```
     ybar    mu1 t   tau1 t
[1,]  110 105.2921 5.841676
[2,]  125 118.0841 7.885174
[3,]  140 135.4134 7.973498
```

Let's compare the posterior moments of θ using the two priors by combining the two R matrices `summ1` and `summ2`.

```
> cbind(summ1,summ2)

     ybar      mu1     tau1 ybar    mu1 t   tau1 t
[1,]  110 107.2442 6.383469  110 105.2921 5.841676
[2,]  125 118.1105 6.383469  125 118.0841 7.885174
[3,]  140 128.9768 6.383469  140 135.4134 7.973498
```

When $\bar{y} = 110$, the values of the posterior mean and posterior standard deviation are similar using the normal and t priors. However, there can be substantial differences in the posterior moments using the two priors when the observed mean score is inconsistent with the prior mean. In the "extreme" case where $\bar{y} = 140$, Figure 3.4 graphs the posterior densities for the two priors.

```
> theta=seq(60, 180, length=500)
> normpost = dnorm(theta, mu1[3], tau1)
> normpost = normpost/sum(normpost)
> plot(theta,normpost,type="l",lwd=3,ylab="Posterior Density")
> like = dnorm(theta,mean=140,sd=sigma/sqrt(n))
> prior = dt((theta - mu)/tscale, 2)
> tpost = prior * like / sum(prior * like)
> lines(theta,tpost)
> legend("topright",legend=c("t prior","normal prior"),lwd=c(1,3))
```

When a normal prior is used, the posterior will always be a compromise between the prior information and the observed data, even when the data result conflicts with one's prior beliefs about the location of Joe's IQ. In contrast, when a t prior is used, the likelihood will be in the flat-tailed portion of the prior and the posterior will resemble the likelihood function.

In this case, the inference about the mean is robust to the choice of prior (normal or t) when the observed mean IQ score is consistent with the prior beliefs. But in the case where an extreme IQ score is observed, we see that the inference is not robust to the choice of prior density.

3.5 Mixtures of Conjugate Priors

In the binomial, Poisson, and normal sampling models, we have illustrated the use of a conjugate prior where the prior and posterior distributions have the same functional form. One straightforward way to extend the family of

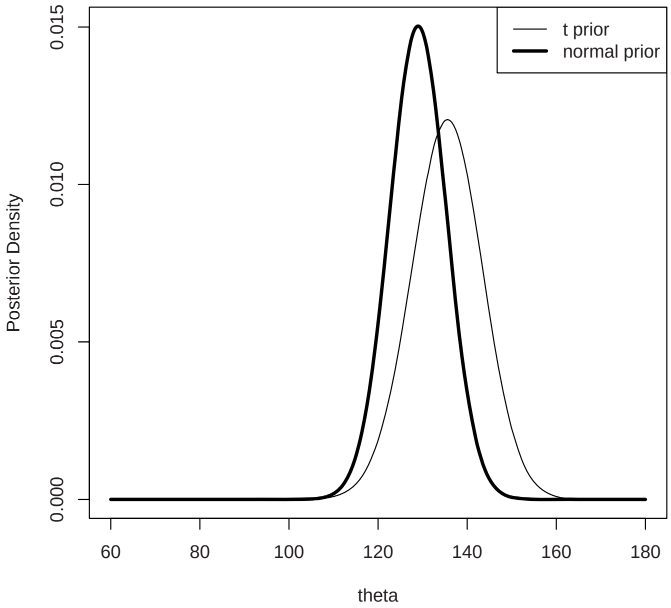

Fig. 3.4. Posterior densities for a person's true IQ using normal and t priors for an extreme observation.

conjugate priors is by using discrete mixtures. Here we illustrate the use of a mixture of beta densities to learn about the probability that a biased coin lands heads.

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.

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

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

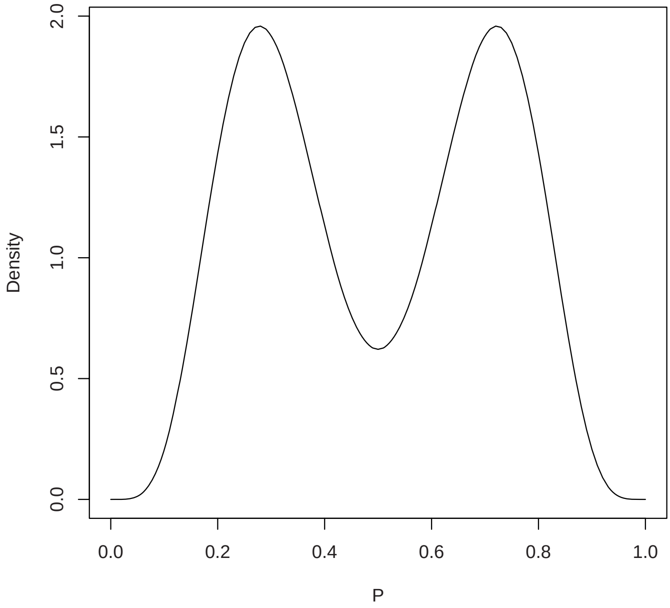

Fig. 3.5. Mixture of beta densities prior distribution that reflects belief that a coin is biased.

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.

The R function `binomial.beta.mix` computes the posterior distribution when the proportion p has a mixture of betas prior distribution. The inputs to this function are `probs`, the vector of mixing probabilities; `betapar`, a matrix of beta shape parameters where each row corresponds to a component of the prior; and `data`, the vector of the number of successes and number of failures in the sample. The output of the function is a list with two components – `probs` is a vector of posterior mixing probabilities and `betapar` is a matrix containing the shape parameters of the updated beta posterior densities.

```
> probs=c(.5,.5)
> beta.par1=c(6, 14)
> beta.par2=c(14, 6)
```

```
> betapar=rbind(beta.par1, beta.par2)
> data=c(7,3)
> post=binomial.beta.mix(probs,betapar,data)
> post

$probs
 beta.par1  beta.par2
0.09269663 0.90730337


$betapar
          [,1] [,2]
beta.par1   13   17
beta.par2   21    9
```

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.

```
> curve(post$probs[1]*dbeta(x,13,17)+post$probs[2]*dbeta(x,21,9),
+  from=0, to=1, lwd=3, xlab="P", ylab="DENSITY")
> curve(.5*dbeta(x,6,12)+.5*dbeta(x,12,6),0,1,add=TRUE)
> legend("topleft",legend=c("Prior","Posterior"),lwd=c(1,3))
```

3.6 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.