \sim \mathrm{N}(16.35525,\ 1/12)$, so $x \sim \mathrm{N}(16.35525,\ 13/12)$.

4. Assumming uniform prior, posterior is $\mathrm{N}(\overline{x}, \phi/n)$, so take $n = k$. If prior variance is ϕ_0, posterior is $\{1/\phi_0 + n/\phi\}^{-1}$, so take n the least integer such that $n \geqslant (k-1)\phi/\phi_0$.

5. Posterior is $k(2\pi/25)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\theta-0.33)^2 \times 25\}$ for $\theta > 0$ and 0 for $\theta < 0$. Integrating we find $1 = k\mathsf{P}(X > 0)$ where $X \sim \mathrm{N}(0.33,\ 1/25)$, so $k = \{1 -$

$\Phi(-1.65)\}^{-1} = 1.052$. We now seek θ_1 such that $p(\theta \leqslant \theta_1|\boldsymbol{x}) = 0.95$ or equivalently $k\mathsf{P}(0 < X \leqslant \theta_1) = 1$ with k and X as before. This results in $0.95 = 1.052\{\Phi(5\theta_1 - 1.65) - \Phi(-1.65)\}$, so $\Phi(5\theta_1 - 1.65) = 0.95/1.052 + 0.0495 = 0.9525$ leading to $5\theta_1 - 1.65 = 1.67$ and so to $\theta_1 = 0.664$. The required interval is thus $[0, 0.664)$.n

6. From the point of view of someone starting from prior ignorance, my beliefs are equivalent to an observation of mean λ and variance ϕ and yours to one of mean μ and variance ψ, so after taking into account my beliefs such a person is able to use Section 2.2 and conclude $\theta \sim \mathrm{N}(\lambda_1, \phi_1)$ where $1/\phi_1 = 1/\phi + 1/\psi$ and $\lambda_1/\phi_1 = \lambda/\phi + \mu/\psi$.

7. The likelihood (after inserting a convenient constant) is

$$l(\boldsymbol{x}|\theta) = (2\pi\phi/n)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\overline{x} - \theta)^2/(\phi/n)\}.$$

Hence by Bayes' Theorem, within I_α

$$\begin{aligned} Ac(1-\varepsilon)(2\pi\phi/n)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\overline{x} - \theta)^2/(\phi/n)\} &\leqslant p(\theta|\boldsymbol{x}) \\ \leqslant Ac(1+\varepsilon)(2\pi\phi/n)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\overline{x} - \theta)^2/(\phi/n)\} & \end{aligned}$$

and outside I_α

$$0 \leqslant p(\theta|\boldsymbol{x}) \leqslant AMc(2\pi\phi/n)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\overline{x} - \theta)^2/(\phi/n)\},$$

where A is a constant equal to $p(\boldsymbol{x})^{-1}$. Using the right hand inequality for the region inside I_α we get

$$\begin{aligned} \int_{I_\alpha} p(\theta|\boldsymbol{x})\,\mathrm{d}\theta &\leqslant Ac(1+\varepsilon) \int_{I_\alpha} (2\pi\phi/n)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\overline{x} - \theta)^2/(\phi/n)\}\,\mathrm{d}\theta \\ &= Ac(1+\varepsilon) \int_{-\lambda_\alpha}^{\lambda_\alpha} (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}t^2)\,\mathrm{d}t, \text{ where } t = (\overline{x} - \theta)/\sqrt{\phi/n} \\ &= Ac(1+\varepsilon)[\Phi(\lambda_\alpha) - \Phi(-\lambda_\alpha)] = Ac(1+\varepsilon)(1-\alpha) \end{aligned}$$

since $\Phi(-\lambda_\alpha) = 1 - \Phi(\lambda_\alpha)$. Similarly, the same integral exceeds $AQc(1-\varepsilon)(1-\alpha)$, and, if J_α is the outside of I_α,

$$0 \leqslant \int_{J_\alpha} p(\theta|\boldsymbol{x})\,\mathrm{d}\theta \leqslant AMc\alpha.$$

Combining these results we have, since $\int_{I_\alpha \cup J_\alpha} p(\theta|\boldsymbol{x})\,\mathrm{d}\theta = 1$,

$$Ac(1-\varepsilon)(1-\alpha) \leqslant 1 \leqslant Ac[(1+\varepsilon)(1-\alpha) + M\alpha],$$

and hence

$$\frac{1}{(1+\varepsilon)(1-\alpha) + M\alpha} \leqslant Ac \leqslant \frac{1}{(1-\varepsilon)(1-\alpha)}.$$

The result now follows on remarking that the maximum value of the exponential in J_α occurs at the end-points $\theta \pm \lambda_\alpha\sqrt{\phi/n}$, where it has the value $\exp(-\frac{1}{2}\lambda_\alpha^2)$.

8. The likelihood is

$$\begin{aligned} l(\theta|\boldsymbol{x}) &\propto h(\theta)^n \exp\left\{\sum t(x_i)\psi(\theta)\right\} \\ &\propto h(\theta)^n \exp\left\{\exp\left[\log\psi(\theta) - \log\left(1\Big/\sum t(x_i)\right)\right]\right\}. \end{aligned}$$

This is of the data-translated form $g(\psi(\theta) - t(\boldsymbol{x}))$ *if* the function $h(x) \equiv 1$.

9. Take prior $S_0\chi_\nu^{-2}$ where $\nu = 8$ and $S_0/(\nu-2) = 100$ so $S_0 = 600$ and take $n = 30$ and $S = 13.2^2(n-1) = 5052.96$. Then (cf. Section 2.7) posterior is $(S_0+S)\chi_{\nu+n}^{-2}$ where $\nu + n = 38$ and $S_0 + S = 5652.96$. Values of χ^2 corresponding to a 90% HDR for $\log\chi^2_{38}$ are (interpolating between 35 and 40 degrees of freedom) 25.365 and 54.269 so 90% interval is $(104, 223)$.

10. $n = 10$, $\overline{x} = 5.032$, $S = 3.05996$, $s/\sqrt{n} = 0.1844$. Posterior for mean θ is such that $(\theta - 5.032)/0.1844 \sim \mathrm{t}_\nu$, so as $\mathrm{t}_{9,0.95} = 1.833$ a 90% HDR is $(4.69, 5.37)$. Posterior for variance ϕ is $S\chi_\nu^{-2}$, so as values of χ^2 corresponding to a 90% HDR for $\log\chi^2_9$ are 3.628 and 18.087 required interval is $(0.17, 0.84)$.

11. Sufficient statistic is $\sum_{i=1}^n x_i$, or equivalently $\overline{x}$.

12. Sufficient statistic is $(\sum_{i=1}^n x_i, \prod_{i=1}^n x_i)$ or equivalently $(\overline{x}, \check{x})$ where $\check{x}$ is the geometric mean.

13. $p(\beta) \propto \beta^{-\alpha_0}\exp(-\xi/\beta)$.

14. If $\theta \sim \mathrm{U}(-\pi/2, \pi/2)$ and $x = \tan\theta$, then by the usual change-of-variable rule $p(x) = \pi^{-1}|\mathrm{d}\theta/\,\mathrm{d}x| = \pi^{-1}\left(1+x^2\right)^{-1}$, and similarly if $\theta \sim \mathrm{U}(\pi/2, 3\pi/2)$. The result follows.

15. Straightforwardly

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\pi}) &= \frac{n!}{x!\,y!\,z!}\pi^x\rho^y\sigma^z \\ &= \left(\frac{n!}{x!\,y!\,(n-x-y)!}\right) \times \exp\{n\log(1-\pi-\rho)\} \\ &\qquad \times \exp[x\log\{\pi/(1-\pi-\rho\} + y\log\{\rho/(1-\pi-\rho)\}] \\ &= g(x,y) \times h(\pi,\rho) \times \exp[t(x,y)\psi(\pi,\rho) + u(x,y)\chi(\pi,\rho)] \end{aligned}$$

16. $\nu_1 = \nu_0 + n = 104$; $n_1 = n_0 + n = 101$; $\theta_1 = (n_0\theta_0 + n\overline{x})/n_1 = 88.96$; $S = (n-1)s^2 = 2970$; $S_1 = S_0 + S + (n_0^{-1} + n^{-1})^{-1}(\theta_0 - \overline{x})^2 = 3336$; $s_1/\sqrt{n_1} =$

$\sqrt{3336/(101\times 104)} = 0.56$. It follows that *a posteriori*

$$\frac{\mu-\theta_1}{s_1/\sqrt{n_1}} \sim \mathrm{t}_{\nu_1}, \qquad \phi \sim S_1\chi_{\nu_1}^{-2}.$$

For prior, find 75% point of t_4 from, e.g., Neave's Table 3.1 as 0.741. For posterior, as degrees of freedom are large, can approximate t by normal, noting that the 75% point of the standard normal distribution is 0.6745. Hence a 50% prior HDR for μ is $85 \pm 0.741\sqrt{(350/4\times 1)}$, that is $(75.6, 94.4)$, while a 50% posterior HDR for μ is $88.96 \pm 0.6745 \times 0.56$, that is, $(88.58, 89.34)$.

17. With $p_1(\theta)$ being $\mathrm{N}(0,1)$ we see that $p_1(\theta|x)$ is $\mathrm{N}(2,2)$, and with $p_2(\theta)$ being $\mathrm{N}(1,1)$ we see that $p_2(\theta|x)$ is $\mathrm{N}(3,1)$. As

$$\begin{aligned}
\int p(x|\theta)p_1(\theta)\,\mathrm{d}\theta &= \int \frac{1}{2\pi}\exp\{-\tfrac{1}{2}(x-\theta)^2 - \tfrac{1}{2}\theta^2\}\,\mathrm{d}\theta \\
&= \frac{\sqrt{\frac{1}{2}}}{\sqrt{2\pi}}\exp\{-\tfrac{1}{4}x^2\} \\
&\quad \times \int \frac{1}{\sqrt{2\pi\frac{1}{2}}}\exp\{-\tfrac{1}{2}(\theta-\tfrac{1}{2}x)^2/\tfrac{1}{2}\}\,\mathrm{d}\theta \\
&= \frac{\sqrt{\frac{1}{2}}}{\sqrt{2\pi}}\exp\{-\tfrac{1}{4}x^2\} = \frac{1}{2\sqrt{\pi}}\exp\{-2\}
\end{aligned}$$

and similarly

$$\begin{aligned}
\int p(x|\theta)p_1(\theta)\,\mathrm{d}\theta &= \int \frac{1}{2\pi}\exp\{-\tfrac{1}{2}(x-\theta)^2 - \tfrac{1}{2}(\theta-1)^2\}\,\mathrm{d}\theta \\
&= \frac{\sqrt{\frac{1}{2}}}{\sqrt{2\pi}}\exp\{-\tfrac{1}{4}(x-1)^2\} \\
&\quad \times \int \frac{1}{\sqrt{2\pi\frac{1}{2}}}\exp\{-\tfrac{1}{2}(\theta-\tfrac{1}{2}(x+1))^2/\tfrac{1}{2}\}\,\mathrm{d}\theta \\
&= \frac{\sqrt{\frac{1}{2}}}{\sqrt{2\pi}}\exp\{-\tfrac{1}{4}(x-1)^2\} = \frac{1}{2\sqrt{\pi}}\exp\{-\tfrac{1}{4}\}
\end{aligned}$$

so that just as in Section 2.13 we see that the posterior is an α' to β' mixture of $\mathrm{N}(2,1)$ and $\mathrm{N}(3,1)$ where $\alpha' \propto \frac{2}{3}\mathrm{e}^{-2} = 0.09$ and $\beta' \propto \frac{1}{3}\mathrm{e}^{-1/4} = 0.26$, so that $\alpha' = 0.26$ and $\beta' = 0.74$. It follows that

$$\mathsf{P}(\theta > 1) = 0.26 \times 0.1587 + 0.74 \times 0.0228 = 0.058.$$

18. Elementary manipulation gives

$$
\begin{aligned}
& n\overline{X}^2 + n_0\theta_0^2 - (n+n_0)\left(\frac{n\overline{X}+n_0\theta_0}{n+n_0}\right)^2 \\
&= \frac{1}{n+n_0}[\{n(n+n_0)-n^2\}\overline{X}^2 + \{n_0(n+n_0)-n_o^2\}\theta_0^2 - 2(nn_0)\overline{X}\theta] \\
&= \frac{nn_0}{n+n_0}[\overline{X}^2+\theta_0^2-2\overline{X}\theta_0] = (n_0^{-1}+n^{-1})^{-1}(\overline{X}-\theta)^2.
\end{aligned}
$$

D.3 Exercises on Chapter 3

1. Using Bayes postulate $p(\pi)=1$ for $0 \leqslant \pi \leqslant 1$ we get a posterior $(n+1)\pi^n$ which has mean $(n+1)/(n+2)$.

2. From Table B.5, $\underline{F}_{40,24} = 0.55$ and $\overline{F}_{40,24} = 1.87$, so for $\text{Be}(20,12)$ take lower limit $20 \times 0.55/(12+20\times 0.55) = 0.48$ and upper limit $20 \times 1.87/(12+20\times 1.87) = 0.76$ so 90% HDR $(0.48.0.76)$. Similarly by interpolation $\underline{F}_{41,25} = 0.56$ and $\overline{F}_{41,25} = 1.85$, so for $\text{Be}(20.5,12.5)$ quote $(0.48.0.75)$. Finally by interpolation $\underline{F}_{42,26} = 0.56$ and $\overline{F}_{42,26} = 1.83$, so for $\text{Be}(21,13)$ quote $(0.47.0.75)$. It does not seem to make much difference whether we use a $\text{Be}(0,0)$, a $\text{Be}(\frac{1}{2},\frac{1}{2})$ or a $\text{Be}(1,1)$ prior.

3. Take $\alpha/(\alpha+\beta) = 1/3$ so $\beta = 2\alpha$ and

$$\mathcal{V}\pi = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{2}{3^2(3\alpha+1)}$$

so $\alpha = 55/27 \cong 2$ and $\beta = 4$. Posterior is then $\text{Be}(2+8, 4+12)$, that is $\text{Be}(10,16)$. The 95% values for $F_{32,20}$ are 0.45 and 2.30 by interpolation, so those for $F_{20,32}$ are 0.43 and 2.22. An appropriate interval for π is from $10 \times 0.43/(16+10\times 0.43$ to $10 \times 2.22/(16+10\times 2.22)$, that is $(0.21, 0.58)$.

4. Take $\alpha/(\alpha+\beta) = 0.4$ and $\alpha+\beta = 12$, so approximately $\alpha = 5$, $\beta = 7$. Posterior is then $\text{Be}(5+12, 7+13)$, that is, $\text{Be}(17,20)$.

