Normal Random Effects Model

Statistics 220

Spring 2005

Normal-Normal Hierarchical Model

Have J independent groups, with known variance σ^2

$$y_{ij}|\theta_j \overset{ind}{\sim} N(\theta_j, \sigma^2),\ i = 1, \ldots, n_j;\ j = 1, \ldots, J$$

Except for the fixed measurement variance, this is the basis for the 1-way ANOVA model. So following the analysis for this model, the sample mean for each group be

$$\bar{y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

Its sampling variance is

$$\sigma_j^2 = \frac{\sigma^2}{n_j}$$

So

$$\bar{y}_{.j}|\theta_j \overset{ind}{\sim} N(\theta_j, \sigma_j^2)$$

For what follows, we are going to base it on the above normal model, independent observations with (potentially) different but known variances.

Note: In most situations, the assumption of known measurement error variances is dubious. However it is not always. The book discusses two examples where assuming that these variances are effective known is reasonable. Both involve situations where the data to be analyzed comes from summary measures from analyzes of large data sets.

If this assumption is not reasonable, we can put a prior distribution on σ^2. In this case, the analysis isn't quite as nice as what follows, but is tractable. We'll come back to it in Chapter 15.

We now need a model for $\theta_1, \ldots, \theta_J$. A popular choice is

$$\theta_j | \mu, \tau^2 \overset{iid}{\sim} N(\mu, \tau^2)$$

When combined with the original data model, this gives us the standard normal random effects model used in ANOVA.

Next we need to put a prior on μ and τ^2. While we could put an informative prior on these, say by following semi-conjugate ideas discussed earlier, lets follow the text and use a non-informative prior. For many problems fitting into this framework, the data swamps the prior in the analysis.

One reasonable choice is to have μ and τ^2 independent ($p(\mu, \tau^2) = p(\mu)p(\tau^2)$). With this, the obvious prior on μ is

$$p(\mu) \propto 1$$

i.e. uniform.

For τ^2, one valid choice is

$$p(\tau) \propto 1$$

i.e. again uniform.

Note that the Jeffreys' prior for $\tau (p(\log \tau) \propto 1, p(\tau) \propto \frac{1}{\tau})$ won't work as it leads to an improper posterior distribution.

- Joint posterior distribution

$$\begin{aligned} p(\theta, \mu, \tau|y) &\propto p(\mu, \tau)p(\theta|\mu, \tau)p(y|\theta) \\ &\propto p(\mu, \tau) \prod_{j=1}^{J} N(\theta_j|\mu, \tau^2) \prod_{j=1}^{J} N(\theta_j|\mu, \tau^2) \end{aligned}$$

- Conditional posterior distribution of the normal means θ_j

  Given the structure of the problem (independence of θ_j's given μ and τ and the independence of the $\bar{y}_{.j}$'s given the θ_j's), the conditional posterior $p(\theta|\mu,\tau,y)$ factors into J independent pieces.

  Notice that for each θ, this is similar to the case of a single normal mean with the conjugate prior.

$$\begin{aligned} p(\theta_j|\mu,\tau,y) &\propto p(\theta_j|\mu,\tau^2)p(\bar{y}_{.j}|\theta_j,\sigma^2) \\ &\propto N(\theta_j|\mu,\tau^2)N(\theta_j|\mu,\tau^2) \\ &= N(\theta_j|\hat{\theta}_j,V_j) \end{aligned}$$

  where

$$\hat{\theta}_j = \frac{\frac{1}{\sigma_j^2}\bar{y}_{.j} + \frac{1}{\tau^2}\mu}{\frac{1}{\sigma_j^2}+\frac{1}{\tau^2}} \qquad V_j = \frac{1}{\frac{1}{\sigma_j^2}+\frac{1}{\tau^2}}$$

