Problem 1 Calculate the predictive probability for k future heads out of m tosses based on r heads observed in n tosses, using a beta prior density for the success probability.

Solution 1 The posterior density for the success probability θ is

$$\pi(\theta \mid R = r) = \frac{\theta^{a+r-1}(1-\theta)^{b+n-r-1}}{B(a+r, b+n-r)}, \quad 0 < \theta < 1,$$

and the conditional probability of k further heads in m further independent tosses is

$$\mathrm{P}(K = k \mid \theta) = \binom{m}{k} \theta^k (1-\theta)^{m-k}, \quad k \in \{0, \dots, m\},$$

so the required predictive probability is

$$\begin{aligned} \mathrm{P}(K = k \mid R = r) &= \int_0^1 \mathrm{P}(K = k \mid \theta) \pi(\theta \mid R = r) \, \mathrm{d}\theta \\ &= \binom{m}{k} \frac{B(a+r+k, b+n+m-r-k)}{B(a+r, b+n-r)}, \quad k \in \{0, \dots, m\}. \end{aligned}$$

Problem 2 How would you express prior ignorance about an angle? About the position of a star in the firmament?

Solution 2 Total ignorance would presumably correspond to uniform distributions on the circle an on the sphere, i.e.,

$$\pi(\theta) = \frac{1}{2\pi}, \quad 0 < \theta \leq 2\pi, \quad \pi(\theta, \phi) = \frac{1}{4\pi}, \quad 0 < \theta \leq 2\pi, -\pi/2 \leq \phi \leq \pi/2.$$

Taking into account the positions of the Milky Way or other astronomical features would be harder (and would depend on the time of day ...).

Problem 3 Verify this table of conjugate prior densities:

$f(y \mid \theta)$	Parameter	Prior
Binomial	success probability	beta
Poisson	mean	gamma
Exponential	rate	gamma
Normal	mean (known variance)	normal
Normal	variance (known mean)	inverse gamma

Solution 3

For the binomial model, we have $y \in \{0, 1, \dots, m\}$ and

$$f(y; \pi) = \binom{m}{y} \theta^y (1-\theta)^{m-y} \propto \theta^a (1-\theta)^b, \quad 0 < \theta < 1,$$

so the density proportional to $\theta^{a-1}(1-\theta)^{b-1}$ and with $a, b > 0$ is conjugate. This is a beta density.

For the Poisson model, we have $y \in \{0, 1, \dots\}$ and

$$f(y; \mu) = \frac{1}{y!} \mu^y e^{-\mu} \propto \mu^a e^{-b\mu}, \quad \mu > 0,$$

so the density proportional to $\mu^{a-1} e^{-b\mu}$ with $a, b > 0$ is conjugate. This is a gamma density.

For the exponential model, we have $y > 0$ and

$$f(y; \lambda) = \lambda e^{-y\lambda}, \quad \lambda > 0,$$

so the density proportional to $\lambda^{a-1} e^{-b\lambda}$ with $a, b > 0$ is conjugate. This is a gamma density.

For the normal model with known variance, we have y real and

$$f(y; \mu) = \exp\left\{-(y-\mu)^2/\sigma^2 - \tfrac{1}{2}\log(2\pi\sigma^2)\right\} \propto \exp\left\{y\mu/\sigma^2 - \mu^2/(2\sigma^2)\right\} = \exp\left(a\mu - b\mu^2/2\right)$$

with a real and b positive. This corresponds to the $\mathcal{N}(\mu_0, \tau^2)$ density with $a = \mu_0/\tau^2$ and $b = 1/\tau^2$.