5. Take beta prior with $\alpha/(\alpha+\beta) = \frac{1}{3}$ and $\alpha+\beta = \frac{1}{4}11 = 2.75$. It is convenient to take integer α and β, so take $\alpha = 1$ and $\beta = 2$, giving $\alpha/(\alpha+\beta) = \frac{1}{3}$ and $(\alpha+\beta)/11 = 0.273$. Variance of prior is $2/36 = 0.0555$ so standard deviation is 0.236. Data is such that $n = 11$ and $X = 3$, so posterior is $\text{Be}(4,10)$. From tables, values of F corresponding to a 50% HDR for $\log \text{F}_{20,8}$ are $\underline{F} = 0.67$ and $\overline{F} = 1.52$. It follows that the appropriate interval of values of $\text{F}_{8,20}$ is $(1/\overline{F}, 1/\underline{F})$, that is $(0.66, 1.49)$. Consequently and appropriate interval for the proportion π required is $4\times 0.66/(10+4\times 0.66) \leqslant \pi \leqslant 4 \times 1.49/(10+4\times 1.49)$, that is $(0.21, 0.37)$. The posterior mean is $4/(4+10) = 0.29$, the posterior mode is $(4-1)/(4+10-2) =$

0.25 and using the relationship median $\cong (2\,\mathrm{mean} + \mathrm{mode})/3$ the posterior mode is approximately 0.28. The actual overall proportion $86/433 = 0.20$ is consistent with the above. [Data from the York A.U.T. Membership List for 1990/91.]

6. By Appendix A.13, $\mathsf{E}x = n(1-\pi)/\pi$ and $\mathcal{V}x = n(1-\pi)/\pi^2$, so $g'(\mathsf{E}x) = \frac{1}{2}n^{-1}[(1-\pi)/\pi^2]^{-\frac{1}{2}}$, so $\mathcal{V}x = 1/4n$ (cf. Section 1.5). By analogy with the argument that an arc-sine prior for the binomial distribution is approximately data-translated, this suggests a prior uniform in $\sinh^{-1}\sqrt{\pi}$ so with density $\frac{1}{2}\pi^{-\frac{1}{2}}(1+\pi)^{-\frac{1}{2}}$. But note Jeffreys' Rule suggests $\mathrm{Be}(0, \frac{1}{2})$ as remarked in Section 7.4.

7. As this is a rare event, we use the Poisson distribution. We have $n = 280$, $T = \sum x_i = 196$. Posterior is $S_1^{-1}\chi^2_{\nu'}$, where $S_1 = S_0 + 2n$, $\nu' = \nu + 2T$. For reference prior take $\nu = 1$, $S_0 = 0$ we get a posterior which is $560^{-1}\chi^2_{393}$. Using the approximation in Appendix A.3, we find that a 95% interval is $\frac{1}{2}(\sqrt{785} \pm 1.96)^2/560$, that is, $(0.61, 0.80)$.

8. Prior is such that $\nu/S_0 = 0.66$, $2\nu/S_0^2 = 0.115^2$, so $\nu = 66$, $S_0 = 100$, so posterior has $S_1 = 660$, $\nu' = 458$. This gives a posterior 95% interval $\frac{1}{2}(\sqrt{915} \pm 1.96)^2/660$, that is, $(0.61, 0.79)$.

9. $\partial p(x|\alpha)/\partial\alpha = 3/2\alpha - \frac{1}{2}x^2$ and $\partial^2 p(x|\alpha)/\partial\alpha^2 = -3/2\alpha^2$, so $I(\alpha|x) = 3/2\alpha^2$ and we take $p(\alpha) \propto 1/\alpha$ or equivalently a uniform prior in $\psi = \log\alpha$.

10. We have $\partial^2 L/\partial\pi^2 = -x/\pi^2 - z/(1-\pi-\rho)^2$, etc., so

$$I(\pi, \rho \,|\, x, y, z) = \begin{pmatrix} n/\pi + n/(1-\pi-\rho) & n/(1-\pi-\rho) \\ n/(1-\pi-\rho) & n/\rho + n/(1-\pi-\rho) \end{pmatrix}$$

and so after a little manipulation $\det I(\pi, \rho \,|\, x, y, z) = n\{\pi\rho(1-\pi-\rho)\}^{-1}$, suggesting a prior

$$p(\pi, \rho) \propto \pi^{-\frac{1}{2}}\rho^{-\frac{1}{2}}(1-\pi-\rho)^{-\frac{1}{2}}$$

which is evidently related to the arc-sine distribution.

11. $\partial p(x|\gamma)/\partial\gamma = 1/\gamma + \log\xi - \log x$ and $\partial^2 p(x|\gamma)/\partial\gamma^2 = -1/\gamma^2$, so $I(\gamma|x) = 1/\gamma^2$ and we take $p(\gamma) \propto 1/\gamma$ or equivalently a uniform prior in $\psi = \log\gamma$.

12. Using Appendix A.16, coefficient of variation is $\sqrt{\{2/(\gamma+1)(\gamma-2)\}}$. This is less than 0.01 if $\frac{1}{2}(\gamma+1)(\gamma-2) > 1/0.01^2$ or $\gamma^2 - \gamma - 20{,}002 > 0$, so if $\gamma > \frac{1}{2}(1 + \sqrt{80{,}009}) = 141.9$. Taking the reference prior $p(\alpha, \beta) \propto (\beta - \alpha)^{-2}$, that is, $\mathrm{Pabb}(-\infty, \infty, -1)$ (cf. Section 3.6), we need $\gamma' = n - 1 > 141.9$, that is, n at least 143.

13. Take prior $p(\theta) = (d-1)\theta^{-d}$ for $\theta_0 < \theta < \infty$. Then posterior is $p(\theta|\boldsymbol{x}) = (d'-1)\theta^{-d'}$ for $\theta_1 < \theta < \infty$ where $d' = d + n(c+1)$ and $\theta_1 = \max\{\theta_0, M\}$ where $M = \max\{x_i\}$.

14. Prior $p(\nu) \propto 1/\nu$ as before, but likelihood is now $p(71, 100 \,|\, \nu) = 1/\nu^2$ for $\nu \geqslant 100$, so posterior is approximately

$$p(\nu \,|\, 71, 100) \propto \nu^{-3} \Big/ \left(\sum_{\mu \geqslant 100} \mu^{-3} \right).$$

Approximating sums by integrals, the posterior probability $\mathsf{P}(\nu \geqslant \nu_0 \,|\, 71, 100) = 100^2/\nu_0^2$, and in particular the posterior median is $100\sqrt{2}$ or about 140.

15. We have $p(\theta) = 1/\theta$ and setting $\psi = \log\theta$ we get $p(\psi) = p(\theta)\,|\mathrm{d}\theta/\mathrm{d}\psi| = 1$. Thus we expect all pages to be more or less equally dirty.

16. The group operation is $(x, y) \mapsto x + y \mod 2\pi$ and the appropriate Haar measure is Lebesgue measure on the circle, that is, arc length around the circle.

17. Table of $c^2\{c^2 + (x_1 - \mu)^2\}^{-1}\{c^2 + (x_2 - \mu)^2\}^{-1}$:

$\mu \backslash c$	1	2	3	4	4
0	0.00	0.01	0.01	0.01	0.01
2	0.06	0.05	0.04	0.03	0.02
4	0.04	0.06	0.05	0.04	0.03
6	0.06	0.05	0.04	0.03	0.02
8	0.00	0.01	0.01	0.01	0.01

Integrating over c using Simpson's Rule (and ignoring the constant) for the above values of μ we get 0.11, 0.48, 0.57, 0.48 and 0.11 respectively. Integrating again we get for intervals as shown:

$(-1, 1)$	$(1, 3)$	$(3, 5)$	$(5, 7)$	$(7, 9)$	Total
0.92	2.60	3.24	2.60	0.92	10.28

so the required posterior probability is $3.24/10.28 = 0.31$.

18. First part follows as sum of concave functions is concave and a concave function has a unique maximum. For the example, note that

$$\begin{aligned} p(x|\theta) &= \exp(\theta - x)/\{1 + \exp(\theta - x)\}^2 \quad (-\infty < x < \infty) \\ &= \tfrac{1}{4}\operatorname{sech}^2 \tfrac{1}{2}(\theta - x) \quad (-\infty < x < \infty) \end{aligned}$$

(which is symmetrical about $x = \theta$), so that the log-likelihood is

$$L(\theta|x) = \theta - x - 2\log\{1 + \exp(\theta - x)\}.$$

Hence

$$\begin{aligned}
L'(\theta|x) &= 1 - 2\exp(\theta - x)/\{1 + \exp(\theta - x)\} \\
&= \{1 - \exp(\theta - x)\}/\{1 + \exp(\theta - x)\} \\
&= 1 - 2/\{1 + \exp(x - \theta)\} \\
L''(\theta|x) &= -2\exp(x - \theta)/\{1 + \exp(x - \theta)\}^2 \\
&= -2\exp(\theta - x)/\{1 + \exp(\theta - x)\}^2.
\end{aligned}$$

As this is clearly always negative, the log-likelihood is concave. Also

$$\begin{aligned}
L'(\theta|x)/L''(\theta|x) &= \tfrac{1}{2}\{\exp(\theta - x) - \exp(x - \theta)\} \\
I(\theta|x) &= 2\int_{-\infty}^{\infty} \exp(2(\theta - x))/\{1 + \exp(\theta - x)\}^4 \,\mathrm{d}x \\
&= (1/8)\int_{-\infty}^{\infty} \operatorname{sech}^4(\theta - x)\,\mathrm{d}x \\
&= (1/8)\int_{-\infty}^{\infty} \operatorname{sech}^4 y\,\mathrm{d}y \\
&= (1/24)[\sinh y \operatorname{sech}^3 y + 2\tanh y]_{-\infty}^{\infty} = 1/6.
\end{aligned}$$

(The integration can be checked by differentiation of the result). Now proceed as in Section 3.10.

D.4 Exercises on Chapter 4

1. $1 - (1 - p_0) = p_0 = \left[1 + (1 - \pi_0)\pi_0^{-1}B^{-1}\right]^{-1} = 1 - (1 - \pi_0)\pi_0^{-1}B^{-1} + (1 - \pi_0)^2\pi_0^{-2}B^{-2}$. Result then follows on noting $\pi_0^{-1} = \{1 - (1 - \pi_0)\}^{-1} \cong 1 + (1 - \pi_0)$.

2. Substituting in the formulae in the text we get

$$\begin{aligned}
&\phi_1 = (0.9^{-2} + 1.8^{-2})^{-1} = 0.648 = 0.80^2; \\
&\theta_1 = 0.648(93.3/0.9^2 + 93.0/1.8^2) = 93.2; \\
&\pi_o = \Phi((93.0 - 93.3)/0.9) = \Phi(-0.333) = 0.3696; \;\; \pi_0/(1 - \pi_0) = 0.59; \\
&p_0 = \Phi((93.0 - 93.2)/0.8) = \Phi(-0.25) = 0.4013; \;\; p_0/(1 - p_0) = 0.67; \\
&B = 0.67/0.59 = 1.14.
\end{aligned}$$

3. $n = 12$, $\overline{x} = 118.58$, $S = 12969$, $s/\sqrt{n} = 9.91$, $(\overline{x} - 100)/(s/\sqrt{n}) = 1.875$. Taking a normal approximation this gives a P-value of $1 - \Phi(1.875) = 0.0303$.

4. $n = 300$, $n\pi = 900/16 = 56.25$, $n\pi(1 - \pi) = 45.70$, $(56 - 56.25)/\sqrt{45.70} = -0.037$, so certainly not significant.

5. Posterior with reference prior is $S_1^{-1}\chi^2_{\nu'}$ with $S_1 = 12$ and $\nu' = 31$. Values of χ^2 corresponding to a 90% HDR for $\log\chi^2_{31}$ are (by interpolation) 19.741 and 45.898, so a 90% posterior HDR is from 1.65 to 3.82 which includes 3. So not appropriate to reject null hypothesis.

6. If $k = 0.048$ then $\exp(2k) = 1.101 = 1/0.908$, so take θ within $\pm\varepsilon = k\sqrt{(\phi/n)}\mathrm{z} = 0.048\sqrt{(\phi/10)}/2.5 = 0.006\sqrt{\phi}$.

7. Using standard power series expansions

$$\begin{aligned}
B &= (1+1/\lambda)^{\frac{1}{2}}\exp[-\tfrac{1}{2}(1+\lambda)^{-1}] \\
&= \lambda^{-\frac{1}{2}}(1+\lambda)^{\frac{1}{2}}\exp(-\tfrac{1}{2}z^2)\exp[\tfrac{1}{2}\lambda z^2(1+\lambda)^{-1}] \\
&= \lambda^{\frac{1}{2}}(1+\tfrac{1}{2}\lambda+\dots)\exp(-\tfrac{1}{2}z^2)[1+\tfrac{1}{2}z^2(1+\lambda)^{-1}+\dots] \\
&= \lambda^{\frac{1}{2}}\exp(-\tfrac{1}{2}z^2)(1+\tfrac{1}{2}\lambda(z^2+1)+\dots).
\end{aligned}$$

8. Likelihood ratio is

$$\begin{aligned}
\frac{\{2\pi(\phi+\varepsilon)\}^{-n/2}\exp[-\frac{1}{2}\sum x_i^2/(\phi+\varepsilon)]}{\{2\pi(\phi-\varepsilon)\}^{-n/2}\exp[-\frac{1}{2}\sum x_i^2/(\phi-\varepsilon)]} &\cong \left(1+\frac{\varepsilon}{\phi}\right)^{-n}\exp\left[\varepsilon\sum x_i^2/\phi^2\right] \\
&\cong \exp\left[\varepsilon\left(\sum x_i^2 - n\phi\right)/\phi^2\right] \\
&\cong \exp\left[\frac{n\varepsilon}{\phi}\left(\frac{\sum x_i^2/n}{\phi}-1\right)\right].
\end{aligned}$$

9. $\partial p_1(\overline{x})/\partial\psi = \{-\frac{1}{2}(\psi+\phi/n)^{-1}+\frac{1}{2}(\overline{x}-\theta)^2/(\psi+\phi/n)^2\}p_1(\overline{x})$ which vanishes if $\psi+\phi/n = (\overline{x}-\theta)^2 = z^2\phi/n$ so if $\psi = (z^2-1)\phi/n$. Hence $p_1(\overline{x}) \leqslant (2\pi\phi/n)^{-\frac{1}{2}}z^{-1}\exp(-\frac{1}{2})$. Consequently $B = (2\pi\phi/n)^{-\frac{1}{2}}\exp(-\frac{1}{2}z^2)/p_1(\overline{x}) \geqslant \sqrt{\mathrm{e}}\,z\exp(-\frac{1}{2}z^2)$. For last part use $p_0 = \left[1+(\pi_1/\pi_0)B^{-1}\right]$.

10. Posterior probability is a minimum when B is a minimum, hence when $2\log B$ is a minimum, and

$$\mathrm{d}\,(2\log B)/\mathrm{d}n = \{1+n\}^{-1} + z^2\{1+1/n\}^{-2}(-1/n^2)$$

which vanishes when $n = z^2-1$. Since z^2-1 is not necessarily an integer, this answer is only approximate.

11. Test against $\mathrm{B}(7324, 0.75)$ with mean 5493 and variance 1373, so $z = (5474 - 5493)/\sqrt{1373} = -0.51$. Not significant so theory confirmed.

