5.2

(a) Are the parameters exchangeable?
Yes, they are exchangeable. The joint distribution is

\[
p(\theta_1, \ldots, \theta_{2J}) = \left( \frac{2J}{J} \right)^{-1} \sum_p \left( \prod_{j=1}^{J} N(\theta_{p(j)}|1, 1) \prod_{j=J+1}^{2J} N(\theta_{p(j)}|1, 1) \right)
\]

where the sum is over all permutations \( p \) of \((1, \ldots, 2J)\). The density is obviously invariant to permutations of the indexes \((1, \ldots, 2J)\).

(b) Show that the distribution cannot be written as a mixture of iid components.
Pick any \( i, j \). The covariance of \( \theta_i, \theta_j \) is negative. You can see this because if \( \theta_i \) is large, then it probably comes from the \( N(1, 1) \) distribution, which means that it is more likely than not that \( \theta_j \) comes from the \( N(1, 1) \) distribution (because we know that exactly half of the \( 2J \) parameters come from each of the two distributions), which means that \( \theta_j \) will probably be negative. Conversely, if \( \theta_i \) is negative, then \( \theta_j \) is most likely positive.

Then, by Exercise 5.3, \( p(\theta_1, \ldots, \theta_{2J}) \) cannot be written as a mixture of iid components.

The above argument can be made formal and rigorous by defining \( \phi_1, \ldots, \phi_{2J} \), where half of the \( \phi_j \)'s are 1 and half are -1, and then setting \( \theta_j | \phi_j \sim N(\phi_j, 1) \).

\[
cov(\theta_i, \theta_j) = E(\theta_i \theta_j) - E(\theta_i)E(\theta_j)
= E(E(\theta_i \theta_j | (\phi_i, \phi_j))) - E(E(\theta_i | \phi_i))E(E(\theta_j | \phi_j))
\]

\[
= E(\theta_i \theta_j | (\phi_i = 1, \phi_j = 1))p(\phi_i = 1, \phi_j = 1)
+ E(\theta_i \theta_j | (\phi_i = 1, \phi_j = -1))p(\phi_i = 1, \phi_j = -1)
+ E(\theta_i \theta_j | (\phi_i = -1, \phi_j = 1))p(\phi_i = -1, \phi_j = 1)
+ E(\theta_i \theta_j | (\phi_i = -1, \phi_j = -1))p(\phi_i = -1, \phi_j = -1)
- [E(\theta_i | \phi_i = 1)p(\phi_i = 1) + E(\theta_i | \phi_i = -1)p(\phi_i = -1)] \\
\cdot [E(\theta_j | \phi_j = 1)p(\phi_j = 1) + E(\theta_j | \phi_j = -1)p(\phi_j = -1)]
\]

\[
= 1 \cdot \frac{J}{2J} \cdot \frac{J-1}{2J-1} + (-1) \cdot \frac{J}{2J} \cdot \frac{J}{2J-1} \\
\cdot \frac{J-1}{2J-1}
\]

\[
= \frac{-1}{2J-1} < 0
\]

So the statement is proved.

(c) Show the reason why it cannot be a counter example.
In the limit as \( J \to \infty \), the negative correlation between \( \theta_i \) and \( \theta_j \) approaches zero, and the joint distribution approaches iid. To put it another way, as \( J \to \infty \), the distinction disappears between (1) independently assigning each \( \theta_j \) to one of two groups, and (2) picking exactly half of the \( \theta_j \)'s for each group.
5.3 Prove that the covariances are all nonnegative.

Let $\mu(\phi) = E(\theta_j|\phi)$. From (1.8) on page 24 (also see Exercise 1.2),

$$\text{cov}(\theta_i, \theta_j) = E(\text{cov}(\theta_i, \theta_j)|\phi) + \text{cov}(E(\theta_i|\phi), E(\theta_j|\phi))$$

$$= 0 + \text{cov}(\mu(\phi), \mu(\phi))$$

$$= \text{var}(\mu(\phi))$$

$$\geq 0$$

The first term is equal to zero because given $\phi$, $\theta_i$ and $\theta_j$ for every $i, j$ are independent according the the pdf.

5.7

(a) Show the unnormalized posterior density has an infinite integral.

Consider the limit $(\alpha + \beta) \to \infty$ with $\alpha/\beta$ fixed at any nonzero value. The likelihood (see equation (5.8)) is

$$p(y | \alpha, \beta) \propto \prod_{j=1}^{J} \frac{[\alpha \cdots (\alpha + y_j - 1)][\beta \cdots (\beta + n_j - y_j - 1)]}{(\alpha + \beta) \cdots (\alpha + \beta + n_j - 1)}$$

which is a constant (if we are considering $y$, $n$, and $\alpha/\beta$ to be fixed), so the prior density determines whether the posterior density has a finite integral in this limit. A uniform prior density on $(\log(\alpha/\beta), \log(\alpha + \beta))$ has an infinite integral in this limit, and so the posterior density does also in this case.