For the normal model with known mean, we have y real and

$$f(y; \mu) = \exp\left\{-(y-\mu)^2/\sigma^2 - \tfrac{1}{2}\log(2\pi\sigma^2)\right\} \propto (\sigma^{-2})^a e^{-b/\sigma^2} = \tau^a e^{-b\tau},$$

say, where $\tau = 1/\sigma^2$ and $a > -1$ and $b > 0$. This is again the gamma density, but for $\tau = 1/\sigma^2$, so the density of σ^2 is said to be inverse gamma (more properly *reciprocal gamma*, but it's too late to correct this ...). If τ has the gamma density with parameters $a, b > 0$, then

$$\pi(\tau) = \frac{b^a \tau^{a-1}}{\Gamma(a)} e^{-b\tau}, \quad \tau > 0,$$

then the density of $\sigma^2 = 1/\tau$ is

$$\pi(\sigma^2) = \frac{b^a (\sigma^2)^{1-a}}{\Gamma(a)} e^{-b/\sigma^2} \left|\frac{\mathrm{d}\tau}{\mathrm{d}\sigma^2}\right| = \frac{b^a}{(\sigma^2)^{a+1}\Gamma(a)} e^{-b/\sigma^2}, \quad \sigma^2 > 0.$$

Problem 4 Find elements $\tilde{\theta}$ and $\tilde{J}(\tilde{\theta})$ of the normal approximation to a beta density, and hence check the formulae in Example 11.11. Find also the posterior mean and variance of θ. Give an approximate 0.95 credible interval for θ. How does this differ from a 0.95 confidence interval? Comment.

Solution 4 The log density is

$$\log \pi(\theta) \equiv (a-1)\log\theta + (b-1)\log(1-\theta), \quad 0 < \theta < 1,$$

and its first and second derivatives are

$$\frac{a-1}{\theta} - \frac{b-1}{1-\theta}, \quad -\frac{a-1}{\theta^2} - \frac{b-1}{(1-\theta)^2}.$$

so $\tilde{\theta} = (a-1)/(a+b-2)$ sets the first derivative to zero, and minus the second derivative evaluated at $\tilde{\theta}$ equals

$$J(\tilde{\theta}) = \frac{(a+b-2)^3}{(a-1)(b-1)}.$$

Updating using r successes out of m trials replaces a and b by $a+r$ and $b+m-r$, which yields the formulae in the example.

The posterior mean and variance of θ are $\mu(r) = (a+r/)(m+a+b)$ and $\sigma^2(r) = (a+r)(b+m-r)/(a+b+m)^3$, from which an approximate 95% credible interval can be found as $\mu(r) \pm 1.96\sigma(r)$. This differs from a 95% confidence interval in two ways: first, the presence of a and b in the formulae, which has little numerical effect unless $r=0$ or $r=m$, in which case the usual confidence interval (which has $a=b=0$) doesn't work, because it has length zero; and second, the interpretation. In the credible interval θ is treated as random and the data as fixed, whereas in the confidence interval the parameter is fixed and r is treated as the realisation of a random variable, so the probability is with respect to repeated sampling from the model with θ fixed. Despite this different in interpretation, there is essentially no numerical difference between the intervals unless r is very small or very close to m, so in most cases a Bayesian and frequentist would agree about the inference.

Problem 5 Two independent samples $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ and $X_1, \dots, X_m \overset{\text{iid}}{\sim} \mathcal{N}(\mu, c\sigma^2)$ are available, where $c > 0$ is known. Find posterior densities for μ and σ based on prior $\pi(\mu, \sigma) \propto 1/\sigma$.

Solution 5 The likelihood is

$$\prod_{j=1}^{n} f(y_j \mid \mu, \sigma^2) \times \prod_{j=1}^{m} f(x_j \mid \mu, c\sigma^2), \quad \mu \in \mathbb{R}, \sigma^2 > 0,$$

where as usual in Bayesian settings we condition on the parameters, which are regarded as random variables. The prior is proportional to $1/\sigma$, so inspection of the product of the prior and likelihood implies that we can write

$$\pi(\mu, \sigma^2 \mid y, x) \propto \frac{1}{(\sigma^2)^{(a+1)/2}} \exp\left\{-A(\mu - B)^2/(2\sigma^2) - C/(2\sigma^2)\right\},$$

where $a = m + n$ and A, B and C are to be determined. This implies that the posterior marginal density of σ^2 is

$$\pi(\sigma^2 \mid y, x) = \int \pi(\mu, \sigma \mid y, x)\, \mathrm{d}\mu \propto \frac{1}{(\sigma^2)^{a/2}} \exp\{-C/(2\sigma^2)\}$$