- Marginal posterior distribution of the posterior distribution of the hyperparameters μ and τ

$$p(\mu, \tau|y) \propto p(\mu, \tau)p(y|\mu, \tau)$$

As the book mentions, this decomposition isn't usually helpful as $p(y|\mu, \tau)$ usually doesn't have a nice form. However for normal-normal model this can be determined as the integral

$$\begin{aligned} p(y|\mu, \tau) &= \int p(y, \theta|\mu, \tau)d\theta \\ &= \int \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right) \frac{1}{\tau\sqrt{2\pi}} \exp\left(-\frac{1}{2\tau^2}(\theta-\mu)^2\right) d\theta \end{aligned}$$

can be calculated and seen to be nice. Given the quadratic structure of the exponential piece, it must be a normal distribution. The integration can be done by completing the square for θ (giving a normal density to integrate out) or by getting the mean and variance of $y|\mu, \tau$ by

$$\begin{aligned} E[y] &= E[E[y|\theta]] = E[\theta] = \mu \\ \mathrm{Var}(y) &= \mathrm{Var}(E[y|\theta]) + E[\mathrm{Var}(y|\theta)] \\ &= \mathrm{Var}(\theta) + E[\sigma^2] = \tau^2 + \sigma^2 \end{aligned}$$

So

$$p(\mu, \tau|y) \propto p(\mu, \tau) \prod_{j=1}^{J} N(\theta_j|\mu, \sigma_j^2 + \tau^2)$$

Note: In the general situation, let ϕ be the hyperparameter. While the use of conjugate priors will often given nice forms for $p(y|\phi)$, they don't combine well with the prior. For example, in the rat tumor example

$$p(y|\alpha, \beta) = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+y)\Gamma(\beta+n-y)}{\Gamma(\alpha+\beta+n)}$$

(i.e. the Beta-Binomial distribution). The posterior density can be calculated (as we did last class), but there isn't a nice conjugate density to this distribution which allows for easy calculation in the future steps.

This sort of situation is commonly the case. The reason why things work nicely for the normal-normal model is that is the conjugate to itself.

Now lets use the fact $p(\mu, \tau) \propto 1$

Similarly to before

$$\mu|\tau, y \sim N(\hat{\mu}, V_\mu)$$

where

$$\hat{\mu} = \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2} \bar{y}_{.j}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2}} \qquad \text{and} \qquad V_\mu^{-1} = \sum_{j=1}^{J} \frac{1}{\sigma_j^2 + \tau^2}$$

The marginal posterior of $\tau|y$ isn't quite as nice, though a useful form for the density can be found, based on the idea

$$\begin{aligned} p(\tau|y) &= \frac{p(\mu,\tau|y)}{p(\mu|\tau,y)} \\ &\propto \frac{p(\tau)\prod_{j=1}^{J} N(\bar{y}_{.j}|\mu,\sigma^2+\tau^2)}{N(\mu|\hat{\mu},V_\mu)} \end{aligned}$$

As noted before, this must hold for any choice of μ, so pick one to make this easy to work with. In this case evaluate at $\mu = \hat{\mu}$ giving,

$$
\begin{aligned}
p(\tau|y) &\propto \frac{p(\tau)\prod_{j=1}^{J} N(\bar{y}_{.j}|\hat{\mu}, \sigma^2+\tau^2)}{N(\hat{\mu}|\hat{\mu}, V_\mu)} \\
&\propto p(\tau) V_\mu^{1/2} \prod_{j=1}^{J} \frac{1}{\sqrt{\sigma^2+\tau^2}} \exp\left(-\frac{(\bar{y}_{.j}-\hat{\mu})^2}{2(\sigma^2+\tau^2)}\right) \\
&= V_\mu^{1/2} \prod_{j=1}^{J} \frac{1}{\sqrt{\sigma^2+\tau^2}} \exp\left(-\frac{(\bar{y}_{.j}-\hat{\mu})^2}{2(\sigma^2+\tau^2)}\right)
\end{aligned}
$$

Comment on prior $p(\tau)$: As mentioned earlier, the Jeffreys' prior ($p(\tau) \propto \tau$) leads to an improper posterior. To show this, you can integrate the density and show that it is infinite. Effectively what is happening is that their are few degrees of freedom for estimating τ. The Jeffreys' prior puts too much weight on larger τs, which leads to the integral to blowup.

