Single Parameter Models

Normal Distribution, Known Mean, Unknown Variance

Gelman et al. (2003) consider a problem of estimating an unknown variance using American football scores. The focus is on the difference d between a game outcome (winning score minus losing score) and a published point spread. We observe $d_1, ..., d_n$, the observed differences between game outcomes and point spreads for n football games. If these differences are assumed to be a random sample from a normal distribution with mean 0 and unknown variance σ^2, the likelihood function is given by

$$L(\sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\sum_{i=1}^{n} d_i^2/(2\sigma^2)\right\}, \ \sigma^2 > 0.$$

Normal Distribution, Known Mean, Unknown Variance

Suppose the noninformative prior density $p(\sigma^2) \propto 1/\sigma^2$ is assigned to the variance. This is the standard vague prior placed on a variance – it is equivalent to assuming that the logarithm of the variance is uniformly distributed on the real line. Then the posterior density of σ^2 is given, up to a proportionality constant, by

$$g(\sigma^2|\text{data}) \propto (\sigma^2)^{-n/2-1} \exp\{-v/(2\sigma^2)\},$$

where $v = \sum_{i=1}^{n} d_i^2$. If we define the precision parameter $P = 1/\sigma^2$, then it can be shown that P is distributed as U/v, where U has a chi-squared distribution with n degrees of freedom. Suppose we are interested in a point estimate and a 95% probability interval for the standard deviation σ.

Normal Distribution, Known Mean, Unknown Variance

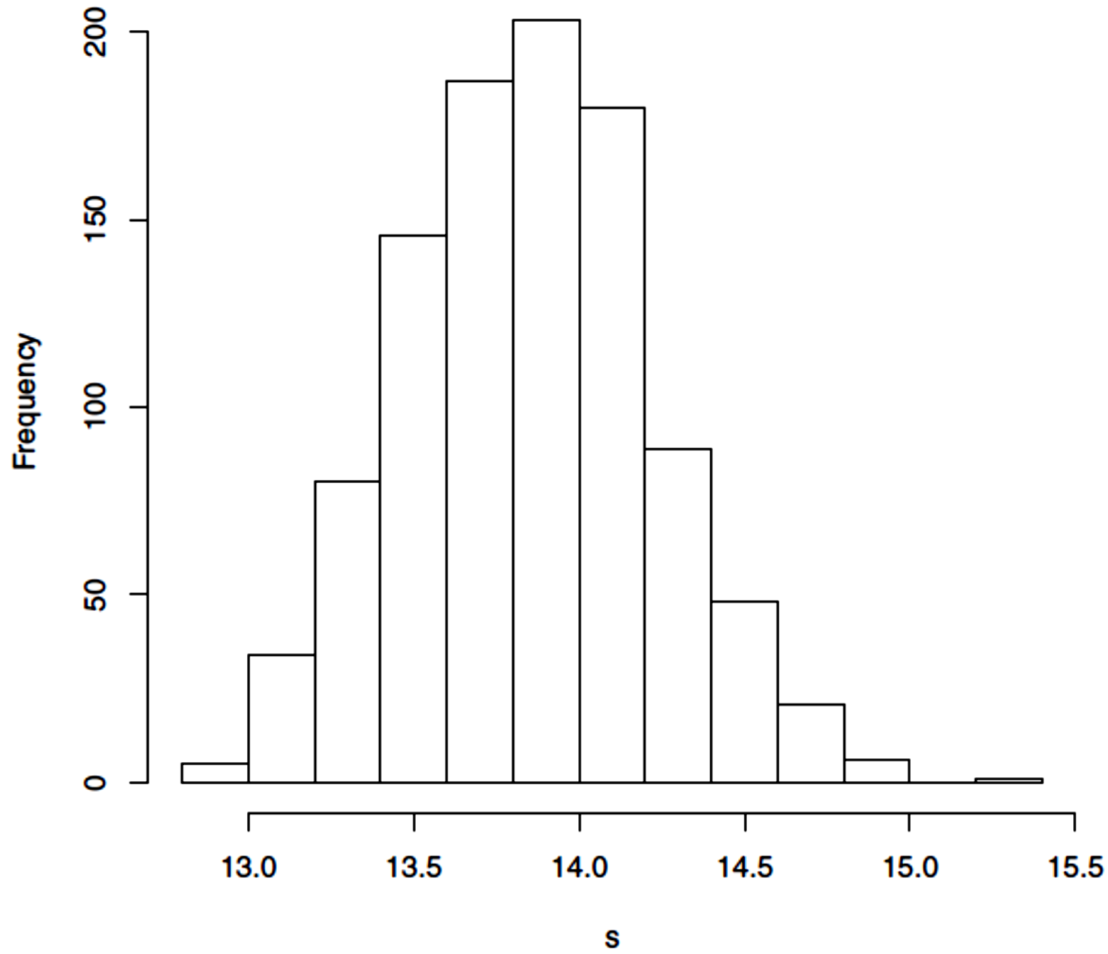

Fig. 3.1. Histogram of simulated sample of the standard deviation σ of differences between game outcomes and point spreads.

