```
> py = dpois(y, lam * ex) * dgamma(lam, shape = alpha,
+     rate = beta)/dgamma(lam, shape = alpha + y,
+     rate = beta + ex)
> cbind(y, round(py, 3))

       y
 [1,]  0 0.172
 [2,]  1 0.286
 [3,]  2 0.254
 [4,]  3 0.159
 [5,]  4 0.079
 [6,]  5 0.033
 [7,]  6 0.012
 [8,]  7 0.004
 [9,]  8 0.001
[10,]  9 0.000
[11,] 10 0.000

> lambdaB = rgamma(1000, shape = alpha + yobs, rate = beta + ex)
```

To see the impact of the prior density on the inference, it is helpful to display the prior and posterior distributions on the same graph. In Figure 3.2, density estimates of the simulated draws from the posterior distributions of the rates are shown for hospitals A and B. The gamma prior density is also displayed in each case. We see that for hospital A, with relatively little experience in surgeries, the prior information is significant and the posterior distribution resembles the prior distribution. In contrast, for hospital B, with many surgeries, the prior information is less influential and the posterior distribution resembles the likelihood function.

```
> par(mfrow = c(2, 1))
> plot(density(lambdaA), main="HOSPITAL A",
+   xlab="lambdaA", lwd=3)
> curve(dgamma(x, shape = alpha, rate = beta), add=TRUE)
> legend("topright",legend=c("prior","posterior"),lwd=c(1,3))
> plot(density(lambdaB), main="HOSPITAL B",
+   xlab="lambdaB", lwd=3)
> curve(dgamma(x, shape = alpha, rate = beta), add=TRUE)
> legend("topright",legend=c("prior","posterior"),lwd=c(1,3))
```

3.4 An Illustration of Bayesian Robustness

In practice, one may have incomplete prior information about a parameter in the sense that one's beliefs won't entirely define a prior density. There may be a number of different priors that match the given prior information. For

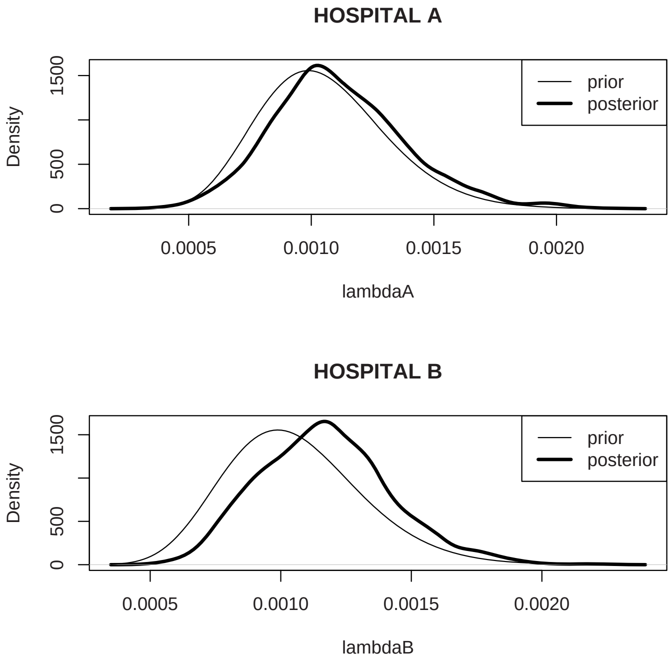

Fig. 3.2. Prior and posterior densities for heart transplant death rate for two hospitals.

example, if you believe a priori that the median of a parameter θ is 30 and its 80th percentile is 50, certainly there are many prior probability distributions that can be chosen that match these two percentiles. In this situation where different priors are possible, it is desirable that inferences from the posterior not be dependent on the exact functional form of the prior. A Bayesian analysis is said to be *robust* to the choice of prior if the inference is insensitive to different priors that match the user's beliefs.

To illustrate this idea, suppose you are interested in estimating the true IQ θ for a person we'll call Joe. You believe Joe has average intelligence, and the median of your prior distribution is 100. Also, you are 90% confident that Joe's IQ falls between 80 and 120. By using the function `normal.select`, we find the values of the mean and standard deviation of the normal density that match the beliefs that the median is 100 and the 95th percentile is 120.

```
quantile1=list(p=.5,x=100); quantile2=list(p=.95,x=120)
normal.select(quantile1, quantile2)

$mu
[1] 100
```

```
$sigma
[1] 12.15914
```

We see from the output that the normal density with mean $\mu = 100$ and $\tau = 12.16$ matches this prior information.

Joe takes four IQ tests and his scores are y_1, y_2, y_3, y_4. Assuming that an individual score y is distributed as $N(\theta, \sigma)$ with known standard deviation $\sigma = 15$, the observed mean score $\bar{y}$ is $N(\theta, \sigma/\sqrt{4})$.