or equivalently that $\phi = 1/\sigma^2$ has density proportional to $\phi^{a/2-2} \exp(-C\phi/2)$, i.e., that the posterior density of ϕ is gamma with shape parameter $(a-2)/2$ and scale parameter $C/2$, or equivalently that the posterior density of σ^2 is inverse gamma with the same parameters. For the posterior marginal density of μ we likewise have

$$\begin{aligned}
\pi(\mu \mid y, x) = \int \pi(\mu, \phi \mid y, x)\, \mathrm{d}\phi \quad &\propto \quad \int \phi^{(a+1)/2-2} \exp\left[-\phi\left\{A(\mu - B)^2/2 + C\right\}/2\right] \mathrm{d}\phi \\
&\propto \quad \left\{A(\mu - B)^2/2 + C\right\}^{-(a-1)/2} \\
&\propto \quad \left\{1 + Ab(\mu - B)^2/(2Cb)\right\}^{-(b+1)/2},
\end{aligned}$$

where $b = n + m - 2$, which implies that $T = (\mu - B)/(C/Ab)^{1/2}$ has a t_b distribution, conditional on the data.

To find A, B and C, we note that the first and second derivatives of the outer sides of

$$\sum_{j=1}^{n}(\mu - y_j)^2 + \sum_{i=1}^{m}(\mu - x_i)^2/c = n(\mu - \overline{y})^2 + \sum_{j=1}^{n}(y_j - \overline{y})^2 + m(\mu - \overline{x})^2/c + \sum_{i=1}^{m}(x_i - \overline{x})^2/c = A(\mu - B)^2 + C$$

with respect to μ yield the equations

$$(n + m/c)\mu - n\overline{y} - m\overline{x}/c = A(\mu - B), \quad n + m/c = A,$$

so $B = (cn\overline{y} + m\overline{x})/(cn + m)$ and $C = \sum_{j=1}^{n}(B - y_j)^2 + \sum_{i=1}^{m}(B - x_i)^2/c$.

Problem 6 Two balls are drawn successively without replacement from an urn containing three white and two red balls. Are the outcomes of the first and second draws independent? Are they exchangeable?

Solution 6 Let W_1 and W_2 denote the indicator variables that the two balls are white. Clearly $\mathrm{P}(W_1 = 1) = 3/5$ and

$$\mathrm{P}(W_1 = W_2 = 1) = \binom{3}{2}\binom{2}{0}\Big/\binom{5}{2} = 3 \times 1/10.$$

To compute $\mathrm{P}(W_2 = 1)$ we either argue by symmetry, or condition on the outcome of W_1:

$$\mathrm{P}(W_2 = 1) = \mathrm{P}(W_2 = 1 \mid W_1 = 1)\mathrm{P}(W_1 = 1) + \mathrm{P}(W_2 = 1 \mid W_1 = 0)\mathrm{P}(W_1 = 0) = \frac{2}{4} \times \frac{3}{5} + \frac{3}{4} \times \frac{2}{5} = \frac{3}{5}.$$

Similar computations (or symmetry) show that $\mathrm{P}(W_1 = 1, W_2 = 0) = \mathrm{P}(W_1 = 0, W_2 = 1)$, so the two outcomes are exchangeable but not independent, because $\mathrm{P}(W_1 = W_2 = 1) \neq \mathrm{P}(W_1 = 1)\mathrm{P}(W_2 = 1)$.

Problem 7 Under what conditions are the Bernoulli random variables Y_1 and $Y_2 = 1 - Y_1$ exchangeable? What about $Y_1, \ldots, Y_n$ given that $Y_1 + \cdots + Y_n = m$?

Solution 7 Let $\mathrm{P}(Y_1 = 1) = p$. For Y_1 and Y_2 to be exchangeable we must have

$$\mathrm{P}(Y_1 = y, Y_2 = y') = \mathrm{P}(Y_1 = y', Y_2 = y), \quad y, y' \in \{0, 1\}.$$

This probability equals zero when $y = y'$, and if $y = 1, y' = 0$ then it equals $\mathrm{P}(Y_1 = 1) = p$, and if $y = 0, y' = 1$ then it equals $\mathrm{P}(Y_1 = 0) = 1 - p$. Hence they are exchangeable only if $p = 1/2$.