Estimating a Heart Transplant Mortality Rate

The standard estimate of λ is the maximum likelihood estimate $\hat{\lambda} = y/e$. Unfortunately, this estimate can be poor when the number of deaths y is close to zero. In this situation when small death counts are possible, it is desirable to use a Bayesian estimate that uses prior knowledge about the size of the mortality rate. A convenient choice for a prior distribution is a member of the gamma(α, β) density of the form

$$p(\lambda) \propto \lambda^{\alpha-1} \exp(-\beta\lambda), \ \lambda > 0.$$

Estimating a Heart Transplant Mortality Rate

A convenient source of prior information is heart transplant data from a small group of hospitals that we believe has the same rate of mortality as the rate from the hospital of interest. Suppose we observe the number of deaths z_j and the exposure o_j for ten hospitals $(j = 1, ..., 10)$, where z_j is Poisson with mean $o_j\lambda$. If we assign λ the standard noninformative prior $p(\lambda) \propto \lambda^{-1}$, then the updated distribution for λ, given these data from the ten hospitals, is

$$p(\lambda) \propto \lambda^{\sum_{j=1}^{10} z_j - 1} \exp\left(-(\sum_{j=1}^{10} o_j)\lambda \right).$$

Estimating a Heart Transplant Mortality Rate

Using this information, we have a gamma(α, β) prior for λ, where $\alpha = \sum_{j=1}^{10} z_j$ and $\beta = \sum_{j=1}^{10} o_j$. In this example, we have

$$\sum_{j=1}^{10} z_j = 16, \sum_{j=1}^{10} o_j = 15174,$$

and so we assign λ a gamma(16, 15174) prior.

Estimating a Heart Transplant Mortality Rate

If the observed number of deaths from surgery y_{obs} for a given hospital with exposure e is Poisson $(e\lambda)$ and λ is assigned the gamma(α, β) prior, then the posterior distribution will also have the gamma form with parameters $\alpha + y_{\text{obs}}$ and $\beta + e$. Also the (prior) predictive density of y (before any data are observed) can be computed using the formula

$$f(y) = \frac{f(y|\lambda)g(\lambda)}{g(\lambda|y)},$$

where $f(y|\lambda)$ is the Poisson$(e\lambda)$ sampling density and $g(\lambda)$ and $g(\lambda|y)$ are, respectively, the prior and posterior densities of λ.

Estimating a Heart Transplant Mortality Rate

By the model-checking strategy of Box (1980), both the posterior density $g(\lambda|y)$ and the predictive density $f(y)$ play important roles in a Bayesian analysis. By using the posterior density, one performs inference about the unknown parameter conditional on the Bayesian model that includes the assumptions of sampling density and the prior density. One can check the validity of the proposed model by inspecting the predictive density. If the observed data value y_{obs} is consistent with the predictive density $p(y)$, then the model seems reasonable. On the other hand, if y_{obs} is in the extreme tail portion of the predictive density, then this casts doubt on the validity of the Bayesian model, and perhaps the prior density or the sampling density has been misspecified.