12. Likelihood ratio is

$$\frac{(2\pi 2\phi)^{-\frac{1}{2}}\exp[-\frac{1}{2}u^2/(2\phi)](2\pi\psi)^{-\frac{1}{2}}\exp[-\frac{1}{2}(z-\mu)/\psi]}{(2\pi 2\psi)^{-\frac{1}{2}}\exp[-\frac{1}{2}u^2/(2\psi)](2\pi\psi/2)^{-\frac{1}{2}}\exp[-\frac{1}{2}(z-\mu)/(\psi/2)]}$$
$$= (\psi/2\phi)^{\frac{1}{2}}\exp[-\tfrac{1}{2}u^2(1//2\phi - 1/2\psi) + \tfrac{1}{2}(z-\mu)^2/\psi]$$
$$\cong (\psi/2\phi)^{\frac{1}{2}}\exp[-\tfrac{1}{2}u^2/(2\phi) + \tfrac{1}{2}(z-\mu)^2/\psi]$$

With $\sqrt{(\psi/\phi)} = 100$, $u = 2\times\sqrt{(2\phi)}$ and $z = \mu$, we get $B = \left(100/\sqrt{2}\right)\exp(-2) = 9.57$, although 2 standard deviation is beyond 1.96, the 5% level.

13. $p_1(\overline{x}) = \int \rho_1(\theta)p(\overline{x}|\theta)\,\mathrm{d}\theta$ which is $1/\tau$ times the integral between $\mu+\tau/2$ and $\mu-\tau/2$ of an $\mathrm{N}(\theta, \phi/n)$ density and so as $\tau \gg \phi/n$ is nearly the whole integral. Hence $p_1(\overline{x}) \cong 1/\tau$, from which it easily follows that $B = (2\pi\phi/n\tau^2)^{-\frac{1}{2}}\exp(-\frac{1}{2}z^2)$. In Section 4.5 we found $B = (1 + n\psi/\phi)^{\frac{1}{2}}\exp[-\frac{1}{2}z^2(1+\phi/n\psi)^{-1}]$, which for large n is about $B = (\phi/n\psi)^{-\frac{1}{2}}\exp(-\frac{1}{2}z^2)$. This agrees with the first form found here if $\tau^2 = 2\pi\phi$. As the variance of a uniform distribution on $(\mu-\tau/2, \mu+\tau/2)$ is $\tau^2/12$, this may be better expressed as $\tau^2/12 = (\pi/6)\phi = 0.52\phi$.

14. Jeffreys considers the case where both the mean θ and the variance ϕ are unknown, and wishes to test $\mathrm{H}_0 : \theta = 0$ versus $\mathrm{H}_1 : \theta \neq 0$ using the conventional choice of prior odds $\pi_0/\pi_1 = 1$, so that $B = p_0/p_1$. Under both hypotheses he uses the standard conventional prior for ϕ, namely $p(\phi) \propto 1/\phi$. He assumes that the prior for θ is dependent on $\sigma = \sqrt{\phi}$ as a scale factor. Thus if $\gamma = \theta/\sigma$ he assumes that $\pi(\gamma,\theta) = p(\gamma)p(\theta)$ so that γ and θ are independent. He then assumes that

(i) if the sample size $n = 1$ then $B = 1$, and

(ii) if the sample size $n \geqslant 2$ and $S = \sum(X_i - \overline{X})^2 = 0$, then $B = 0$, that is, $p_1 = 1$.

From (i) he deduces that $p(\gamma)$ must be an even function with integral 1, while he shows that (ii) is equivalent to the condition

$$\int_0^\infty p(\gamma)\gamma^{n-2}\,\mathrm{d}\gamma = \infty.$$

He then shows that the simplest function satisfying these conditions is $p(\gamma) = \pi^{-1}(1+\gamma^2)^{-1}$ from which it follows that $p(\theta) = p(\gamma)|\mathrm{d}\gamma/\mathrm{d}\theta| = \pi^{-1}\sigma(\sigma^2+\theta^2)^{-1}$. Putting $\sigma = \sqrt{\phi}$ and generalizing to the case where H_0 is that $\theta = \theta_0$ we get the distribution in the question.

There is some arbitrariness in Jeffreys' "simplest function"; if instead he had taken $p(\gamma) = \pi^{-1}\kappa(\kappa^2+\gamma^2)^{-1}$ he would have ended up with $p(\theta) = \pi^{-1}\tau(\tau^2 + (\theta-\theta_0)^2)^{-1}$ where $\tau = \kappa\sigma$. However, even after this generalization, the argument is not overwhelmingly convincing.

15. (a) Maximum likelihood estimator $\widehat{\theta}$ of θ is x/n, so

$$B \geqslant \left(\frac{\theta_0}{\widehat{\theta}}\right)^x \left(\frac{1-\theta_0}{1-\widehat{\theta}}\right)^{n-x}; \qquad p_0 \geqslant \left[1 + \frac{1-\pi_0}{\pi_0}\left(\frac{\widehat{\theta}}{\theta_0}\right)^x \left(\frac{1-\widehat{\theta}}{1-\theta_0}\right)^{n-x}\right]^{-1}.$$

(b) From tables (e.g. D. V. Lindley and W. F. Scott, *New Cambridge Elementary Statistical Tables*, Cambridge: University Press 1995 [1st edn (1984)], Table 1, or H. R. Neave, *Statistics Tables for mathematicians, engineers, economists and the behavioural and management sciences*, London: George Allen & Unwin (1978), Table 1.1) the probability of a binomial observation $\leqslant 14$ from $\mathrm{B}(20, 0.5)$ is 0.9793, so the appropriate (two-tailed) P-value is $2(1-0.9793) = 0.0414 \cong 1/24$. The maximum likelihood estimate is $\widehat{\theta} = 15/20 = 0.75$, so the lower bound on B is $(0.5/0.75)^{15}(0.5/0.25)^5 = 2^{20}/3^{15} = 0.0731$, implying a lower bound on p_0 of 0.0681 or just over 1/15.

16. (a) $n = 12$, $\nu = 11$, $t = \overline{x}/(s/\sqrt{n}) = 1.2/\sqrt{(1.1/12)} = 3.96$ and if we take $k = 1$ the Bayes factor B is

$$\frac{(1+t^2/\nu)^{-(\nu+1)/2}}{(1+nk)^{-\frac{1}{2}}(1+t^2(1+nk)^{-1}/\nu)^{-(\nu+1)/2}} = \frac{0.004910}{(0.2774)(0.5356)} = 0.033.$$

(b) $z = \overline{x}/(s/\sqrt{n}) = 1.2/\sqrt{12} = 4.16$ and (taking $\psi = \phi$ as usual)

$$\begin{aligned} B &= (1+n)^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}z^2(1+1/n)^{-1}\right] \\ &= (1+12)^{\frac{1}{2}} \exp\left[-\tfrac{1}{2}(4.16)^2(1+1/12)^{-1}\right] = 0.001. \end{aligned}$$

17. Two-tailed P-value is 0.0432 (cf. D. V. Lindley and W. F. Scott, *New Cambridge Elementary Statistical Tables*, Cambridge: University Press 1995 [1st edn (1984)], Table 9), while Bayes factor is

$$\frac{(1+t^2/\nu)^{-(\nu+1)/2}}{(1+nk)^{-\frac{1}{2}}(1+t^2(1+nk)^{-1}/\nu)^{-(\nu+1)/2}} = \frac{0.08712}{(0.3162)(0.7313)} = 0.377$$

so $F = 1/B = 2.65$. Range $(1/30P, 3/10P)$ is $(0.77, 6.94)$, so F *is* inside it and we do not need to "think again".

18. For P-value 0.1 think again if F not in $(\frac{1}{3}, 3)$. As $p_0 = [1+F]^{-1}$ this means if p_0 not in $(0.250, 0.750)$, so if $n = 1000$. Similarly if P-value 0.05, if p_0 not in $(0.143, 0.600)$, so if $n = 1000$ (and the case $n = 100$ is marginal); if P-value 0.01, if p_0 not in $(0.032, 0.231)$, so if $n = 100$ or $n = 1000$; if P-value 0.001, if p_0 not in $(0.003, 0.029)$, so if $n = 50$, $n = 100$ or $n = 1000$.

D.5 Exercises on Chapter 5

1. Mean difference $\overline{w} = 0.05\dot{3}$, $S = 0.0498$, $s = 0.0789$. Assuming a standard reference prior for θ and a variance known equal to 0.0789^2, the posterior distribution of the effect θ of Analyst A over Analyst B is $\mathrm{N}(0.05\dot{3}, 0.0789^2/9)$ leading to a 90% interval $0.05\dot{3} \pm 1.6449 \times 0.0789/3$, that is, $(0.010, 0.097)$. If the variance is not assumed known then the normal distribution should be replaced by t_8.

2. If variance assumed known, $z = 0.05\dot{3}/(0.0789/3) = 2.028$ so with $\psi = \phi$

$$\begin{aligned} B &= (1+n)^{\frac{1}{2}} \exp[-\tfrac{1}{2}z^2(1+1/n)^{-1}] \\ &= (1+9)^{\frac{1}{2}} \exp[-\tfrac{1}{2}(2.028)^2(1+\tfrac{1}{9})^{-1}] = 0.497 \end{aligned}$$

and $p_0 = \left[1 + 0.497^{-1}\right]^{-1} = 0.33$ with the usual assumptions. If the variance is not assumed known,

$$B = \frac{\{1 + 2.028^2/8\}^{-9/2}}{10^{-\frac{1}{2}}\{1 + 2.028^2 10^{-1}/8\}^{-9/2}} = \frac{0.1546}{(0.3162)(0.7980)} = 0.613$$

and $p_0 = \left[1 + 0.613^{-1}\right]^{-1} = 0.38$.

3. We know $\sum(x_i+y_i) \sim \mathrm{N}(2\theta, 2\sum\phi_i)$ and $\sum(x_i-y_i)$ is independently $\mathrm{N}(0, 2\sum\phi_i)$, so $\sum(x_i-y_i)^2/2\sum\phi_i \sim \chi_n^2$ and hence if $\theta = 0$ then

$$\sum(x_i+y_i)\Big/\sqrt{\sum(x_i-y_i)^2} \sim \mathrm{t}_n.$$

Hence test whether $\theta = 0$. If $\theta \neq 0$, it can be estimated as $\frac{1}{2}\sum(x_i+y_i)$.

4. In that case

$$\begin{aligned} B &= (1 + 440.5/99)^{\frac{1}{2}} \exp[-\tfrac{1}{2}(1.91)^2\{1 + 99/440.5\}^{-1}] \\ &= 5.4495^{\frac{1}{2}} \exp(-1.4893) = 0.53. \end{aligned}$$

If the prior probability of the null hypothesis is taken as $\pi_0 = \frac{1}{2}$, then this gives a posterior probability of $p_0 = (1 + 0.53^{-1})^{-1} = 0.35$.

5. $m = 9$, $n = 12$, $\overline{x} = 12.42$, $\overline{y} = 12.27$, $s_x = 0.1054$, $s_y = 0.0989$.

(a) Posterior of δ is $\mathrm{N}(12.42 - 12.27, 0.1^2(1/9 + 1/12))$ so 90% interval is $0.15 \pm 1.6449 \times \sqrt{0.00194}$, that is, $(0.077, 0.223)$.

(b) $S = 8 \times 0.1054^2 + 11 \times 0.0989^2 = 0.1965$, $s = \sqrt{0.1965/19} = 0.102$, so $s\sqrt{(m^{-1}+n^{-1})} = 0.045$, so from tables of t_{19} a 90% interval is $0.15 \pm 1.729 \times 0.045$, that is $(0.072, 0.228)$.

6. With independent priors uniform in λ, μ and $\log\phi$, that is, $p(\lambda,\mu,\phi)\propto 1/\phi$,

$$\begin{aligned}
p(\lambda,\mu,\phi\,|\,\boldsymbol{x},\boldsymbol{y}) &\propto p(\lambda,\mu,\phi)\,p(\boldsymbol{x}|\lambda,\phi)\,p(\boldsymbol{y}|\mu,\phi)\\
&\propto (1/\phi)(2\pi\phi)^{-(m+n)/2}\exp\left[-\tfrac{1}{2}\left\{\sum(x_i-\lambda)^2+\tfrac{1}{2}\sum(y_i-\mu)^2\right\}/\phi\right]\\
&\propto \phi^{-(m+n)/2-1}\exp[-\tfrac{1}{2}\{S_x+m(\overline{x}-\lambda)^2+\tfrac{1}{2}S_y+\tfrac{1}{2}n(\overline{y}-\mu)^2\}/\phi]
\end{aligned}$$

Writing $S=S_x+\frac{1}{2}S_y$ and $\nu=m+n-2$ we get

$$\begin{aligned}
p(\lambda,\mu,\phi|\boldsymbol{x},\boldsymbol{y}) &\propto \phi^{-\nu/2-1}\exp[-\tfrac{1}{2}S/\phi]\ (2\pi\phi/m)^{-\frac{1}{2}}\exp[-\tfrac{1}{2}m(\lambda-\overline{x})^2)/2\phi]\\
&\qquad\times(2\pi2\phi/m)^{-\frac{1}{2}}\exp[-\tfrac{1}{2}m(\mu-\overline{y})^2)/2\phi]\\
&\propto p(\phi|S)\,p(\lambda|\phi,\overline{x})\,p(\mu|\phi,\overline{y})
\end{aligned}$$

where

$p(\phi\,\|\,S)$	is an $S\chi_\nu^{-2}$ density,
$p(\lambda\,\|\,\phi,\overline{x})$	is an $\mathrm{N}(\overline{x},\phi/m)$ density,
$p(\mu\,\|\,\phi,\overline{x})$	is an $\mathrm{N}(\overline{y},2\phi/m)$ density,

It follows that, for given ϕ, the parameters λ and μ have independent normal distributions, and so that the joint density of $\delta=\lambda-\mu$ and ϕ is

$$p(\delta,\phi|\boldsymbol{x},\boldsymbol{y})=p(\phi|S)\,p(\delta|\overline{x}-\overline{y},\phi)$$

where $p(\delta\,|\,\overline{x}-\overline{y},\phi)$ is an $\mathrm{N}(\overline{x}-\overline{y},\ \phi(m^{-1}+2n^{-1})$ density. This variance can now be integrated out just as in Sections 2.12 and 5.2, giving a very similar conclusion, that is, that if

$$t=\frac{\delta-(\overline{x}-\overline{y})}{s\{m^{-1}+2n^{-1}\}^{\frac{1}{2}}}$$

where $s^2=S/\nu$, then $t\sim\mathrm{t}_\nu$.

7. We find that

$$S_1=S_0+S_x+S_y+m_0\lambda_0^2+m\overline{x}^2+n_0\mu_0^2+n\overline{y}^2-m_1\lambda_1^2-n_1\mu_1^2$$

and then proceed as in Exercise 18 on Chapter 2 to show that

$$S_1=S_0+S_x+S_y+(m_0^{-1}+m^{-1})^{-1}(\overline{x}-\lambda_0)^2+(n_0^{-1}+n^{-1})^{-1}(\overline{y}-\mu_0)^2.$$

8. $m=10$, $n=9$, $\overline{x}=22.2$, $\overline{y}=23.1$, $s_x=1.253$, $s_y=0.650$. Consequently $\sqrt{(s_x^2/m+s_y^2/n)}=0.452$ and $\tan\theta==(s_y/\sqrt{n})/(s_x/\sqrt{m})=0.547$, so $\theta\cong 30°$. Interpolating in Table B.1 with $\nu_1=8$ and $\nu_2=9$ the 90% point of the Behrens' distribution is 1.42, so a 90% HDR is $22.2-23.1\pm1.42\times0.452$, that is, $(-1.54,-0.26)$.

