Single Parameter Models:

Mixture of Conjugate Priors and Fairness of a Coin

Mixtures of Conjugate Priors

Suppose a special coin is known to have a significant bias, but we don't know if the coin is biased toward heads or tails. If p represents the probability that the coin lands heads, we believe that either p is in the neighborhood of 0.3 or in the neighborhood of 0.7 and it is equally likely that p is in one of the two neighborhoods. This belief can be modeled using the prior density

$$g(p) = \gamma g_1(p) + (1 - \gamma) g_2(p),$$

where g_1 is beta(6, 14), g_2 is beta(14, 6), and the mixing probability is $\gamma = 0.5$. Figure 3.5 displays this prior that reflects a belief in a biased coin.

Mixtures of Conjugate Priors

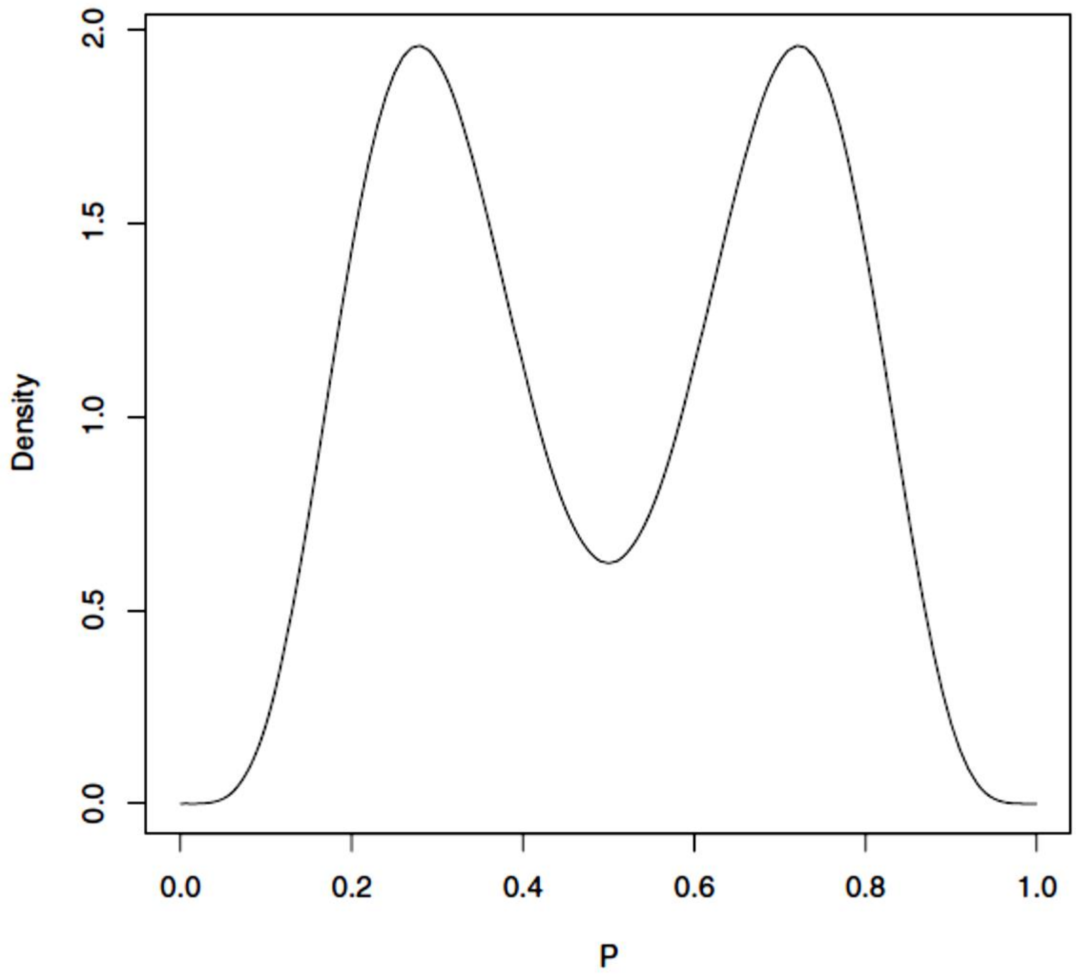

Fig. 3.5. Mixture of beta densities prior distribution that reflects belief that a coin is biased.