Estimating a Heart Transplant Mortality Rate

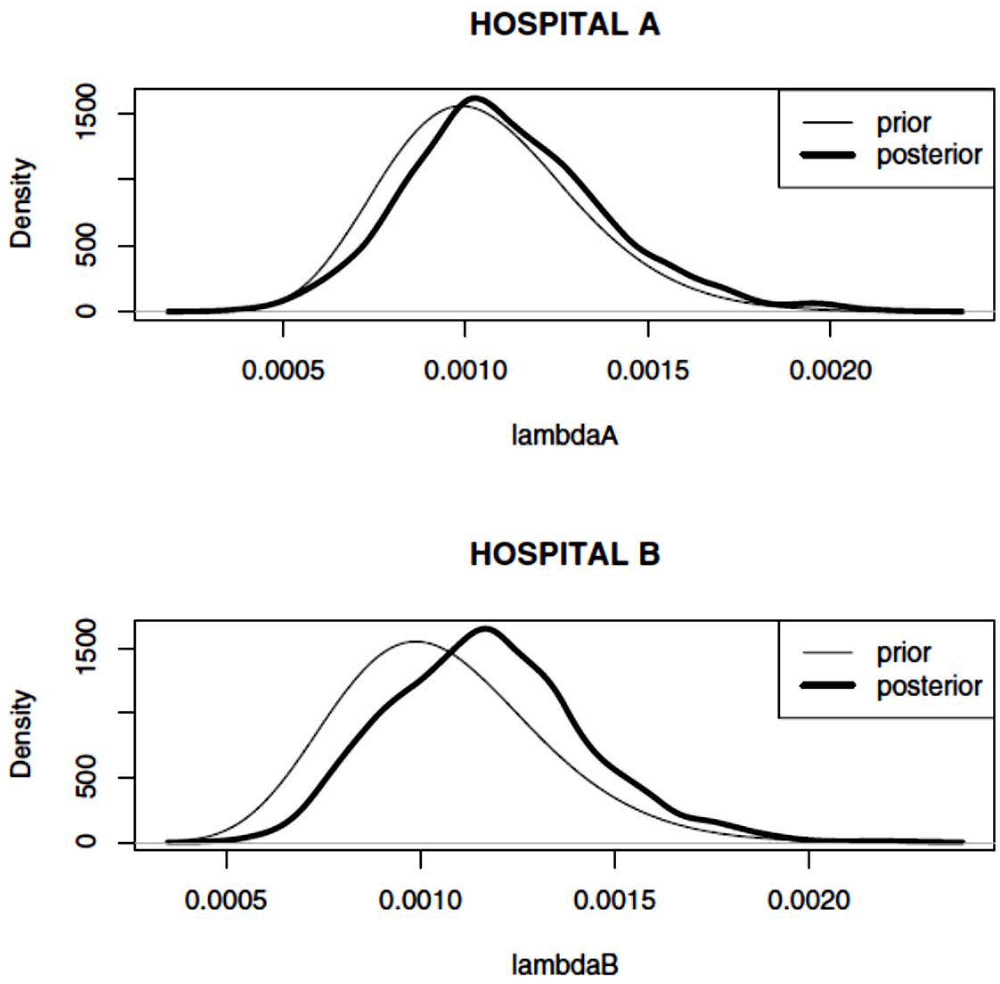

Fig. 3.2. Prior and posterior densities for heart transplant death rate for two hospitals.

An Illustration of Bayesian Robustness

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

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

An Illustration of Bayesian Robustness

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.

An Illustration of Bayesian Robustness

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.