9. Evidently $f_1=(m-1)/(m-3)$ and

$$f_2=\frac{(m-1)^2}{(m-3)^2(m-5)}(\sin^4\theta+\cos^4\theta).$$

Also $1 = (\sin^2\theta + \cos^2\theta)^2 = \sin^4\theta + \cos^4\theta + 2\sin^2\theta\cos^2\theta$, so that $\sin^4\theta + \cos^4\theta = 1 - \sin^2 2\theta = \cos^2 2\theta$. The result follows.

10. As $T_x \sim \mathrm{t}_{\nu(x)}$ and $T_y \sim \mathrm{t}_{\nu(y)}$ we have

$$p(T_x, T_y | \boldsymbol{x}, \boldsymbol{y}) \propto (1 + T_x^2/\nu(x))^{-(\nu(x)+1)/2}(1 + T_y^2/\nu(y))^{-(\nu(y)+1)/2}$$

Jacobian is trivial, and the result follows (cf. Appendix A.18).

11. Set $y = \nu_1 x/(\nu_2 + \nu_1 x)$ so that $1 - y = 1/(\nu_2 + \nu_1 x)$ and $\mathrm{d}y/\mathrm{d}x = -\nu_1/(\nu_2 + \nu_1 x)^2$. Then using Appendix A.19 for the density of x

$$p(y) = p(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| \propto \frac{x^{\nu_1/2-1}}{(\nu_2 + \nu_1 x)^{(\nu_1+\nu_2)/2-2}} \propto y^{\nu_1/2-1}(1-y)^{\nu_2/2-1}$$

from which the result follows.

12. $k = s_1^2/s_2^2 = 6.77$, so $\eta^{-1} = k/\kappa = 6.77/\kappa \sim \mathrm{F}_{24,14}$. A 95% interval for $\mathrm{F}_{24,14}$ from Table B.5 is $(0.40, 2.73)$ so one for κ is $(2.5, 16.9)$.

13. $k = 3$ with $\nu_x = \nu_y = 9$. A an interval corresponding to a 99% HDR for $\log\mathrm{F}$ for $\mathrm{F}_{9,9}$ is $(0.15, 6.54)$, so a 99% interval for $\sqrt{\kappa}$ is from $\sqrt{3 \times 0.15}$ to $\sqrt{3 \times 6.54}$, that is, $(0.67, 4.43)$.

14. Take $\alpha_0 + \beta_0 = 6$, $\alpha_0/(\alpha_0 + \beta_0) = \frac{1}{2}$, $\gamma_0 + \delta_0 = 6$, $\gamma_0/(\gamma_0 + \delta_0) = \frac{2}{3}$, so $\alpha_0 = 3$, $\beta_0 = 3$, $\gamma_0 = 4$, $\delta_0 = 2$ and hence $\alpha = 3 + a = 11$, $\beta = 3 + b = 15$, $\gamma = 4 + c = 52$, $\delta = 2 + d = 64$ so that

$$\log\{(\alpha - \tfrac{1}{2})(\delta - \tfrac{1}{2})/(\beta - \tfrac{1}{2})(\gamma - \tfrac{1}{2})\} = \log 0.983 = -0.113$$

while $\alpha^{-1} + \beta^{-1} + \gamma^{-1} + \delta^{-1} = 0.192$. Hence posterior of log-odds ratio is $\Lambda - \Lambda' \sim \mathrm{N}(-0.113, 0.192)$. The posterior probability that $\pi > \rho$ is

$$\Phi(-0.113/\sqrt{0.192}) = \Phi(-0.257) = 0.3986.$$

15. Cross-ratio is $(45 \times 29)/(28 \times 30) = 1.554$ so its logarithm is 0.441. More accurately, adjusted value is $(44.5 \times 28.5)/(27.5 \times 29.5) = 1.563$ so its logarithm is 0.446. Value of $a^{-1} + b^{-1} + c^{-1} + d^{-1}$ is 0.126, and so the posterior distribution of the log-odds ratio is $\Lambda - \Lambda' \sim \mathrm{N}(0.441, 0.126)$ or more accurately $\mathrm{N}(0.446, 0.126)$. The posterior probability that the log-odds ratio is positive is $\Phi(0.441/\sqrt{0.126}) = \Phi(1.242) = 0.8929$ or more accurately $\Phi(0.446/\sqrt{0.126}) = \Phi(1.256) = 0.8955$. With the same data $\sin^{-1}\sqrt{(45/73)} = 0.903$ radians and $\sin^{-1}\sqrt{(30/59)} = 0.794$ radians, and $1/4m + 1/4n = 0.00766$, so the posterior probability that $\pi > \rho$ is $\Phi((0.903 - 0.794)/\sqrt{0.00766}) = \Phi(1.245) = 0.8934$.

16. $\Lambda - \Lambda' = \log(\pi/\rho) - \log\{(1-\pi)/(1-\rho)\}$ so if $\pi - \rho = \alpha$ we get

$$\begin{aligned}\Lambda - \Lambda' &= \log\{\pi/(\pi-\alpha)\} - \log\{(1-\pi)/(1-\pi+\alpha)\}\\ &= -\log\{1-\alpha)/\pi\} + \log\{1+\alpha/(1-\pi)\}\\ &\cong \alpha/\pi + \alpha/(1-\pi) = \alpha/\{\pi(1-\pi)\}.\end{aligned}$$

17. With conventional priors, posteriors are near $\pi \sim \mathrm{Be}(56, 252)$ and $\rho \sim \mathrm{Be}(34, 212)$, so approximating by normals of the same means and variances $\pi \sim \mathrm{N}(0.1818, 0.0004814)$, $\rho \sim \mathrm{N}(0.1382, 0.0004822)$, so $\pi - \rho \sim \mathrm{N}(0.0436, 0.000963)$ so $\mathsf{P}(\pi - \rho > 0.01) = \Phi((0.0436 - 0.01)/\sqrt{0.000963}) = \Phi(1.083) = 0.8606$ and so the posterior odds are $0.8606/(1 - 0.8606) = 6.174$.

18. Using a normal approximation, $x \sim \mathrm{N}(8.5, 8.5)$ and $y \sim \mathrm{N}(11.0, 11.0)$, so that $x - y \sim \mathrm{N}(-2.5, 19.5)$.

D.6 Exercises on Chapter 6

1. Straightforward substitution gives

Sample	1	2	3	4	5	Total
n	12	45	23	19	30	129
r	0.631	0.712	0.445	0.696	0.535	
$\tanh^{-1} z$	0.743	0.891	0.478	0.860	0.597	
$n \tanh^{-1} z$	8.916	40.095	10.994	16.340	17.910	94.255

Posterior for ζ is $\mathrm{N}(94.255/129, 1/129)$ and 95% HDR is $0.7307 \pm 1.96 \times 0.0880$, that is, $(0.5582, 0.9032)$. A corresponding interval for ρ is $(0.507, 0.718)$.

2. Another straightforward substitution gives

n	45	34	49
$1/n$	0.0222	0.0294	0.0204
r	0.489	0.545	0.601
$\tanh^{-1} z$	0.535	0.611	0.695

Hence $\zeta_1 - \zeta_2 \sim \mathrm{N}(0.535 - 0.611,\ 0.0222 + 0.0294)$, that is, $\mathrm{N}(-0.076, 0.227^2)$. Similarly $\zeta_2 - \zeta_3 \sim \mathrm{N}(-0.084, 0.223^2)$ and $\zeta_3 - \zeta_1 \sim \mathrm{N}(0.160, 0.206^2)$. It follows without detailed examination that there is no evidence of difference.

3. $\zeta \sim \mathrm{N}\left(\sum n_i \tanh^{-1} r_i)/\sum n_i\,,\ 1/\sum n_i\right)$ so required interval is

$$\frac{\sum n_i \tanh^{-1} r_i}{\sum n_i} \pm \frac{1.96}{\sqrt{\sum n_i}}.$$

4. We found in Section 6.1 that

$$p(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}) \propto p(\rho) \frac{(1-\rho^2)^{(n-1)/2}}{(1-\rho r)^{n-(3/2)}}$$

As $p(\rho)(1-\rho^2)^{-\frac{1}{2}}(1-\rho r)^{3/2}$ does not depend on n, for large n

$$p(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}) \propto \frac{(1-\rho^2)^{n/2}}{(1-\rho r)^{n}}$$

so that $\log l(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}) = c + \frac{1}{2} n \log(1-\rho^2) - n\log(1-\rho r)$, and hence

$$(\partial/\partial\rho) \log l(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}) = -n\rho(1-\rho^2)^{-1} + nr(1-\rho r)^{-1}.$$

implying that the maximum likelihood estimator $\widehat{\rho} \cong r$. Further

$$(\partial^2/\partial\rho^2) \log l(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}) = -n(1-\rho^2)^{-1} - 2n\rho^2(1-\rho^2)^{-2} + nr^2(1-\rho r)^{-2},$$

so that if $\rho = r$ we have $(\partial^2/\partial\rho^2) \log l(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}) = -n(1-\rho^2)^{-2}$. This implies that the information should be about $I(\rho \,|\, \boldsymbol{x}, \boldsymbol{y}) = n(1-\rho^2)^{-2}$, and so leads to a prior $p(\rho) \propto (1-\rho^2)^{-1}$.

5. We have

$$p(\rho|\boldsymbol{x}, \boldsymbol{y}) \propto p(\rho)(1-\rho^2)^{(n-1)/2} \int_0^\infty (\cosh t - \rho r)^{-(n-1)}\,\mathrm{d}t$$

Now write $-\rho r = \cos\theta$ so that

$$p(\rho|\boldsymbol{x}, \boldsymbol{y}) \propto p(\rho)(1-\rho^2)^{(n-1)/2} \int_0^\infty (\cosh t + \cos\theta)^{-(n-1)}\,\mathrm{d}t$$

and set

$$I_k = \int_0^\infty (\cosh t + \cos\theta)^{-k}\,\mathrm{d}t$$

We know that $I_1 = \theta/\sin\theta$ (cf. J. A. Edwards, *A Treatise on the Integral Calcuous* (2 vols), London: Macmillan (1921) [reprinted New York: Chelsea (1955)], art. 180). Moreover

$$(\partial/\partial\theta)(\cosh t + \cos\theta)^{-k} = k\sin\theta(\cosh t + \cos\theta)^{-(k+1)}$$

and by induction

$$(\partial/\sin\theta\partial\theta)^k(\cosh t + \cos\theta)^{-1} = k!(\cosh t + \cos\theta)^{-(k+1)}$$

Differentiating under the integral sign, we conclude that

$$(\partial/\sin\theta\partial\theta)^k I_1 = k! I_{k+1} \qquad (k \geqslant 0).$$

Taking $k = n - 2$, or $k + 1 = n - 1$, we get

$$I_k = \int_0^\infty (\cosh t + \cos\theta)^{-k}\,\mathrm{d}t \propto (\partial/\sin\theta\partial\theta)^k(\theta/\sin\theta).$$

(ignoring the factorial). Consequently

$$p(\rho|\boldsymbol{x},\boldsymbol{y}) \propto p(\rho)(1-\rho^2)^{(n-1)/2}(\partial/\sin\theta\partial\theta)^k(\theta/\sin\theta).$$

Since $\mathrm{d}(r\rho)/\mathrm{d}\theta = \mathrm{d}(-\cos\theta)/\mathrm{d}\theta = \sin\theta$ and so $\partial/\sin\theta\partial\theta = \mathrm{d}/\mathrm{d}(r\rho)$, we could alternatively write

$$p(\rho|\boldsymbol{x},\boldsymbol{y}) \propto p(\rho)(1-\rho^2)^{(n-1)/2}\frac{\mathrm{d}^{n-2}}{\mathrm{d}(\rho r)^{n-2}}\left(\frac{\text{arc }\cos(-\rho r)}{(1-\rho^2 r^2)^{\frac{1}{2}}}\right)$$

6. Supposing that the prior is

$$p(\rho) \propto (1-\rho^2)^{k/2}$$

and $r = 0$ then

$$p(\rho|\boldsymbol{x},\boldsymbol{y}) \propto p(\rho)(1-\rho^2)^{(n-1)/2}$$

so with the prior as in the question we get

$$p(\rho|\boldsymbol{x},\boldsymbol{y}) \propto (1-\rho^2)^{(k+n-1)/2}$$

If we define

$$t = \sqrt{(k+n+1)}\frac{\rho}{\sqrt{(1-\rho^2)}}, \qquad \rho = \frac{t}{\sqrt{\{(k+n+1)+t^2\}}}$$

so that

$$1-\rho^2 = \frac{(k+n+1)}{(k+n+1)+t^2}, \qquad 2\rho\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{(k+n+1)2t}{\{(k+n+1)+t^2\}^2}$$
$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{(k+n+1)}{\{(k+n+1)+t^2\}^{3/2}}$$

and hence

$$\begin{aligned}
p(t|\boldsymbol{X},\boldsymbol{Y}) &\propto p(\rho|\boldsymbol{X},\boldsymbol{Y})\,\mathrm{d}\rho/\,\mathrm{d}t \\
&\propto \left[\frac{(k+n+1)}{(k+n-1)+t^2}\right]^{(k+n-1)/2}\frac{(k+n+1)}{\{(k+n+1)+t^2\}^{3/2}} \\
&\propto \left[1+\frac{t^2}{k+n+1}\right]^{-(k+n+2)/2}
\end{aligned}$$

This can be recognized as Student's t distribution on $k+n+1$ degrees of freedom (see Section 2.12).

7. See G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley (1973, Section 8.5.4—the equation given in the question is implicit in their equation (8.5.49)).

8. See G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Reading, MA: Addison-Wesley (1973, Section 8.5.4, and in particular equation (8.5.43)).