(b) Show the combination yields the prior density (5.10).

The prior here we use is

$$p\left(\frac{\alpha}{\alpha + \beta}, (\alpha + \beta)^{-1/2}\right) \propto 1.$$ 

Let

$$u = \frac{\alpha}{\alpha + \beta}, v = (\alpha + \beta)^{-1/2}.$$ 

To get the form in (5.10), let $g = \log(\alpha/\beta)$, $h = \log(\alpha + \beta)$, then

$$u = \frac{e^g}{1 + e^g}, v = e^{-h/2}.$$ 

We need to show

$$p(g, h) = p(u, v) |J| \propto \alpha \beta (\alpha + \beta)^{-5/2},$$

where the Jacobian of the transformation is

$$\begin{vmatrix} \frac{\partial u}{\partial g} & \frac{\partial u}{\partial h} \\ \frac{\partial v}{\partial g} & \frac{\partial v}{\partial h} \end{vmatrix} = \begin{vmatrix} -\frac{e^g}{(1 + e^g)^2} & 0 \\ 0 & -\frac{1}{2} e^{-h/2} \end{vmatrix} = \frac{1}{2} \frac{e^{g-h/2}}{(1 + e^g)^2} = \frac{1}{2} \alpha \beta (\alpha + \beta)^{-5/2}.$$
(c) Show that the resulting posterior density is proper if the condition holds.

There are 4 limits to consider:

1. $\alpha \to 0$ with $\alpha + \beta$ fixed
2. $\beta \to 0$ with $\alpha + \beta$ fixed
3. $\alpha + \beta \to 0$ with $\alpha/\beta$ fixed
4. $\alpha + \beta \to \infty$ with $\alpha/\beta$ fixed

As in Exercise 5.7a, we work with expression (5.8). We have to show that the integral is finite in each of these limits.

Let $J_0$ be the number of experiments with $y_j > 0$, $J_1$ be the number of experiments with $y_j < n_j$, and $J_{01}$ be the number of experiments with $0 < y_j < n_j$.

1. For $\alpha \to 0$, proving convergence is trivial. All the factors in the likelihood (1) go to constants except $\alpha^{J_0}$, so the likelihood goes to 0 (if $J_0 > 0$) or a constant (if $J_0 = 0$). The prior distribution is a constant as $\alpha$ goes to 0. So whether $J_0 = 0$ or $J_0 > 0$, the integral of the posterior density in this limit is finite.

2. For $\beta \to 0$, same proof.

3. First transform to $(\frac{\alpha}{\alpha + \beta}, \alpha + \beta)$, so we only have to worry about the limit in one dimension. Multiplying the prior distribution (5.9) by the Jacobian yields

$$p\left(\frac{\alpha}{\alpha + \beta}, \alpha + \beta\right) \propto (\alpha + \beta)^{-3/2}$$

For $\alpha + \beta \to 0$, the likelihood looks like

$$p(y \mid \alpha, \beta)/(\alpha + \beta)^{-J_{01}},$$

ignoring all factors such as $\gamma(\alpha + \beta + n_j)$ and $\frac{\alpha}{\alpha + \beta}$ that are constant in this limit. So the posterior density in this parameterization is

$$p\left(\frac{\alpha}{\alpha + \beta}, \alpha + \beta \mid y\right) \propto (\alpha + \beta)^{-3/2}(\alpha + \beta)^{J_{01}}.$$  \hspace{1cm} (2)

The function $x^c$ has a finite integral as $x \to 0$ if $c > -1$, so (2) has a finite integral if $J_{01} > \frac{1}{2}$.

Note: the statistical reasoning here is that if $J_{01} = 0$ (so that all the data are of “0 successes out of $n_j$” or “$n_j$ successes out of $n_j$”), then it is still possible that $\alpha = \beta = 0$ (corresponding to all the $\theta_j$’s being 0 or 1), and we have to worry about the infinite integral of the improper prior distribution in the limit of $(\alpha + \beta) \to 0$. If $J_{01} > 0$, the likelihood gives the information that $(\alpha + \beta) = 0$ is not possible. In any case, values of $\alpha$ and $\beta$ near 0 do not make much sense in the context of the problem (modeling rat tumor rates), and it might make sense to just constrain $\alpha \geq 1$ and $\beta \geq 1$.

4. For $\alpha + \beta \to \infty$, the likelihood is constant, and so we just need to show that the prior density has a finite integral (see the solution to Exercise 5.7a). As above,

$$p\left(\frac{\alpha}{\alpha + \beta}, \alpha + \beta\right) \propto (\alpha + \beta)^{-3/2},$$

which indeed has a finite integral as $(\alpha + \beta) \to \infty$. 
5.8 We first note that, since \( p(\mu, \tau | y) \) and \( p(\theta | \mu, \tau, y) \) have proper distributions, the joint posterior density \( p(\theta, \mu, \tau | y) \) is proper if and only if the marginal posterior density \( p(\tau | y) \) from (5.21) is proper—that is, has a finite integral for \( \tau \) from 0 to \( \infty \).