Mixtures of Conjugate Priors

In this situation, it can be shown that we have a conjugate analysis, as the prior and posterior distributions are represented by the same "mixture of betas" functional form. Suppose we flip the coin n times, obtaining s heads and $f = n - s$ tails. The posterior density of the proportion has the mixture form

$$g(p|\text{data}) = \gamma(\text{data})g_1(p|\text{data}) + (1 - \gamma(\text{data}))g_2(p|\text{data}),$$

where g_1 is beta$(6 + s,\ 14 + f)$, g_2 is beta$(14 + s,\ 6 + f)$, and the mixing probability $\gamma(\text{data})$ has the form

$$\gamma(\text{data}) = \frac{\gamma f_1(s, f)}{\gamma f_1(s, f) + (1 - \gamma) f_2(s, f)},$$

where $f_j(s, f)$ is the prior predictive probability of s heads in n flips when p has the prior density g_j.

Mixtures of Conjugate Priors

Suppose we flip the coin ten times and obtain seven heads and three tails. From the R output, we see that the posterior distribution of p is given by the beta mixture

$$g(p|\text{data}) = 0.093\,\text{beta}(13, 17) + 0.907\,\text{beta}(21, 9).$$

The prior and posterior densities for the proportion are displayed (using several `curve` commands) in Figure 3.6. Initially we were indifferent to the direction of the bias of the coin, and each component of the beta mixture had the same weight. Since a high proportion of heads was observed, there is evidence that the coin is biased toward heads and the posterior density places a greater weight on the second component of the mixture.