9. For a bivariate normal distribution

$$\begin{aligned}\log l(\alpha,\beta,\gamma|x,y) = &-\log 2\pi + \tfrac{1}{2}\log\delta \\ &- \tfrac{1}{2}\alpha(x-\lambda)^2 - \gamma(x-\lambda)(y-\mu) - \tfrac{1}{2}\beta(y-\mu)^2\end{aligned}$$

where $\delta = \alpha\beta - \gamma^2 = 1/\Delta$. Then

$$\begin{aligned}(\partial/\partial\alpha)\log l &= \tfrac{1}{2}\beta/\delta - \tfrac{1}{2}(x-\lambda)^2 \\ (\partial/\partial\beta)\log l &== \tfrac{1}{2}\alpha/\delta - \tfrac{1}{2}(y-\mu)^2, \\ (\partial/\partial\gamma)\log l &= -\gamma/\delta - (x-\lambda)(y-\mu),\end{aligned}$$

Consequently the information matrix, i.e. minus the matrix of second derivatives (taking expectations is trivial as all elements are constant) is

$$\begin{aligned}I &= \begin{pmatrix} \tfrac{1}{2}\beta^2\delta^{-2} & -\tfrac{1}{2}\delta^{-1} + \tfrac{1}{2}\alpha\beta\delta^{-2} & -\beta\gamma\delta^{-2} \\ -\tfrac{1}{2}\delta^{-1} + \tfrac{1}{2}\alpha\beta\delta^{-2} & \tfrac{1}{2}\alpha^2\delta^{-2} & -\alpha\gamma\delta^{-2} \\ -\beta\gamma\delta^{-2} & -\alpha\gamma\delta^{-2} & \delta^{-1} + 2\gamma^2\delta^{-2} \end{pmatrix} \\ &= \frac{1}{2\delta^2}\begin{pmatrix} \beta^2 & \gamma^2 & -2\beta\gamma \\ \gamma^2 & \alpha^2 & -2\alpha\gamma \\ -2\beta\gamma & -2\alpha\gamma & 2(\alpha\beta+\gamma^2) \end{pmatrix}\end{aligned}$$

so that its determinant is

$$\det I = \frac{1}{8\delta^6}\begin{vmatrix} \beta^2 & \gamma^2 & -2\beta\gamma \\ \gamma^2 & \alpha^2 & -2\alpha\gamma \\ -2\beta\gamma & -2\alpha\gamma & -2(\alpha\beta+\gamma^2) \end{vmatrix}$$

Adding $\frac{1}{2}\beta/\gamma$ times the last column to the first and $\frac{1}{2}\gamma/\beta$ times the last column to the second we get

$$\begin{aligned}\det I &= \frac{1}{8\delta^6}\begin{pmatrix} 0 & 0 & -2\beta\gamma \\ -\delta & \alpha\beta^{-1}\delta & -2\alpha\gamma \\ \beta\gamma^{-1}\delta & -\beta^{-1}\gamma\delta & 2(\alpha\beta+\gamma^2) \end{pmatrix} \\ &= \frac{1}{8\delta^6}(-2\beta\gamma)\left(-\frac{\alpha}{\gamma}\delta^2 + \frac{\gamma}{\beta}\delta^2\right) = \frac{1}{4\delta^3}\end{aligned}$$

We thus conclude that $\det I = 1/4\delta^3$ implying a reference prior

$$p(\alpha,\beta,\gamma \propto \delta^{-3/2}.$$

Rather similarly we get

$$\frac{\partial(\alpha,\beta,\gamma)}{\partial(\phi,\psi,\kappa)} = \left|\begin{array}{ccc} -\psi^2/\Delta^2 & 1/\Delta-\phi\psi/\Delta^2 & 2\psi\kappa/\Delta^2 \\ 1/\Delta-\phi\psi/\Delta^2 & -\phi^2/\Delta^2 & 2\phi\kappa/\Delta^2 \\ \psi\kappa/\Delta^2 & \phi\kappa/\Delta^2 & -1/\Delta-2\kappa^2/\Delta^2 \end{array}\right|$$
$$= \frac{1}{\Delta^6}\left|\begin{array}{ccc} \psi^2 & \kappa^2 & -2\psi\kappa \\ \kappa^2 & \phi^2 & -2\phi\kappa \\ -\psi\kappa & -\phi\kappa & \phi\psi+\kappa^2 \end{array}\right|$$

and evaluate this as $-1/\Delta^3$. It then follows that

$$p(\phi,\psi,\kappa) \propto |-1/\Delta^3|\, p(\alpha,\beta,\gamma) \propto \Delta^{-3/2}.$$

10. Age is y, weight is x. $n=12$, $\overline{x}=2911$, $\overline{y}=38.75$, $S_{xx}=865,127$, $S_{yy}=36.25$, $S_{xy}=4727$. Hence $a=\overline{y}=38.75$, $b=S_{xy}/S_{xx}=0.005464$, $r=S_{xy}/\sqrt{(S_{xx}S_{yy})}=0.8441$, $S_{ee}=S_{yy}-S_{xy}^2/S_{xx}=S_{yy}(1-r^2)=10.422$, and $s^2=S_{ee}/(n-2)=1.0422$. Take $x_0=3000$ and get $a+b(x_0-\overline{x})=39.24$ as mean age for weight 3000. For a particular baby, note that the 95% point of t_{10} is 1.812 and $s\sqrt{\{1+n^{-1}+(x_0-\overline{x})^2/S_{xx}\}}=1.067$ so a 90% interval is $39.24\pm1.812\times1.067$, that is, $(37.31, 41.17)$. For the mean weight of all such babies note $s\sqrt{\{n^{-1}+(x_0-\overline{x})^2/S_{xx}\}}=0.310$, so interval is $(38.68, 39.80)$.

11. From the formulae in the text

$$S_{ee}=S_{yy}-S_{xy}^2/S_{xx}=S_{yy}-2S_{xy}^2/S_{xx}+S_{xy}^2/S_{xx}=S_{yy}-2bS_{xy}+b^2S_{xx}$$

where $b=S_{xy}/S_{xx}$. Hence

$$\begin{aligned} S_{ee} &= \sum(y_i-\overline{y})^2-2b\sum(y_i-\overline{y})(x_i-\overline{x})+b^2\sum(x_i-\overline{x})^2 \\ &= \sum\{(y_i-\overline{y})-b(x_i-\overline{x})\}^2. \end{aligned}$$

The result now follows as $y_i=a$.

12. Clearly (as stated in the hint) $\sum u_i=\sum v_i=\sum u_iv_i=0$, hence $\overline{u}=\overline{v}=0$ and $S_{uv}=0$. We now proceed on the lines of Section 6.3, redefining S_{ee} as $S_{yy}-Suy^2/S_{uu}-S_{vy}^2/S_{vv}$ and noting that

$$\begin{aligned} &\sum(y_i-\alpha-\beta u_i-\gamma v_i)^2=\sum\left\{(y_i-\overline{y})+(\overline{y}-\alpha)-\beta u_i-\gamma v_i\right\}^2 \\ &=S_{yy}+n(y_i-\overline{y})^2+\beta^2S_{uu}+\gamma^2S_{vv}-2\beta S_{uy}-2\gamma S_{vy} \\ &=S_{yy}-S_{uy}^2/S_{uu}-S_{vy}^2/S_{vv}+n(y_i-\overline{y})^2 \\ &\qquad+S_{uu}(\beta-S_{uy}/S_{uu})^2+S_{vv}(\gamma-S_{vy}/S_{vv})^2 \\ &=S_{ee}+n(y_i-\overline{y})^2+S_{uu}(\beta-S_{uy}/S_{uu})^2+S_{vv}(\gamma-S_{vy}/S_{vv})^2 \end{aligned}$$

We consequently get a density $p(\alpha,\beta,\gamma,\phi\,|\,\boldsymbol{x},\boldsymbol{y})$ from which we integrate out *both* β and γ to get

$$p(\alpha,\phi\,|\,\boldsymbol{x},\boldsymbol{y}) \propto \phi^{-n/2}\exp[-\tfrac{1}{2}\{S_{ee}+n(\alpha-a)^2\}/\phi].$$

The result now follows.

13. $I = 4$, $\sum K_i = 28$, $G = 1779$, $\sum\sum x_{ik}^2 = 114,569$, $C = G^2/N = 113,030$. Further, $S_T = 1539$, $S_t = (426^2+461^2+450^2+442^2)/7 - C = 93$ and hence the analysis of variance table is as follows:

ANOVA Table

Source	Sum of squares	Degrees of freedom	Mean square	Ratio
Treatments	93	6	31	(<1)
Error	1446	24	60	
TOTAL	1539	27		

We conclude that the results from the four samples agree.

14. $d = \theta_2+\theta_4+\theta_6-\theta_3-\theta_5-\theta_7$ with $\widehat{d} = -1.4$ and $K_d = (6/4)^{-1}$ so that $0.82(d+1.4)/2.74 \sim \mathrm{t}_{25}$.

15. The analysis of variance table is as follows:

ANOVA Table

Source	Sum of squares	Degrees of freedom	Mean square	Ratio
Treatments	49,884	2	24,942	13.3
Blocks	149,700	5	29,940	16.0
Error	18,725	10	1,872	
TOTAL	218,309	17		

We conclude that the treatments differ significantly (an $\mathrm{F}_{2,10}$ variable exceeds 9.43 with probability 0.5%).

16. This is probably seen most clearly by example. When $r = t = 2$ we take

$$\boldsymbol{x} = \begin{pmatrix} x_{111} \\ x_{112} \\ x_{121} \\ x_{122} \\ x_{211} \\ x_{212} \\ x_{221} \\ x_{222} \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \\ \kappa_{12} \end{pmatrix}.$$

17. Evidently $\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I}$ from which (a) and (b) and (d) are immediate. For (c), note $\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ which is clearly symmetric.

18. We take

$$\boldsymbol{A} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ & \ddots \\ 1 & x_n \end{pmatrix}, \qquad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

so that writing $S = \sum(x_i - \overline{x})^2$ we see that $\det(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = nS$ and hence

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1} = \frac{1}{nS}\begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}, \qquad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = \begin{pmatrix} \sum y_i \\ \sum x_i y_y \end{pmatrix}$$

from which $\widehat{\boldsymbol{\eta}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$ is easily found. In particular

$$\widehat{\eta}_1 = \frac{1}{nS}\left(-\sum x_i \sum y_i + n\sum x_i y_i\right) = \frac{\sum(y_i - \overline{y})(x_i - \overline{x})}{\sum(x_i - \overline{x})^2}.$$

D.7 Exercises on Chapter 7

1. Recall that t is sufficient for θ given x if $p(x|\theta) = f(t,\theta)g(x)$ (see Section 2.9). Now

$$p(t|\theta) = \begin{cases} p(x|\theta) + p(y|\theta) = p(z|\theta) & \text{if } t = t(z) \\ p(x|\theta) & \text{if } t = t(x) \text{ for } x \neq y \end{cases}$$

so that

$$p(t|\theta) = \begin{cases} 0 & \text{if } x = y \\ p(x|\theta) & \text{if } x = t \end{cases}$$

Setting $f(t,\theta) = p(t|\theta)$ and $g(x) = 1$ $(x \neq y)$, $g(y) = 0$, it follows that t is sufficient for x given θ. It then follows from a naïve appliaction of the weak sufficiency principle that $\mathrm{Ev}(E, y, \theta) = \mathrm{Ev}(E, z, \theta)$. However if $\widetilde{x}$ is a continuous random variable, then $p_{\widetilde{x}}(y|\theta) = 0$ for all y, so we may take y and z as *any* two possible values of $\widetilde{x}$.

2. $l(\theta \,|\, g(x)) = l(\theta \,|\, h(x))$ from which the result follows.

3. The likelihood function is easily seen to be

$$l(\theta|x) = \frac{1}{3} \qquad \text{for} \qquad \theta = \begin{cases} x/2,\ 2x,\ 2x+1 & \text{when } x \text{ is even} \\ (x-1)/2,\ 2x,\ 2x+1 & \text{when } x \neq 1 \text{ is odd} \\ 1,\ 2,\ 3 & \text{when } x = 1. \end{cases}$$

The estimators d_1, d_2 and d_3 corresponding to the smallest, middle and largest possible θ are

$$d_1(x) = \begin{cases} x/2 & \text{when } x \text{ is even} \\ (x-1)/2 & \text{when } x \neq 1 \text{ is odd} \\ 1 & \text{when } x = 1, \end{cases}$$

and $d_2(x) = 2x$, $d_3(x) = 2x + 1$. The values of $p(d_1 = \theta)$ given in the question now follow. However, a Bayesian analyis leads to a posterior

$$p(\theta|x) = \frac{l(\theta|x)\pi(\theta)}{p(x)} = \frac{\pi(\theta)I_A(\theta)}{\pi(d_1(x)) + \pi(d_2(x)) + \pi(d_3(x))}$$

where $\pi(\theta)$ is the prior, $A = \{d_1(x), d_2(x), d_3(x)\}$ and $I_A(\theta)$ is the indicator function of A. Thus, indeed, the data conveys nothing to the Bayesian version except that θ is $d_1(x)$, $d_2(x)$, or $d_3(x)$. However, Bayesians are indifferent between $d_1(x)$, $d_2(x)$, or $d_3(x)$ only if they have equal prior probabilities, which cannot hold for all x if the prior is proper. For a discussion, see J. O. Berger and R. L. Wolpert, *The Likelihood Principle*, Hayward, CA: Institute of Mathematical Statistics (1984 and 1988, Example 34).

4. Computation of the posterior is straightforward. For a discussion, see J. O. Berger and R. L. Wolpert *The Likelihood Principle*, Hayward, CA: Institute of Mathematical Statistics (1984 and 1988, Example 35).

5. Rules (a) and (b) *are* stopping times and rule (c) is *not*.

6. $n = 4, \overline{x} = 1.25$, $z = 1.25\sqrt{4} = 2.50$.

(a) Reject at the 5% level (and indeed possibly do so on the basis of the first observation alone).

(a) Since $B = (1 + n\psi/\phi)^{\frac{1}{2}} \exp[-\frac{1}{2}z^2(1 + \phi/n\psi)^{-1}] = 1.07$ with $\psi = \phi$, so with $\pi_0 = \frac{1}{2}$ we get $p_0 = (1 + B^{-1})^{-1} = 0.52$. Null hypothesis still more probable than not.

7. As we do *not* stop the first four times but *do* the fifth time

$$\begin{aligned} p(\boldsymbol{x}|\lambda) &= \frac{\lambda^{3+1+2+5+7}}{3!\,1!\,5!\,2!\,7!}\exp(-5\lambda)\frac{3}{3}\frac{1}{4}\frac{2}{6}\frac{5}{11}\frac{11}{18} \\ l(\lambda|\boldsymbol{x}) &\propto \lambda^{10}\exp(-5\lambda). \end{aligned}$$

8. $\mathsf{E}S = \mathsf{E}(S+1) - 1$ and after re-arrangement $\mathsf{E}(S+1)$ is $(s+1)(R''-2)/(r''-2)$ times a sum of probabilities for the beta-Pascal distribution with S replaced by $(S+1)$, with s replaced by $(s+1)$, and with r'' replaced by $(r''-1)$. As probabilities sum to unity, the result follows.

9. Up to a constant

$$L(\pi|x, y) = \log l(\pi|x, y) = (x + n)\log\pi + (n - x + y)\log(1 - \pi)$$

so $-(\partial^2 L(\pi)|x,y)/\partial\pi^2) = (x+n)/\pi^2 + (n-x+y)/(1-\pi)^2$. Because the expectations of x and y are $n\pi$ and $n(1-\pi)/\pi$ respectively, we get $I(\pi|x,y) = n(1+\pi)/\pi^2(1-\pi)$, so that Jeffreys' rule leads to

$$p(\pi|x,y) \propto (1+\pi)^{\frac{1}{2}}\pi^{-1}(1-\pi)^{-\frac{1}{2}}.$$

10. $\mathsf{E}u(x) = \sum u(x)p(x) = \sum u(x)2^{-x}$ suggesting you would enter provided $e < \mathsf{E}u(x)$. If $u(x) \propto x$ then $\mathsf{E}u(x) \propto \sum 2^x 2^{-x} = \infty$ resulting in the implausible proposition that you would pay an arbitrarily large entry fee $\pounds e$.

11. By differentiation of the log-likelihood $L(\pi|x) = x\log\theta + (n-x)\log(1-\theta)$ with respect to θ we see that x/n is the maximum likelihood estimator.