An Illustration of Bayesian Robustness

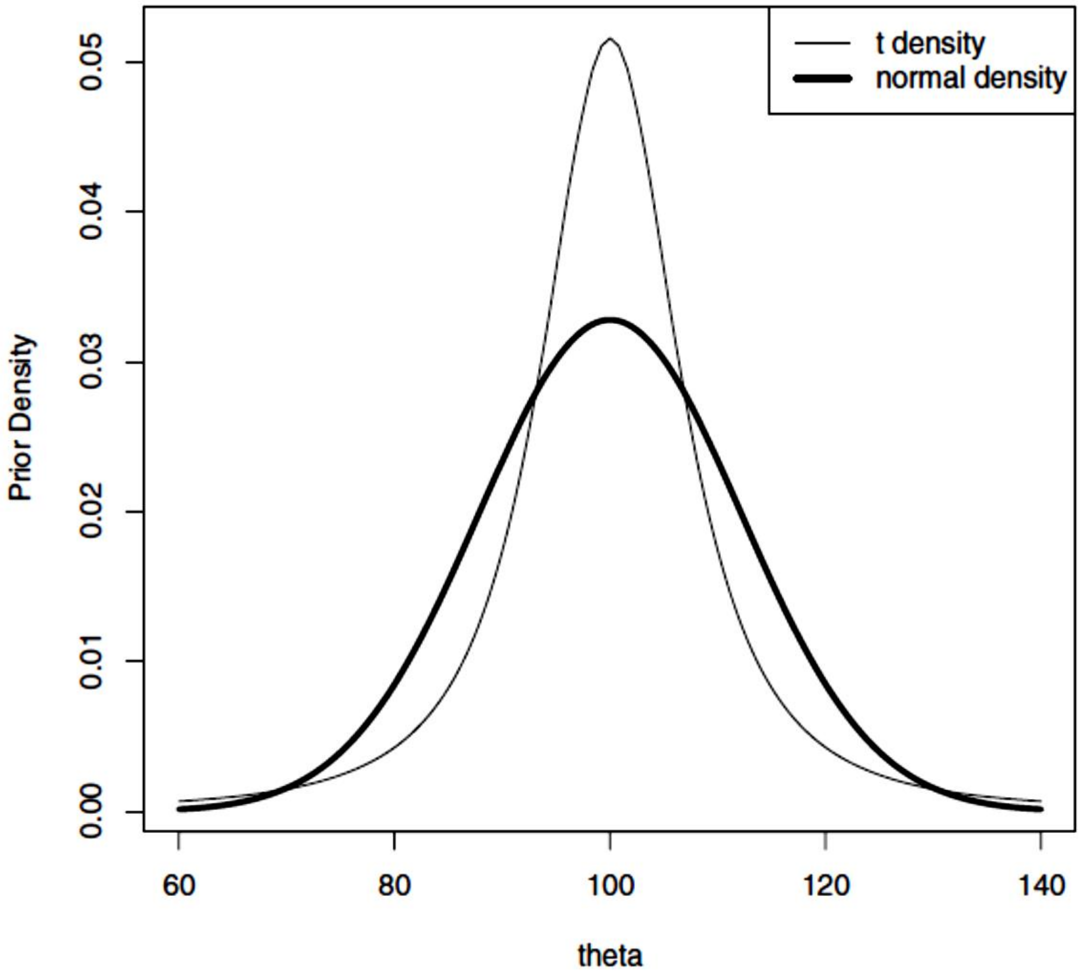

Fig. 3.3. Normal and t priors for representing prior opinion about a person's true IQ score.

An Illustration of Bayesian Robustness

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.

An Illustration of Bayesian Robustness

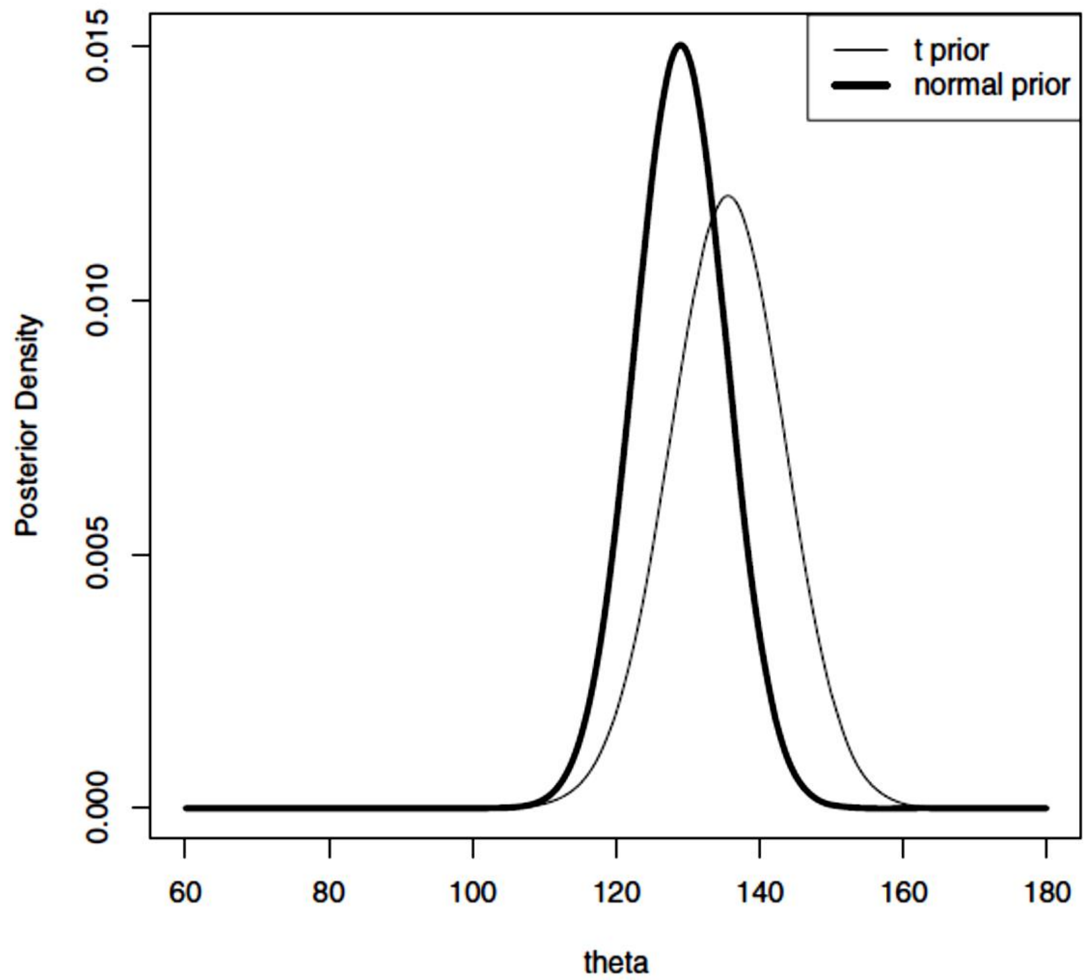

Fig. 3.4. Posterior densities for a person's true IQ using normal and t priors for an extreme observation.