Mixtures of Conjugate Priors

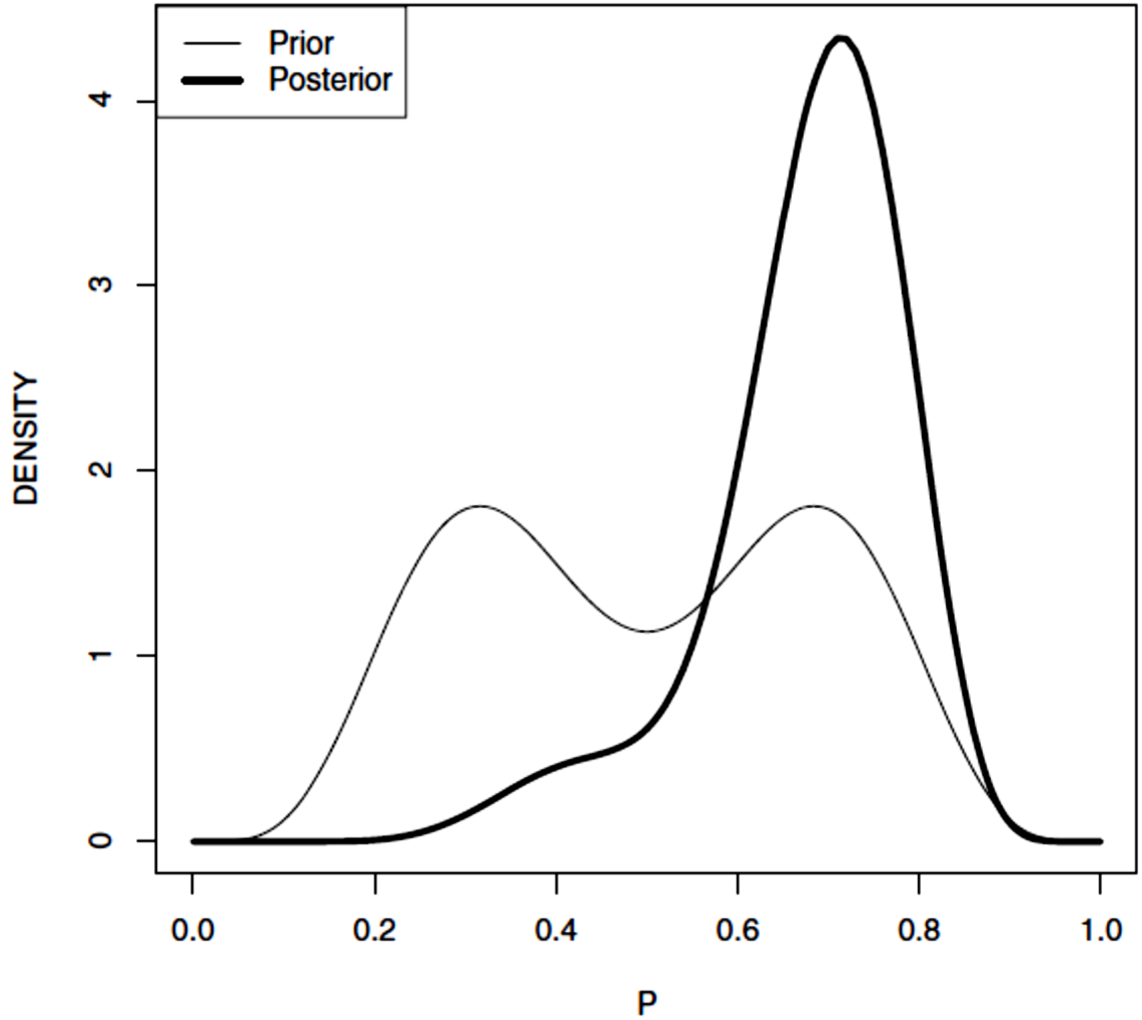

Fig. 3.6. Prior and posterior densities of a proportion for the biased coin example.

A Bayesian Test of the Fairness of a Coin

Mixture of priors is useful in the development of a Bayesian test of two hypotheses about a parameter. Suppose you are interested in assessing the fairness of a coin. You observe y binomially distributed with parameters n and p, and you are interested in testing the hypothesis H that $p = .5$. If y is observed, then it is usual practice to make a decision on the basis of the p-value

$$2 \times \min\{P(Y \leq y), P(Y \geq y)\}.$$

If this p-value is *small*, then you reject the hypothesis H and conclude that the coin is not fair. Suppose, for example, the coin is flipped 20 times and only 5 heads are observed. In R we compute the probability of obtaining five or fewer heads.

A Bayesian Test of the Fairness of a Coin

Let's consider this problem from a Bayesian perspective. There are two possible models here – either the coin is fair ($p = .5$) or the coin is not fair ($p \neq .5$). Suppose that you are indifferent between the two possibilities, so you initially assign each model a probability of 1/2. Now, if you believe the coin is fair, then your entire prior distribution for p would be concentrated on the value $p = .5$. If instead the coin is unfair, you would assign a different prior distribution on (0, 1), call it $g_1(p)$, that would reflect your beliefs about the probability of an unfair coin . Suppose you assign a beta(a, a) prior on p. This beta distribution is symmetric about .5 – it says that you believe the coin is not fair, and the probability is close to $p = .5$. To summarize, your prior distribution in this testing situation can be written as the mixture

$$g(p) = .5I(p = .5) + .5I(p \neq .5)g_1(p),$$

where $I(A)$ is an indicator function equal to 1 if the event A is true and otherwise is equal to 0.

A Bayesian Test of the Fairness of a Coin

After observing the number of heads in n tosses, we would update our prior distribution by Bayes' rule. The posterior density for p can be written as

$$g(p|y) = \lambda(y)I(p = .5) + (1 - \lambda(y))g_1(p|y),$$

where g_1 is a beta$(a+y, a+n-y)$ density and $\lambda(y)$ is the posterior probability of the model where the coin is fair,

$$\lambda(y) = \frac{.5p(y|.5)}{.5p(y|.5) + .5m_1(y)}.$$

In the expression for $\lambda(y)$, $p(y|.5)$ is the binomial density for y when $p = .5$, and $m_1(y)$ is the (prior) predictive density for y using the beta density.

A Bayesian Test of the Fairness of a Coin

In R, the posterior probability of fairness $\lambda(y)$ is easily computed. The R command **`dbinom`** will compute the binomial probability $p(y|.5)$, and the predictive density for y can be computed using the identity

$$m_1(y) = \frac{f(y|p)g_1(p)}{g_1(p|y)}.$$

Assume first that we assign a beta(10, 10) prior for p when the coin is not fair and we observe $y = 5$ heads in $n = 20$ tosses. The posterior probability of fairness is stored in the R variable **`lambda`**.