Because prior for θ is uniform, that is, Be$(1,1)$, posterior is Be$(x+1, n-x+1)$. The question deals with a particular case of weighted quadratic loss, so we find $d(x)$ as

$$\mathsf{E}^w(\theta|x) = \frac{\mathsf{E}((1-\theta)^{-1}|x)}{\mathsf{E}(\theta^{-1}(1-\theta)^{-1}\,|\,x)} = \frac{\mathrm{B}(x+1,\,n-x)}{\mathrm{B}(x+1,\,n-x+1)}\frac{\mathrm{B}(x+1,\,n-x+1)}{\mathrm{B}(x,\,n-x)} = \frac{x}{n}.$$

If $x=0$ or $x=n$ then the posterior loss $\rho(a,x)$ is infinite for all a because the integral diverges at $x=0$ or at $x=n$ so the minimum is not well-defined.

12. The minimum of $\mathsf{E}(\pi-\rho)^2 - 2a(\mathsf{E}(\pi-\rho)+a^2$ clearly occurs when $a = \mathsf{E}(\pi-\rho)$. But since the prior for π is uniform, that is, Be$(1,1)$, its posterior is Be$(x+1, n-x+1)$ and so its posterior mean is $(x+1)/(n+2)$; similarly for y. We conclude that the Bayes rule is

$$d(x,y) = (x-y)/(n+2).$$

13. Posterior mean (resulting from quadratic loss) is a weighted mean of the component means, so with the data in Section 2.10 is

$$\alpha'\frac{10}{10+20} + \beta'\frac{20}{10+20} = \frac{115}{129}\frac{10}{10+20} + \frac{14}{129}\frac{20}{10+20} = 0.370.$$

Posterior median (resulting from absolute error loss) can be found by computing the weighted mean of the distribution functions of the two beta distributions for various values and honing in. Result is 0.343. Posterior mode (resulting from zero-one loss) is not very meaningful in a bimodal case, and even in a case where the posterior is not actually bimodal it is not very useful.

14. If we take as loss function

$$L(\theta,a) = \begin{cases} u(a-\theta) & \text{if } a \leqslant \theta \\ v(\theta-a) & \text{if } a \geqslant \theta \end{cases}$$

Suppose $m(x)$ is a $v/(u+v)$ fractile, so that

$$\mathsf{P}\big(x \leqslant m(x)\big) \geqslant v/(u+v), \qquad \mathsf{P}\big(x \geqslant m(x)\big) \geqslant u/(u+v).$$

Suuppose further that $d(x)$ is any other rule and, for definiteness, that $d(x) > m(x)$ for some particular x (the proof is similar if $d(x) < m(x)$). Then

$$L(\theta, m(x)) - L(\theta, d(x)) = \begin{cases} u[m(x) - d(x)] & \text{if } \theta \leqslant m(x) \\ (u+v)\theta - [um(x) + vd(x)] & \text{if } m(x) < \theta < d(x) \\ v[d(x) - m(x)] & \text{if } \theta \geqslant d(x) \end{cases}$$

while for $m(x) < \theta < d(x)$

$$(u+v)\theta - [um(x) + vd(x)] < u[\theta - m(x)] < u[d(x) - m(x)]$$

so that

$$L(\theta, m(x)) - L(\theta, d(x)) \leqslant \begin{cases} u[m(x) - d(x)] & \text{if } \theta \leqslant m(x) \\ v[d(x) - m(x)] & \text{if } \theta > m(x) \end{cases}$$

and hence on taking expectations over θ

$$\begin{aligned} \rho(m(x), x) - \rho(d(x), x) &\leqslant \big\{u[m(x) - d(x)]\big\}\mathsf{P}(\theta \leqslant m(x) \,|\, x) \\ &\qquad + \big\{v[d(x) - m(x)]\big\}\mathsf{P}(\theta > m(x) \,|\, x) \\ &= \big\{d(x) - m(x)\big\}\big\{-u\mathsf{P}(\theta \leqslant m(x) \,|\, x) \\ &\qquad + v\mathsf{P}(\theta \geqslant m(x) \,|\, x)\big\} \\ &\leqslant \big\{d(x) - m(x)\big\}\big\{-uv/(u+v) + uv/(u+v)\big\} = 0 \end{aligned}$$

from which it follows that taking a $v/(u+v)$ fractile of the posterior distribution does indeed result in the appropriate Bayes rule for this loss function.

15. By integration by parts, if $\theta \sim \mathrm{Be}(2, k)$ then

$$\begin{aligned} \mathsf{P}(\theta < \alpha) &= \int_0^\alpha k(k+1)\theta(1-\theta)^k \,\mathrm{d}\theta \\ &= \big[-(k+1)(1-\theta)^k\theta\big]_0^\alpha + \int_0^\alpha (k+1)(1-\theta)^k \,\mathrm{d}\theta \\ &= [-(k+1)(1-\theta)^k\theta - (1-\theta)^{k+1} = 1 - (1-\theta)^k(1+k\alpha). \end{aligned}$$

In this case, the prior for θ is $\mathrm{Be}(2, 12)$ so that the prior probability of H_0, that is, that $\theta < 0.1$, is 0.379, while the posterior is $\mathrm{Be}(2, 18)$ so that the posterior probability that $\theta < 0.1$ is 0.580.

(a) With "0–1" loss we accept the hypothesis with the greater posterior probability, in this case H_1.

(b) The second suggested loss function is of the "0–K_i" form and the decision depends on the relative sizes of p_1 and $2p_0$. Again this results in a decision in facour of H_1.

16. Posterior expected losses are

$$\rho(a_0, x) = 10p_1, \qquad \rho(a_1, x) = 10p_0, \qquad \rho(a_2, x) = 3p_0 + 3p_1.$$

Choose a_0 if $0 \leqslant p_0 \leqslant 0.3$, choose a_1 if $0.3 \leqslant p_0 \leqslant 0.7$ and choose a_2 if $0.7 \leqslant p_0 \leqslant 1$.

17. Posterior variance is $(225^{-1} + 100^{-1})^{-1} = 69.23 = 8.32^2$ and the posterior mean is $69.23(100/225 + 115/100) = 110.38$. Posterior expected losses are

$$p(a_1, x) = \int_{90}^{110} (\theta - 90)\pi(\theta|x)\,\mathrm{d}\theta + \int_{110}^{\infty} 2(\theta - 90)\pi(\theta|x)\,\mathrm{d}\theta$$

Now note that (with $\phi(x)$ the $\mathrm{N}(0,1)$ density function

$$\begin{aligned}\int_{90}^{110} (\theta - 90)\pi(\theta|x)\,\mathrm{d}\theta &= \sqrt{69.23}\int_{-2.450}^{-0.046} z\phi(z)\,\mathrm{d}\theta + 20.38\int_{-2.450}^{-0.046} \phi(z)\,\mathrm{d}\theta \\ &= -\sqrt{(69.23/2\pi)}\{\exp(-\tfrac{1}{2}0.046^2) - \exp(-\tfrac{1}{2}2.450^2)\} \\ &\quad + 20.39\{\Phi(-0.046) - \Phi(-2.450)\} \\ &= -3.15 + 9.66 = 6.51.\end{aligned}$$

By similar manipulations we find that $\rho(a_1, x) = 34.3$, $\rho(a_2, x) = 3.6$ and $\rho(a_3, x) = 3.3$, and thus conclude that a_3 is the Bayes decision.

18. In the negative binomial case

$$\begin{aligned}p(\pi|x) &= p(\pi)p(x|\pi)/p_{\widetilde{x}}(x) \\ &= \binom{n+x-1}{x}(1-\pi)^n\pi^x/p_{\widetilde{x}}(x)\end{aligned}$$

It follows that the posterior mean is

$$\begin{aligned}\mathsf{E}(\pi|x) &= \int \pi\binom{n+x-1}{x}(1-\pi)^n\pi^x/p_{\widetilde{x}}(x) \\ &= \frac{x+1}{n+x}\binom{n+x}{x+1}(1-\pi)^n\pi^x/p_{\widetilde{x}}(x) \\ &= \frac{(x+1)}{(n+x)}\frac{p_{\widetilde{x}}(x+1)}{p_{\widetilde{x}}(x)}\end{aligned}$$

This leads in the same way as in Section 7.5 to the estimate

$$\delta_n = \frac{(\xi+1)f_n(\xi+1)}{(n+\xi)(f_n(\xi)+1)}.$$

D.8 Exercises on Chapter 8

1. Write $\gamma = \alpha/(\alpha+\beta)$ and $\delta = (\alpha+\beta)^{-1/2}$. The Jacobian is

$$\left|\begin{array}{cc} \partial\gamma/\partial\alpha & \partial\gamma/\partial\beta \\ \partial\delta/\partial\alpha & \delta\gamma/\partial\beta \end{array}\right| = \left|\begin{array}{cc} \beta/(\alpha+\beta)^2 & -\alpha/(\alpha+\beta)^2 \\ -\frac{1}{2}(\alpha+\beta)^{-3/2} & -\frac{1}{2}(\alpha+\beta)^{-3/2} \end{array}\right| = -(\alpha+\beta)^{-5/2}$$

from which the result follows.

2. $p(\phi, \boldsymbol{\theta} \,|\, \boldsymbol{x}) \propto p(\phi)\, p(\boldsymbol{\theta}|\phi)\, p(\boldsymbol{x}|\boldsymbol{\theta})$. So

$$p(\phi, \boldsymbol{\theta} \,|\, \boldsymbol{x}) \propto \phi^n \prod \theta_i^{x_i} \exp(-(\phi+1)\theta_i)$$

and by the usual change–of-variable rule (using $|\mathrm{d}z/\,\mathrm{d}\phi| = 1/(1+\phi)^2$)

$$p(z, \boldsymbol{\theta} \,|\, \boldsymbol{x}) \propto z^{-n-2}(1-z)^n \prod \theta_i^{x_i} \exp(-\theta_i/z).$$

Integrating with respect to all the θ_i using $\int \theta^x \exp(-\theta/z) \propto z^{x+1}$ we get

$$p(z|\boldsymbol{x}) \propto z^{n\overline{x}-2}(1-z)^n$$

or $z \sim \mathrm{Be}(n\overline{x}-1, n+1)$, from which it follows that $\mathsf{E}z = (n\overline{x}-1)/(n\overline{x}+n)$. Now note that

$$p(\theta_i \,|\, x_i, \phi) \propto p(x_i|\theta_i)\, p(\theta_i|\phi) \propto \theta_i^{x_i} \exp(-\theta_i)\, \phi \exp(-\phi\theta_i) \propto \theta_i^{x_i} \exp(-\theta_i/z)$$

which is a $\mathrm{G}(x_i, z)$ or $\frac{1}{2}z\chi^2_{x_i}$ distribution, from which it follows that $\mathsf{E}(\theta_i \,|\, x_i, \phi) = x_i z$ and so

$$\mathsf{E}(\theta_i|\boldsymbol{x}) = x_i \mathsf{E}(z|\boldsymbol{x}) = x_i(n\overline{x}-1)/(n\overline{x}+n).$$

3. (a) Use same estimator $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}^{\boldsymbol{B}}$. Expression for $\rho(\widehat{\boldsymbol{\theta}}^0, \boldsymbol{X})$ has S_1 replaced by a weighted sum of squares about μ.

(b) Use estimator $\widehat{\boldsymbol{\theta}}$ with $\widehat{\theta}_i$ the posterior median of the ith component.

4. We find the mean square error of the transformed data is 3.14 times smaller and of the equivalent probabilities is 3.09 (as opposed to corresponding figures of 3.50 and 3.49 for the Efron-Morris estimator) For a complete analysis, see the program

`http://www-users.york.ac.uk/~pml1/bayes/rprogs/baseball.txt`

or

`http://www-users.york.ac.uk/~pml1/bayes/cprogs/baseball.cpp`

5. Evidently $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$ or equivalently $\boldsymbol{\alpha}_j{}^{\mathrm{T}}\boldsymbol{\alpha}_k = 1$ if $j = k$ and 0 if $j \neq k$. It follows that $W_j \sim \mathrm{N}(0,1)$ and that $\mathcal{C}(W_j, W_k) = 0$ if $j \neq k$ which, since the W_j have a multivariate normal distribution, implies that W_j and W_k are independent. Further, as $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$ we also have $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$ and so

$$\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} = (\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X} - \boldsymbol{\mu}) = (\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{X} - \boldsymbol{\mu}).$$

For the last part, normalize the α_{ij} for fixed j to produce a unit vector and the rest follows as above.

6. We know that

$$R(\theta, \widehat{\boldsymbol{\theta}}^{\boldsymbol{JS}}) - R(\theta, \widehat{\boldsymbol{\theta}}^{\boldsymbol{JS+}}) = \frac{1}{r}\sum\left[\mathsf{E}\left(\widehat{\theta}^{JS^2} - \widehat{\theta}^{JS+^2}\right) - 2\theta_i\mathsf{E}\left(\widehat{\theta}_i^{JS} - \widehat{\theta}_i^{JS+}\right)\right].$$

To show that the experession in brackets is always > 0, calculate the expectation by first conditioning on S_1. For any value $S_1 \leqslant r-2$, we have $\widehat{\theta}_i^{JS} = \widehat{\theta}_i^{JS+}$ so that it is enough to show that the right hand side is positive when conditional on any value $S_1 = s_1 > r-2$. Since in that case $\widehat{\theta}_i^{JS+} = 0$, it is enough to show that for any $s_1 > r-2$

$$\theta_i\mathsf{E}\left[\widehat{\theta}_i^{JS}\,\middle|\, S_1 = s_1\right] = \theta_i\left(1 - \frac{r-2}{s_1}\right)\mathsf{E}(X_i - \mu_i \,|\, S_1 = s_1) \leqslant 0$$

and hence it is enough to show that $\theta_i\mathsf{E}(X_i - \mu_i \,|\, S_1 = s_1) \geqslant 0$. Now $S_1 = s_1$ is equivalent to $(X_1 - \mu)^2 = s_1 - \left\{(X_2-\mu)^2 + \cdots + (X_n - \mu)^2\right\}$. Conditioning further on $X_2, \dots, X_n$ and noting that

$$\mathsf{P}(X_1 = y \,|\, S_1 = s_1, X_2 = x_2, \dots, X_n = x_n) \propto \exp\left(-\tfrac{1}{2}(y - \theta_1)^2\right)$$
$$\mathsf{P}(X_1 = -y \,|\, S_1 = s_1, X_2 = x_2, \dots, X_n = x_n) \propto \exp\left(-\tfrac{1}{2}(y + \theta_1)^2\right)$$

we find that

$$\mathsf{E}\left[\theta_1(X_1 - \mu_1) \,|\, S_1 = s_1, X_2 = x_2, \dots X_n = x_n\right] = \frac{\theta_1 y(\mathrm{e}^{\theta_1 y} - \mathrm{e}^{-\theta_1 y})}{\mathrm{e}^{\theta_1 y} + \mathrm{e}^{-\theta_1 y}}$$

where $y = \sqrt{s_1 - (x_2 - \mu)^2 - \cdots - (x_n - \mu)^2}$. The right hand side is an increasing function of $|\theta_1 y|$, which is zero when $\theta_1 y = 0$, and this completes the proof.