With the use of a normal prior in this case, the posterior density of θ will also have the normal functional form. Recall that the precision is defined as the inverse of the variance. Then the posterior precision $P_1 = 1/\tau_1^2$ is the sum of the data precision $P_D = n/\sigma^2$ and the prior precision $P = 1/\tau^2$,

$$P_1 = P_D + P = 4/\sigma^2 + 1/\tau^2,$$

The posterior standard deviation is given by

$$\tau_1 = 1/\sqrt{P_1} = 1/(\sqrt{4/\sigma^2 + 1/\tau^2}).$$

The posterior mean of θ can be expressed as a weighted average of the sample mean and the prior mean where the weights are proportional to the precisions:

$$\mu_1 = \frac{\bar{y}P_D + \mu P}{P_D + P} = \frac{\bar{y}(4/\sigma^2) + \mu(1/\tau^2)}{4/\sigma^2 + 1/\tau^2}.$$

We illustrate the posterior calculations for three hypothetical test results for Joe. We suppose that the observed mean test score is $\bar{y} = 110$, $\bar{y} = 125$, or $\bar{y} = 140$. In each case, we compute the posterior mean and posterior standard deviation of Joe's true IQ θ. These values are denoted by the R variables `mu1` and `tau1` in the following output.

```
> mu = 100
> tau = 12.16
> sigma = 15
> n = 4
> se = sigma/sqrt(4)
> ybar = c(110, 125, 140)
> tau1 = 1/sqrt(1/se^2 + 1/tau^2)
> mu1 = (ybar/se^2 + mu/tau^2) * tau1^2
> summ1=cbind(ybar, mu1, tau1)
> summ1

     ybar      mu1     tau1
[1,]  110 107.2442 6.383469
[2,]  125 118.1105 6.383469
[3,]  140 128.9768 6.383469
```

Let's now consider an alternative prior density to model our beliefs about Joe's true IQ. Any symmetric density instead of a normal could be used, so we use a t density with location μ, scale τ, and 2 degrees of freedom. Since our prior median is 100, we let the median of our t density be equal to $\mu = 100$. We find the scale parameter τ, so the t density matches our prior belief that the 95th percentile of θ is equal to 120. Note that

$$P(\theta < 120) = P\left(T < \frac{20}{\tau}\right) = .95,$$

where T is a standard t variate with two degrees of freedom. It follows that

$$\tau = 20/t_2(.95),$$

where $t_v(p)$ is the pth quantile of a t random variable with v degrees of freedom. We find τ by using the t quantile function `qt` in R.

```
> tscale = 20/qt(0.95, 2)
> tscale

[1] 6.849349
```

We display the normal and t priors in a single graph in Figure 3.3. Although they have the same basic shape, note that the t density has significantly flatter tails – we will see that this will impact the posterior density for "extreme" test scores.

```
> par(mfrow=c(1,1))
> curve(1/tscale*dt((x-mu)/tscale,2),
+   from=60, to=140, xlab="theta", ylab="Prior Density")
> curve(dnorm(x,mean=mu,sd=tau), add=TRUE, lwd=3)
> legend("topright",legend=c("t density","normal density"),
+   lwd=c(1,3))
```

We perform the posterior calculations using the t prior for each of the possible sample results. Note that the posterior density of θ is given, up to a proportionality constant, by

$$g(\theta|\text{data}) \propto \phi(\bar{y}|\theta, \sigma/\sqrt{n}) g_T(\theta|v, \mu, \tau),$$

where $\phi(y|\theta, \sigma)$ is a normal density with mean θ and standard deviation σ, and $g_T(\mu|v, \mu, \tau)$ is a t density with median μ, scale parameter τ, and degrees of freedom v. Since this density does not have a convenient functional form, we summarize it using a direct "prior times likelihood" approach. We construct a grid of θ values that covers the posterior density, compute the product of the normal likelihood and the t prior on the grid, and convert these products to probabilities by dividing by the sum. Essentially we are approximating the continuous posterior density by a discrete distribution on

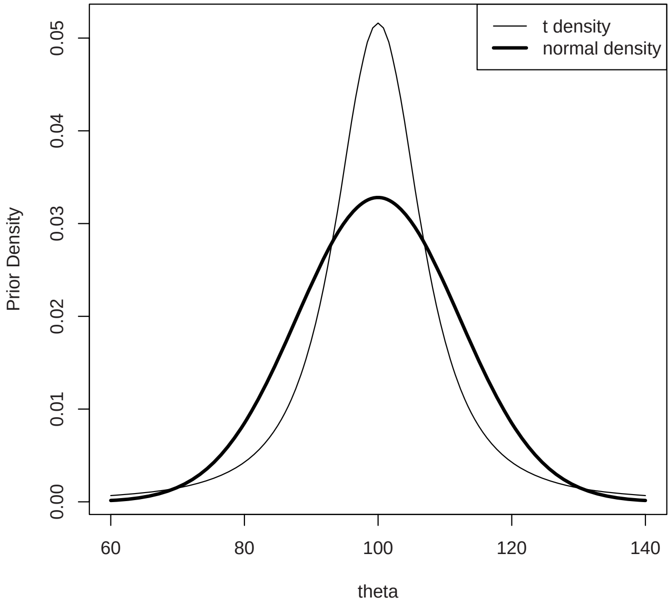

Fig. 3.3. Normal and t priors for representing prior opinion about a person's true IQ score.

this grid. We then use this discrete distribution to compute the posterior mean and posterior standard deviation. We first write a function `norm.t.compute` that implements this computational algorithm for a single value of $\bar{y}$. Then, using `sapply`, we apply this algorithm for the three values of $\bar{y}$, and the posterior moments are displayed in the second and third columns of the R matrix `summ2`.

```
> norm.t.compute=function(ybar) {
+      theta = seq(60, 180, length = 500)
+      like = dnorm(theta,mean=ybar,sd=sigma/sqrt(n))
+      prior = dt((theta - mu)/tscale, 2)
+      post = prior * like
+      post = post/sum(post)
+      m = sum(theta * post)
+      s = sqrt(sum(theta^2 * post) - m^2)
+      c(ybar, m, s) }
> summ2=t(sapply(c(110, 125, 140),norm.t.compute))
> dimnames(summ2)[[2]]=c("ybar","mu1 t","tau1 t")
> summ2
```

```
     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