Mixtures of Conjugate Priors

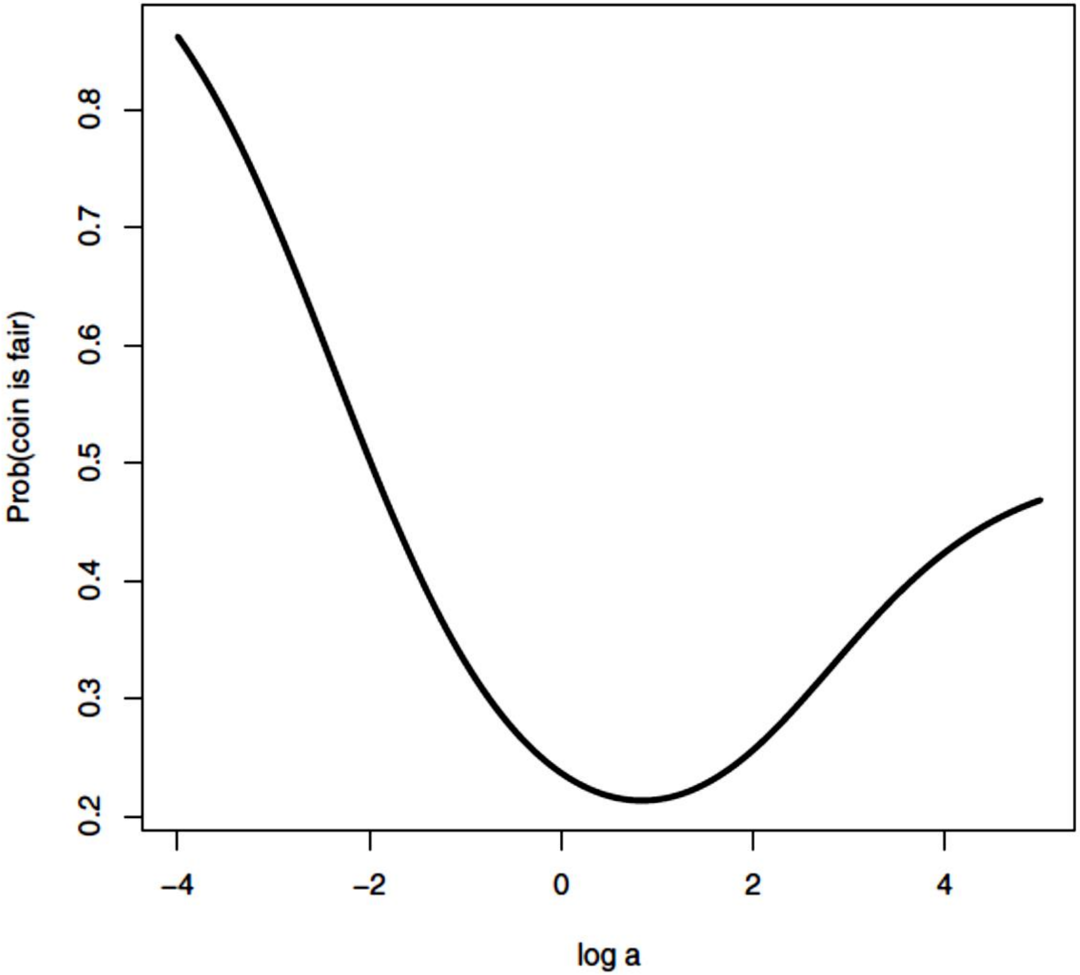

Fig. 3.7. Posterior probability that a coin is fair graphed against values of the prior parameter $\log a$.

A Bayesian Test of the Fairness of a Coin

Another distinction between the frequentist and Bayesian calculations is the event that led to the decision about rejecting the hypothesis that the coin was fair. The *p*-value calculation was based on the probability of the event "5 heads or fewer," but the Bayesian calculation was based solely on the likelihood based on the event "exactly 5 heads." That raises the question: How would the Bayesian answers change if we observed "5 heads or fewer"? One can show that the posterior probability that the coin is fair is given by

$$\lambda(y) = \frac{.5P_0(Y \leq 5)}{.5P_0(Y \leq 5) + .5P_1(Y \leq 5)},$$

where $P_0(Y \leq 5)$ is the probability of five heads or fewer under the binomial model with $p = .5$ and $P_1(Y \leq 5)$ is the predictive probability of this event under the alternative model with a beta$(10, 10)$ prior on p. In the following R output, the cumulative probability of five heads under the binomial model is computed by the R function `pbinom`. The probability of five or fewer heads under the alternative model is computed by summing the predictive density over the six values of y.