7. Note that, as $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + k\boldsymbol{I}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ commute,

$$\begin{aligned}\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_k &= \left\{(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + k\boldsymbol{I}) - (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\right\}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + k\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \\ &= k(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\widehat{\boldsymbol{\theta}}_k.\end{aligned}$$

The bias is

$$\boldsymbol{b}(k) = \left\{(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + k\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} - \boldsymbol{I}\right\}\boldsymbol{\theta},$$

and the sum of the squares of the biases is

$$\mathcal{G}(k) = \boldsymbol{b}(k)^{\mathrm{T}}\boldsymbol{b}(k) = (\mathsf{E}\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^{\mathrm{T}}(\mathsf{E}\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta}).$$

The variance-covariance matrix is

$$\mathcal{V}(\widehat{\boldsymbol{\theta}}_k) = \phi(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + k\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + k\boldsymbol{I})^{-1}$$

so that the sum of the variances of the regression coefficients is

$$\begin{aligned}\mathcal{F}(k) &= \mathsf{E}(\widehat{\boldsymbol{\theta}}_k - \mathsf{E}\boldsymbol{\theta})^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}_k - \mathsf{E}\boldsymbol{\theta})\\ &= \mathrm{Trace}(\mathcal{V}(\widehat{\boldsymbol{\theta}}_k))\\ &= \phi\,\mathrm{Trace}\left\{(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + k\boldsymbol{I})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\,(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + k\boldsymbol{I})^{-1}\right\}\end{aligned}$$

and the residual sum of squares is

$$\begin{aligned}RSS_k &= (\boldsymbol{x} - \boldsymbol{A}\widehat{\boldsymbol{\theta}}_k)^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{A}\widehat{\boldsymbol{\theta}}_k)\\ &= \mathsf{E}\{(\boldsymbol{x} - \boldsymbol{A}\widehat{\boldsymbol{\theta}}) + \boldsymbol{A}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_k)\}^{\mathrm{T}}\{(\boldsymbol{x} - \boldsymbol{A}\widehat{\boldsymbol{\theta}}) + \boldsymbol{A}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_k)\}.\end{aligned}$$

Because $\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{A}\widehat{\boldsymbol{\theta}}) = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0}$, this can be written as

$$\begin{aligned}RSS_k &= (\boldsymbol{x} - \boldsymbol{A}\widehat{\boldsymbol{\theta}})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{A}\widehat{\boldsymbol{\theta}}) + (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_k)^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_k)\\ &= RSS + k^2\,\widehat{\boldsymbol{\theta}}_k{}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\widehat{\boldsymbol{\theta}}_k\end{aligned}$$

while the mean square error is

$$\begin{aligned}MSE_k &= \mathsf{E}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})\\ &= \mathsf{E}\{(\widehat{\boldsymbol{\theta}}_k - \mathsf{E}\widehat{\boldsymbol{\theta}}_k) + (\mathsf{E}\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})\}^{\mathrm{T}}\{(\widehat{\boldsymbol{\theta}}_k - \mathsf{E}\widehat{\boldsymbol{\theta}}_k) + (\mathsf{E}\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})\}\\ &= \mathsf{E}(\widehat{\boldsymbol{\theta}}_k - \mathsf{E}\widehat{\boldsymbol{\theta}}_k)^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}_k - \mathsf{E}\widehat{\boldsymbol{\theta}}_k) + (\mathsf{E}\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^{\mathrm{T}}(\mathsf{E}\widehat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})\\ &= \mathcal{F}(k) + \mathcal{G}(k)\end{aligned}$$

using the fact that $\mathsf{E}(\widehat{\boldsymbol{\theta}}_k - \mathsf{E}\widehat{\boldsymbol{\theta}}_k) = 0$.

It can be shown that $\mathcal{F}(k)$ is a continuous monotonic decreasing function of k with $\mathcal{F}'(0) < 0$, while $\mathcal{G}(k)$ is a continuous monotonic increasing function with $\mathcal{G}(0) = \mathcal{G}'(0) = 0$ which approaches $\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}$ as an upper limit as $k \to \infty$. It follows that there always exists a $k > 0$ such that $MSE_k < MSE_0$. In fact, this is always true when $k < 2\phi/\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}$ (cf. C. M. Theobold, "Generalizations of mean square error applied to ridge regression", *Journal of the Royal Statistical Society Series* **B**, **36** (1974), 103–106).

8. We defined

$$\boldsymbol{H}^{-1} = \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\boldsymbol{B}\left(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{B}\right)\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}$$

so that

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{H}^{-1}\boldsymbol{B} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{B} = \boldsymbol{0}.$$

If $\boldsymbol{B}$ is square and non-singular then

$$\boldsymbol{H}^{-1} = \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\boldsymbol{B}\boldsymbol{B}^{-1}\boldsymbol{\Psi}\left(\boldsymbol{B}^{\mathrm{T}}\right)^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1} = \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1} = \boldsymbol{0}.$$

9. It is easily seen that

$$\begin{aligned}
\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1} &= \begin{pmatrix} \psi_\alpha^{-1} & \psi_\alpha^{-1} & 0 & 0 \\ 0 & 0 & \psi_\beta^{-1} & \psi_\beta^{-1} \end{pmatrix}, \\
\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{B} &= \begin{pmatrix} 2\psi_\alpha^{-1} & 0 \\ 0 & 2\psi_\beta^{-1} \end{pmatrix}
\end{aligned}$$

so that

$$\begin{aligned}
\boldsymbol{H}^{-1} &= \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\boldsymbol{B}\left(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{B}\right)\boldsymbol{B}^{\mathrm{T}}\boldsymbol{\Psi}^{-1} \\
&= \begin{pmatrix} \frac{1}{2}\psi_\alpha^{-1} & -\frac{1}{2}\psi_\alpha^{-1} & 0 & 0 \\ -\frac{1}{2}\psi_\alpha^{-1} & \frac{1}{2}\psi_\alpha^{-1} & 0 & 0 \\ 0 & 0 & \frac{1}{2}\psi_\beta^{-1} & -\frac{1}{2}\psi_\beta^{-1} \\ 0 & 0 & -\frac{1}{2}\psi_\beta^{-1} & \frac{1}{2}\psi_\beta^{-1} \end{pmatrix}
\end{aligned}$$

while

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{pmatrix} 4 & 0 & 2 & 2 \\ 0 & 4 & 2 & 2 \\ 2 & 2 & 4 & 0 \\ 2 & 2 & 0 & 4 \end{pmatrix}.$$

Now

$$\begin{aligned}
\boldsymbol{K}^{-1} &= \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\boldsymbol{A} + \boldsymbol{H}^{-1} \\
&= \begin{pmatrix} 4\phi + \frac{1}{2}\psi_\alpha^{-1} & -\frac{1}{2}\psi_\alpha^{-1} & 2\phi & 2\phi \\ -\frac{1}{2}\psi_\alpha^{-1} & 4\phi + \frac{1}{2}\psi_\alpha^{-1} & 2\phi & 2\phi \\ 2\phi & 2\phi & 4\phi + \frac{1}{2}\psi_\beta^{-1} & -\frac{1}{2}\psi_\beta^{-1} \\ 2\phi & 2\phi & -\frac{1}{2}\psi_\beta^{-1} & 4\phi + \frac{1}{2}\psi_\beta^{-1} \end{pmatrix}
\end{aligned}$$

10. See D. V. Lindley and A. F. M. Smith, "Bayes estimates for the linear model" (with discussion), *Journal of the Royal Statistical Society Series* **B**, **34** (1971), 1–41 [reprinted in N. Polson and G. C. Tiao, *Bayesian Inference* (2 vols), (*The International Library of Critical Writings in Econometrics*, No. 7), Aldershot: Edward Elgar (1995)].

11. We have all $a_i = a$ so $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = -mb/a$ so $1 + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = (a - mb)/a$ and thus

$$\begin{aligned}
\Sigma_{ii} &= \frac{1}{a}\left(1 - \frac{1}{1 + \boldsymbol{u}^{\mathrm{T}}\boldsymbol{u}}\frac{b}{a}\right) = \frac{1}{a}\left(1 - \frac{b}{a - mb}\right) = \frac{a - (m+1)b}{a(a - mb)} \\
\Sigma_{ij} &= -\frac{(m+1)b}{a(a - mb)}
\end{aligned}$$

where $a = n/\phi + 1/\phi$ and $b = -1/r\psi$.

D.9 Exercises on Chapter 9

1. A simulation in R, the program for which can be seen on

```
http://www-users.york.ac.uk/~pml1/bayes/tprogs/integral.txt
```

resulted in values 1.684199, 1.516539, 1.7974, 1.921932, 1.595924, 1.573164, 1.812976, 1.880691, 1.641073, 1.770603 with a mean of 1.719450 as as opposed to the theoretical value of $\mathrm{e} - 1 = 1.71828$. The theoretical variance of a single value of e^X where $X \sim \mathrm{U}(0, 1)$ is $\frac{1}{2}(\mathrm{e} - 1)(3 - \mathrm{e}) = 0.24204$ so that the standard deviation of the mean of 10 values is $\sqrt{0.24204/10} = 0.15557$ which compares with a sample standard deviation of 0.1372971.

A simulation in C++, the program for which can be seen on

```
http://www-users.york.ac.uk/~pml1/bayes/cprogs/integral.cpp
```

resulted in values 1.74529, 1.86877, 1.68003, 1.95945, 1.62928, 1.62953, 1.84021, 1.87704, 1.49146, 1.55213 with a mean of 1.72732 and a sample standard deviation of 0.155317.

2. The easiest way to check the value of $\boldsymbol{A}^t$ is by induction. Then note that for large t the matrix $\boldsymbol{A}^t$ is approximately

$$\begin{pmatrix} 2/5 & 3/5 \\ 2/5 & 3/5 \end{pmatrix}$$

from which the result follows.

3. In this case

$$Q = \mathsf{E}[(y_1 + y_2 - 1)\log\eta + (y_3 + y_4 - 1)\log(1 - \eta)]$$

giving

$$\eta^{(t+1)} = \frac{\mathsf{E}(y_1 \,|\, \eta^{(t)}, \boldsymbol{x}) + x_2 - 1}{\mathsf{E}(y_1 \,|\, \eta^{(t)}, \boldsymbol{x}) + x_2 + x_3 + \mathsf{E}(y_4 \,|\, \eta^{(t)}, \boldsymbol{x}) - 1}$$

where $y_1 \sim \mathrm{B}(x_1, \eta/(\eta + 1))$ and $y_4 \sim \mathrm{B}(x_4, (1 - \eta)/(2 - \eta))$, so that

$$\begin{aligned}
\eta^{(t+1)} &= \frac{x_1\eta^{(t)}/(\eta^{(t)} + 1) + x_2 - 1}{x_1\eta^{(t)}/(\eta^{(t)} + 1) + x_2 + x_3 + x_4(1 - \eta^{(t)})/(2 - \eta^{(t)}) - 2} \\
&= \frac{461\eta^{(t)}/(\eta^{(t)} + 1) + 130 - 1}{461\eta^{(t)}/(\eta^{(t)} + 1) + 130 + 161 + 515(1 - \eta^{(t)})/(2 - \eta^{(t)}) - 2}.
\end{aligned}$$

Starting with $\eta^{(0)} = 0.5$, successive iterations give 0.4681, 0.4461, 0.4411, 0.4394, 0.4386, 0.4385, and thereafter 0.4384. This compares with a value of 0.439 found by maximum likelihood in C. A. B. Smith, *op. cit.*

4. The proof that $p(\eta^{(t+1)}\,|\,\boldsymbol{x}) \geqslant p(\eta^{(t+1)}\,|\,\boldsymbol{x})$ in the subsection "Why the *EM* algorithm works" only uses the properties of a *GEM* algorithm. As for the last part, if convergence has occurred there is no further room for increasing Q, so we must have reached a maximum.

5. Take prior for variance $S_0\chi^2_\nu$, that is, $350\chi^2_4$, and prior for mean which is $\mathrm{N}(\theta_0, \phi_0)$ where $\theta_0 = 85$ and $\phi_0 = S_0/n_0(\nu-2) = 350/(1\times 2) = 175$. We then have data with $n = 100$, $\overline{x} = 89$ and $S_1 = (n-1)s^2 = 2970$. Successive iterations using the *EM* algorithm give the posterior mean as 88.9793 and at all subsequent stages 88.9862. This compares with 88.96 as found in Exercise 16 on Chapter 2.

6. Means for the four looms are 97.3497, 91.7870, 95.7272 and 96.8861, overall mean is 97.4375. Variance of observations from the same loom is 1.6250 and variance of means from different looms in the population is 5.1680.

7. (a) For the example on genetic linkage in Section 9.3, see

`http://www-users.york.ac.uk/~pml1/bayes/rprogs/dataaug.txt`

or

`http://www-users.york.ac.uk/~pml1/bayes/cprogs/dataaug.cpp`

(b) For the example on chained data sampling due to Casella and George, see a similar file with `chained` in place of `dataaug`.

(c) For the example on the semi-conjugate prior with a normal likelihood (using both the *EM* algorithm and chained data augmentation, see a similar file with `semiconj` in place of `dataaug`.

(d) For the example on a change-point analysis of data on coal mining disasters, see similar files with `coalnr.cpp` or `coal.cpp` in place of `dataaug.cpp` (the difference is that the former uses only routines in W. H. Press *et al.*, *Numerical Recipes in C* (2nd edn), Cambridge: University Press 1992, whereas the latter program, which is in many ways preferable, uses gamma variates with non-integer parameters).

8. See the program referred to in part (a) of the previous answer.

9. See the program referred to in part (b) of the answer to the question before last.

10. See the program

http://www-users.york.ac.uk/~pml1/bayes/rprogs/semicon2.txt

or

http://www-users.york.ac.uk/∼pml1/bayes/cprogs/semicon2.cpp

11. B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis* (2nd edn), Chapman and Hall 2000, p. 149, remark that

"In our dataset, each rat was weighed once a week for five consecutive weeks. In fact the rats were all the same age at each weighing $x_{i1} = 8$, $x_{i2} = 15$, $x_{i3} = 22$, $x_{i4} = 29$, and $x_{i5} = 36$ for all i. As a result we may simplify our computations by rewriting the likelihood as

$$Y_{ij} \sim N(\alpha_i + \beta_i(x_{ij} - \overline{x})\,,\,\sigma^2),\ i = 1, ..., k,\ j = 1, ..., n_i,$$

so that it is now reasonable to think of α_i and β_i as independent *a priori*. Thus we may set $\Sigma = \mathit{Diag}(\sigma^2_\alpha, \sigma^2_\beta)$, and replace the Wishart prior with a product of idependent inverse gamma priors, say $\mathit{IG}(a_\alpha, b_\alpha)$ and $\mathit{IG}(a_\beta, b_\beta)$."

This being so, it probably suffices to proceed as in the 'Rats' example supplied with WinBUGS.