(a) Show the posterior is improper.

Everything multiplying \( p(\tau) \) in (5.21) approaches a nonzero constant limit as \( \tau \) tends to zero; call that limit \( C(y) \). Thus the behavior of the posterior density near \( \tau = 0 \) is determined by the prior density. The function \( p(\tau) \propto 1/\tau \) is not integrable for any small interval including \( \tau = 0 \), and so it leads to a nonintegrable posterior density. (We can make this more formal: for any \( \delta > 0 \) we can identify an interval including zero on which \( p(\tau | y) \geq p(\tau)(C - \delta) \).

(b) Show the posterior is proper if \( J > 2 \).

The argument from 5.8a shows that if \( p(\tau) \propto 1 \) then the posterior density is integrable near zero. We need to examine the behavior as \( \tau \to \infty \) and find an upper bound that is integrable. The exponential term is clearly less than or equal to 1. We can rewrite the remaining terms as \( \prod_{j=1}^{J} ((\sum_{K\neq j} \sigma_k^2 + \tau^2)^{-1/2} \). For \( \tau > 1 \) we make this quantity bigger by dropping all of the \( \sigma^2 \) to yield \( \prod_{j=1}^{J} (J\tau^{2(J-1)})^{-1/2} \). An upper bound on \( p(\tau | y) \) for \( \tau \) large is \( p(\tau)J^{-1/2}/\tau^{J-1} \). When \( p(\tau) \propto 1 \), this upper bound is integrable if \( J > 2 \), and so \( p(\tau | y) \) is integrable if \( J > 2 \).

(c) Analyze SAT coaching data.

There are several reasonable options here. One approach is to abandon the hierarchical model and just fit the two schools with independent noninformative prior distributions (as in Exercise 3.3 but with the variances known).

Another approach would be to continue with the uniform prior distribution on \( \tau \) and try to analytically work through the difficulties with the resulting improper posterior distribution. In this case, the posterior distribution has all its mass near \( \tau = \infty \), meaning that no shrinkage will occur in the analysis (see (5.17) and (5.20)). This analysis is in fact identical to the first approach of analyzing the schools independently.

If the analysis based on noninformative prior distributions is not precise enough, it might be worthwhile assigning an informative prior distribution to \( \tau \) (or perhaps to \( (\mu, \tau) \)) based on outside knowledge such as analyses of earlier coaching experiments in other schools. However, one would have to be careful here: with data on only two schools, inferences would be sensitive to prior assumptions that would be hard to check from the data at hand.

5.9

(a) Write the joint posterior density.

\[
p(\theta, \mu, \tau | y) \propto p(\theta | \mu, \tau)p(y | \theta, \mu, \tau),
\]

where we may note that the latter density is really just \( p(y | \theta) \). Since \( p(\theta, \mu, \tau) = p(\theta | \mu, \tau)p(\mu, \tau) \),

\[
p(\theta, \mu, \tau | y) \propto p(\theta, \mu, \tau) \prod_{j=1}^{J} \left[ \theta_j^{-1}(1 - \theta_j)^{-1} \tau^{-1} \exp\left(-\frac{1}{2}(\logit(\theta_j) - \mu)^2/\tau^2\right) \prod_{j=1}^{J} \theta_j^{n_j}(1 - \theta_j)^{n_j-n_j} \right].
\]

(The factor \( \theta_j^{-1}(1 - \theta_j)^{-1} \) is \( d(\logit(\theta_j))/d\theta_j \); we need this since we want to have \( p(\theta | \mu, \tau) \) rather than \( p(\logit(\hat{\theta}) | \mu, \tau) \).)
(b) Show the integral (5.4) has no closed form expression.

Even though we can look at each of the $J$ integrals individually, the integrand separates into independent factors; there is no obvious analytic technique which permits evaluation of these integrals. In particular, we cannot recognize a multiple of a familiar density function inside the integrals. One might try to simplify matters with a substitution like $u_j = \text{logit}(\theta_j)$ or $v_j = \theta_j/(1 - \theta_j)$, but neither substitution turns out to be helpful.

(c) Explain the reason.

In order for expression (5.5) to be useful, we would have to know $p(\theta \mid \mu, \tau, y)$. Knowing it up to proportionality in $\theta$ is insufficient because our goal is to use this density to find another density that depends on $\mu$ and $\tau$. In the rat tumor example (page 128, equation (5.7)), the conjugacy of the beta distribution allowed us to write down the relevant “constant” of proportionality, which of course appears in equation (5.8).