For the second part of the question we use de Finetti's theorem, writing $S = Y_1 + \cdots + Y_n$ and

$$\begin{aligned}
\mathrm{P}(Y_1 = y_1, \ldots, Y_n = y_n \mid S = m) &= \frac{\mathrm{P}(S = m \mid Y_1 = y_1, \ldots, Y_n = y_n)\mathrm{P}(Y_1 = y_1, \ldots, Y_n = y_n)}{\mathrm{P}(S = m)} \\
&= \frac{I(\sum Y_j = m)\mathrm{P}(Y_1 = y_1, \ldots, Y_n = y_n)}{\mathrm{P}(S = m)} \\
&= I\left(\sum Y_j = m\right) \frac{\int \mathrm{P}(Y_1 = y_1, \ldots, Y_n = y_n \mid p) f(p)\,\mathrm{d}p}{\mathrm{P}(S = m)} \\
&= I\left(S = m\right) \int \mathrm{P}(Y_1 = y_1, \ldots, Y_n = y_n \mid S = m)\mathrm{P}(S = m \mid p) f(p)\,\mathrm{d}p \\
&= I(S = m)\mathrm{P}(Y_1 = y_1, \ldots, Y_n = y_n \mid S = m) \\
&= \int \mathrm{P}(Y_1 = y_1, \ldots, Y_n = y_n \mid S = m) I(S = m)\,\mathrm{d}m,
\end{aligned}$$

say, because S is minimal sufficient for p. Thus this distribution has a representation in terms of a mixture (of a single distribution!) and hence is exchangeable.

Problem 8 In Example 11.29, suppose that $v_j' = \tau^2 v_j$. Show that an unbiased estimator of τ^2 is then $SS/(n-p) - 1$, where SS is the residual sum of squares and p is the dimension of β, and explain why a better estimator is $\max\{SS/(n-p)-1, 0\}$.

Find also the profile log likelihood when $v_j' = \tau^2$.

Solution 8

If $v_j' = \tau^2 v_j$, then the computations in Example 11.29 imply that marginally $y_j \overset{\text{ind}}{\sim} \mathcal{N}\{x_j^{\mathrm{T}}\beta, (1+\tau^2)v_j\}$. In this case the unbiased estimator of $\sigma^2 = 1+\tau^2$ is the scaled sum of squares

$$\tilde\sigma^2 = (n-p)^{-1}(W^{1/2}y)^{\mathrm{T}}\left\{I - W^{1/2}X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W^{1/2}\right\}(W^{1/2}y) = SS/(n-p),$$

and thus $\widehat\tau^2 = \tilde\sigma^2 - 1$ is unbiased for τ^2. To avoid a negative estimate, it is better to take $\max(\widehat\tau^2, 0)$.

When $v_j' = \tau^2$, the log likelihood function is

$$\ell(\beta, \tau^2) \equiv -\frac{1}{2}\sum_{j=1}^n \log(v_j + \tau^2) - \frac{1}{2}\sum_{j=1}^n \frac{(y_j - x_j^{\mathrm{T}}\beta)^2}{v_j + \tau^2}, \quad \tau^2 \geq 0,$$

and with τ^2 fixed the least squares estimate of β is obtained by weighted least squares regression of y on the columns of X using weight matrix $W_{\tau^2} = \mathrm{diag}\{1/(v_1+\tau^2), \dots, 1/(v_n+\tau^2)\}$, which results in residual sum of squares

$$SS(\tau^2) = (W_{\tau^2}^{1/2}y)^{\mathrm{T}}\left\{I - W_{\tau^2}^{1/2}X(X^{\mathrm{T}}W_{\tau^2}X)^{-1}X^{\mathrm{T}}W_{\tau^2}^{1/2}\right\}(W_{\tau^2}^{1/2}y).$$

The profile log likelihood is therefore

$$\ell_{\mathrm{p}}(\tau^2) = -\frac{1}{2}\sum_{j=1}^n \log(v_j + \tau^2) - \frac{SS(\tau^2)}{2}, \quad \tau^2 \geq 0.$$