A more detailed description of the general set-up is to be found in Taken from "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling", Alan E. Gelfand, Susan E. Hills, Amy Racine-Poon, and Adrian F. M. Smith, *Journal of the American Statistical Association* **85** (412) (1990), 972–985, Section 6, pp. 978–979, which can be found at

http://www-users.york.ac.uk/~pml1/bayes/rprogs/gelfand.pdf

(LaTeXsource at gelfand.htm). A program for generating random matrices with a Wishart distribution based on the algorithm of Odell and Feiveson described in W. J. Kennedy and J. E. Gentle, Statistical Computing, Marcel Dekker 1980, pp. 231–232 can be found at

http://www-users.york.ac.uk/~pml1/bayes/rprogs/wishart.txt

while the relevant extract from Kennedy and Gentle is at

http://www-users.york.ac.uk/~pml1/bayes/rprogs/rwishart.pdf

12. See the program at

http://www-users.york.ac.uk/~pml1/bayes/rprogs/linkagemh.txt

13. See the program at

```
http://www-users.york.ac.uk/~pml1/bayes/winbugs/wheat.txt
```

or

```
http://www-users.york.ac.uk/~pml1/bayes/rprogs/wheat.txt
```

14. See the program at

```
http://www-users.york.ac.uk/~pml1/bayes/winbugs/logisti2.txt
```

or

```
http://www-users.york.ac.uk/~pml1/bayes/rprogs/logisti2.txt
```

D.10 Exercises on Chapter 10

1. We note that, since $\mathsf{E}X^2 \geqslant (\mathsf{E}X)^2$,

$$\begin{aligned}\mathsf{E}w(x)^2 &= \int \left(\frac{f(x)q(x)}{p(x)}\right)^2 p(x)\,\mathrm{d}x \geqslant \left(\int \frac{|f(x)|q(x)}{p(x)}p(x)\,\mathrm{d}x\right)^2 \\ &= \left(\int |f(x)|q(x)\,\mathrm{d}x\right)^2.\end{aligned}$$

If we take

$$p(x) = \frac{|f(x)|q(x)}{\int |f(\xi)|q(\xi)\,\mathrm{d}\xi}$$

then

$$\mathsf{E}w(x)^2 = \int \left(\int |f(\xi)|q(\xi)\,\mathrm{d}\xi\right)^2 p(x)\,\mathrm{d}x. = \left(\int |f(x)|q(x)\,\mathrm{d}x\right)^2$$

so that this choice of $p(x)$ minimizes $\mathsf{E}w(x)^2$ and, since we always have $\mathsf{E}w(x) = \theta$, consequently minimizes $\mathcal{V}w(x)$.

2. For sampling with the Cauchy distribution we proceed thus:

(a) Substituting $x = \tan\theta$ we find

$$\begin{aligned}\eta = \mathsf{P}(x > 2) &= \int_2^\infty \frac{1}{\pi}\frac{1}{1+x^2}\,\mathrm{d}x = \tfrac{1}{2} - \tan^{-1}(2)/\pi = \tan^{-1}(\tfrac{1}{2})/\pi \\ &= 0.147\,583\,6.\end{aligned}$$

The variance is

$$\eta(1-\eta)/n = 0.125\,802\,7/n.$$

(b) Use the usual change of variable rule to deduce the density of y. Then note that

$$\frac{q(y_i)}{p(y_i)} = \frac{1/(\pi(1+y_i^2))}{2/y_i^2} = \frac{1}{2\pi}\frac{y_i^2}{1+y_i^2}$$

and that we can take $f(y) = 1$ as all values of y satisfy $y \geqslant 2$.

(c) Substitute $x_i = 2/y_i$ in the above to get

$$\frac{1}{2\pi}\frac{4}{4+x_i^2}.$$

(d) On writing $x = 2\tan\theta$ and $\theta_0 = \tan^{-1}(\frac{1}{2})$ we find

$$\mathsf{E}\widehat{\eta} = \int_0^1 \frac{1}{2\pi}\frac{4}{4+x^2}\,\mathrm{d}x = \int_0^1 \frac{1}{2\pi}\frac{1}{\sec^2\theta}2\sec^2\theta\,\mathrm{d}\theta = \theta_0/\pi = \eta$$

Similarly, noting that $\sin(2\theta_0) = 4/5$ the integral for $\mathsf{E}\widehat{\eta}^2$ transforms to

$$\begin{aligned}\int_0^1 \frac{1}{4\pi^2}\frac{1}{\sec^4\theta}2\sec^2\theta\,\mathrm{d}\theta &= \int_0^1 \frac{1}{4\pi^2}2\cos^2\theta\,\mathrm{d}\theta \\ &= \int_0^1 \frac{1}{4\pi^2}\{1+\cos(2\theta)\}\,\mathrm{d}\theta \\ &= \left[\theta + \tfrac{1}{2}\sin(2\theta)\right]_0^{\theta_0}/(4\pi)^2 \\ &= \left\{\tan^{-1}(\tfrac{1}{2}) + \tfrac{2}{5}\right\}/(4\pi^2) \\ &= 0.021\,876\,4\end{aligned}$$

so that

$$\mathcal{V}\widehat{\eta} = \mathsf{E}\widehat{\eta}^2 - (\mathsf{E}\widehat{\eta})^2 = 0.000\,095\,5.$$

3. A suitable program is

```
n <- 10000
alpha <- 2
beta <- 3
x <- runif(n)
p <- function(x) dunif(x)
q <- function(x) dbeta(x,alpha,beta)
w <- q(x)/p(x)
pi <- w/sum(w)
samp <- sample(x,size=n,prob=pi,replace=T)
print(mean(samp))
print(var(samp))
```

A run of this program resulted in a value of the mean of 0.399 as compared with the true value of 0.4 and a value of the mean of 0.0399 compared with a true value of 0.04.

4. Continue the program (avoiding ambiguity in the definition of α) by

```
theta <- 0.1
le <- (1-theta)*n
lo <- 1:(n-le)
hi <- (le+1):n
y <- sort(samp)
r <- y[hi]-y[lo]
rm <- min(r)
lom <- min(lo[r==rm])
him <- min(hi[r==rm])
dd <- function(x) dbeta(x,alpha,beta)
plot(dd,xlim=c(0,1),ylim=c(0,2))
par(new=T)
plot(density(samp),xlim=c(0,1),ylim=c(0,2),
  main="",xlab="",ylab="")
abline(v=y[lom])
abline(v=y[him])
print(y[lom])
print(y[him])
```

The lower limit resulting from a run of the program was 0.061 and the upper limit was 0.705. Referring to the tables in the Appendix we find that values of F corresponding to a 90% HDR of $\log \mathrm{F}_{6,4}$ are $\underline{\mathrm{F}} = 0.21$ and $\overline{\mathrm{F}} = 5.75$. It follows that an appropriate interval of values of $F_{4,6}$ is $(1/\overline{\mathrm{F}}, 1/\underline{\mathrm{F}})$, that is $(0.17, 4.76)$, so that an appropriate interval for π is

$$\frac{2 \times 0.17}{3 + 2 \times 0.17} \leqslant \pi \leqslant \frac{2 \times 4.76}{3 + 2 \times 4.76}$$

that is $(0.10, 0.76)$ (but note that in Section 3.1 we were looking for intervals in which the distribution of the log-odds is higher than anywhere outside, rather than for HDRs for the beta distribution itself).

5. A suitable program is

```
r <- 10
phi <- rep(NA,r)
mu <- rep(NA,r)
S <- rep(NA,r)
n <- 100; xbar <- 89; SS <- (n-1)*30
mu0 <- 85; S0 <- 350; nu0 <- 4; n0 <- 1
phi0 <- S0/(n0*(nu0-2))
S[1] <- S0
nustar <- nu0 + n
```

```
for (i in 2:r) {
  phi[i] <- (1/phi0 + n*nustar/S[i-1])^{-1}
  mu[i] <- phi[i]*(mu0/phi0 + n*xbar*nustar/S[i-1])
  S[i] <- S0 + (SS+n*xbar^2) - 2*n*xbar*mu[i+1] +
             n*(mu[i]^2 + phi[i])
  cat("i",i,"phi",phi[i],"mu",mu[i],"S",S[i],"\n")
}
mustar <- mu[r]; phistar <- phi[r]; Sstar <- S[r]
cat("mu has mean",mustar,"and s.d.",sqrt(phistar),"\n")
cat("phi has mean",Sstar/(nustar-2),"and s.d.",
     (Sstar/(nustar-2))*sqrt(2/(nustar-4)),"\n")
```

We get convergence to $\mu = 88.993$ and $\phi = 0.322$.

6. In the discrete case the Kullback-Leibler divergence is

$$\begin{aligned}
\mathcal{I}(q:p) &= \sum q(x)\log\{q(x)/p(x)\} \\
&= \sum \binom{n}{x}\rho^x(1-\rho^{n-x})\log\left\{\binom{n}{x}\rho^x(1-\rho^{n-x}) \Big/ \binom{n}{x}\pi^x(1-\pi^{n-x})\right\} \\
&= \sum \binom{n}{x}\rho^x(1-\rho^{n-x})\,\{x\log(\rho/\pi)+(n-x)\log[(1-\rho)/(1-\pi)]\} \\
&= n\rho\log(\rho/\pi)+n(1-\rho)\log[(1-\rho)/(1-\pi)]
\end{aligned}$$

using the fact that the mean of $\mathrm{B}(n,\rho)$ is $n\rho$.

For $\mathcal{I}(q:p) = \mathcal{I}(p:q)$ we need

$$n\rho\log(\rho/\pi)+n(1-\rho)\log[(1-\rho)/(1-\pi)] = n\pi\log(\pi/rho)+n(1-\pi)\log[(1-\pi)/(1-\rho)]$$

so

$$(\pi-\rho)\log\left\{\frac{\pi}{1-\pi}\Big/\frac{\rho}{1-\rho}\right\}$$

from which it is clear that this can happen if and only if $\pi = \rho$ (when $\mathcal{I}(q:p) = \mathcal{I}(p:q) = 0$).

7. In the continuous case the Kullback-Leibler divergence is

$$\begin{aligned}
\mathcal{I}(q:p) &= \int q(x)\log\{q(x)/p(x)\}\,\mathrm{d}x \\
&= \int (2\pi\psi)^{-1/2}\exp\{-\tfrac{1}{2}(x-\nu)^2/\psi\}\times \\
&\quad \Big\{\log\Big[(2\pi\psi)^{-1/2}\exp\{-\tfrac{1}{2}(x-\nu)^2/\psi\}\Big] \\
&\qquad -\log\Big[(2\pi\phi)^{-1/2}\exp\{-\tfrac{1}{2}(x-\mu)^2/\phi\}\Big]\Big\}
\end{aligned}$$

$$= \tfrac{1}{2}\log(\psi/\phi) - \mathsf{E}(x-\nu)^2/\phi + \mathsf{E}(x-\mu)^2/\phi$$

where $x \sim \mathrm{N}(\nu,\phi)$. By writing $(x-\mu)^2 = \{(x-\nu)+(\nu-\mu)\}^2$ it is easily concluded that

$$2\mathcal{I}(q:p) = \log(\phi/\psi) + (\phi-\psi)/\psi + (\nu-\mu)^2/\phi.$$

In particular if $\phi = \psi$ then

$$\mathcal{I}(q:p) = \tfrac{1}{2}(\nu-\mu)^2/\phi.$$

8. We find

$$\begin{aligned}\mathcal{I}(q:p) &= \int q(x)\log\{q(x)/p(x)\}\,\mathrm{d}x \\ &= \int_0^\infty \beta^{-1}\exp(-x/\beta)\log\left\{\left(\beta^{-1}\exp(-x/\beta)\right)\Big/2(2\pi)^{-1/2}\exp(-\tfrac{1}{2}x^2)\right\} \\ &= \tfrac{1}{2}\log(8\pi) - \log\beta - \mathsf{E}x/\beta + \tfrac{1}{2}\mathsf{E}x^2\end{aligned}$$

where $x \sim \mathrm{E}(\beta)$. It follows that

$$\mathcal{I}(q:p) = \tfrac{1}{2}\log(8\pi) - \log\beta - 1 + \beta^2$$

and hence

$$\begin{aligned}\frac{\partial\mathcal{I}(q:p)}{\partial\beta} &= -\frac{1}{\beta} + 2\beta \\ \frac{\partial^2\mathcal{I}(q:p)}{\partial\beta^2} &= \frac{1}{\beta^2} + 2\end{aligned}$$

It follows that $\mathcal{I}(q:p)$ is a minimum when $\beta = \sqrt{2}$.

9. See the paper by Corduneaunu and Bishop (2001).

10. The model is

$$(\tfrac{1}{4}+\tfrac{1}{4}\eta,\ \tfrac{1}{4}\eta,\ \tfrac{1}{4}(1-\eta),\ \tfrac{1}{4}(1-\eta)+\tfrac{1}{4})$$

and the values quoted are $x_1 = 461$, $x_2 = 130$, $x_3 = 161$ and $x_4 = 515$. An R program for the *ABC-REJ* algorithm is

```
N <- 100000
etastar <- c(461,130,161,515)
n <- sum(etastar)
eps <- 22
d <- rep(NA,N)
trial <- rep(NA,N)
```

```
for (j in 1:N) {
etatrial <- runif(1)
  inds <- sample(4,n,replace=T,p=c(0.25+etatrial/4,etatrial/4,
            (1-etatrial)/4,(1-etatrial)/4+0.25))
  samp <- c(sum(inds==1),sum(inds==2),sum(inds==3),sum(inds==4))
  d <- sqrt(sum((etastar-samp)^2))
  if (d <= eps) trial[j] <- etatrial
}
eta <- trial[!is.na(trial)]
k <- length(eta)
m <- mean(eta)
s <- sd(eta)
cat("k",k,"m",m,"s",s,"\n")
```

Similar modifications can be made for the other algorithms.

11. Let M_i denote model i ($i = 1, 2, 3, 4$) with model parameters θ_i, where $\theta_1 = (\alpha, \tau)$, $\theta_2 = (\alpha, \beta, \tau)$, $\theta_3 = (\alpha, \lambda, \tau)$ and $\theta_4 = (\alpha, \beta, \lambda, \tau)$.

The reversible jump MCMC algorithm is as follows:

- Initiate the algorithm by choosing an initial model, m to start in and initial parameter values θ_m.
- Each iteration t:
  - Update the model parameters $\theta_{m_{t-1}}$, where m_{t-1} is the model which the algorithm is residing in at the end of iteration $t-1$.
  - Propose to move to a new model M_t' with parameter values θ'. Calculate the acceptance probability for the move and decide whether to accept or reject the new model (and parameter values).

The models are nested with model 4 being the full model with sub-models model 2 and model 3 and the simplest model 1. Each model either includes or excludes β and λ (exclusion of a parameter is equivalent to setting it equal to 0).

Given that the current model j, employ the following reversible jump:

Choose randomly either β and λ, and propose to move to model i has the chosen parameter included (excluded) if the parameter is excluded (included) in model j.

A bijection is required for transforming between the models. Therefore for inclusion of a parameter propose $u \sim \mathrm{N}(0, \sigma^2)$ for suitably chosen σ^2 and propose to set the parameter equal to u. (This works best if x_i and t_i have been standardised to have mean 0 and variance 1.) Leave the other parameters unchanged. Thus the Jacobian of the transformation is simply 1.

Let $L_j(\theta_j)$ denote the likelihood of θ_j given model M_j. Then the acceptance probability for the inclusion of a parameter (either β or τ set equal to u) is

$$\min\left\{1,\ \frac{L_i(\theta_i)}{L_j(\theta_j)} \times \frac{\pi(\theta_i, M_i)}{\pi(\theta_j, M_j)} \times \frac{1/2}{1/2} \times \frac{1}{\exp(-\frac{1}{2}u^2/\sigma^2)/\sqrt{2\pi\sigma^2}}\right\}.$$