Computation:

As $p(\tau|y)$ doesn't correspond to a standard distribution, analyzing the joint posterior is usually done by the following simulation scheme

1. Sample τ_k from $p(\tau|y)$

2. Sample μ_k from $p(\mu_k|\tau_k, y) = N(\mu_k|\hat{\mu}_k, V_{\mu_k})$ where

$$\hat{\mu}_k = \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau_k^2} \bar{y}_{.j}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau_k^2}} \qquad \text{and} \qquad V_{\mu_k}^{-1} = \sum_{j=1}^{J} \frac{1}{\sigma_j^2 + \tau_k^2}$$

3. Sample θ_k from $p(\theta_k|\mu_k, \tau_k, y)$. In this case, the individual components are conditionally independent given μ_k, τ_k, and y giving

$$\theta_{j,k} \sim N(\hat{\theta}_{j,k}, V_{j,k})$$

where

$$\hat{\theta}_{j,k} = \frac{\frac{1}{\sigma_j^2}\bar{y}_{.j} + \frac{1}{\tau_k^2}\mu_k}{\frac{1}{\sigma_j^2} + \frac{1}{\tau_k^2}} \qquad V_{j,k} = \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{\tau_k^2}}$$

Note the conditional independence of the θ_js holds in many hierarchical model. For example, it also held the rat tumor example. It also holds for many of the homework problems (e.g. Chapter 5, # 11,12). This situation will be found to be useful when we get to Gibbs sampling for doing the calculations.

Posterior predictive distributions:

There are two situations where the posterior predictive distribution may need to be calculated. These can be fit into the simulations already done

1. $\tilde{y}$ from a group j already observed.

   Sample $\tilde{y}_{j,k}$ from $N(\theta_{j,k}, \sigma^2)$

   If m observation are needed, draw m values of $\tilde{y}$ from the above distribution.

2. $\tilde{y}$ from a new group $\tilde{j}$

   Sample $\theta_{\tilde{j},k}$ from $N(\theta|\mu_k, \tau_k)$ (draw from prior for θ, not the posterior)

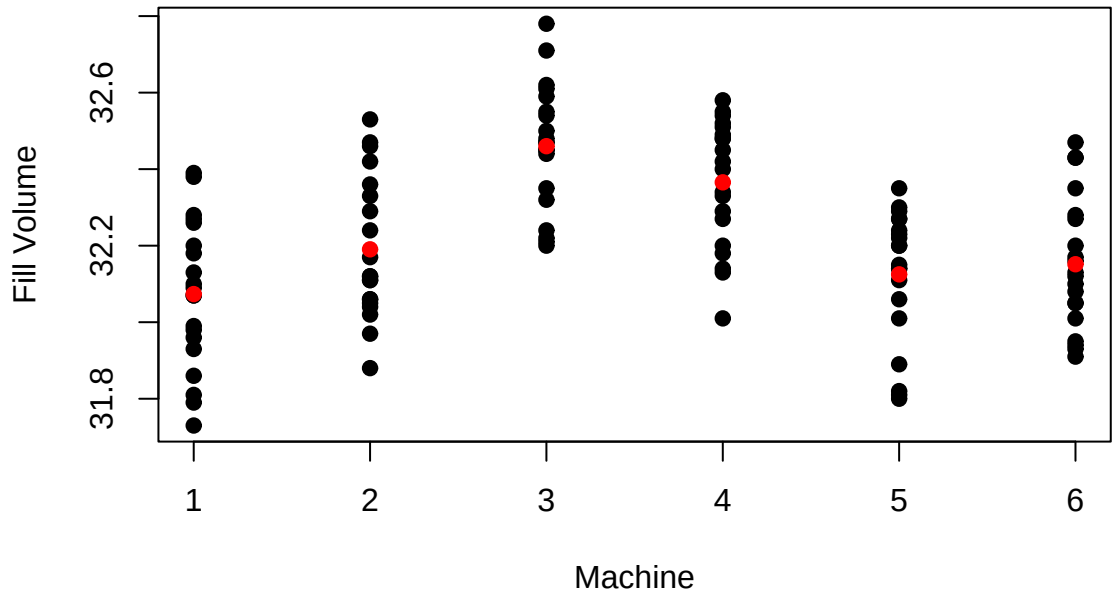
   Sample $\tilde{y}_{\tilde{j},k}$ from $N(\theta_{\tilde{j},k}, \sigma^2)$. Similarly to above if m samples are needed.

The key difference is do we need to draw a new θ or use one we already have. The second situation will lead to more variable samples as there is less information about the corresponding θ in this case.

Examples

Example 1: Detergent Filling Machines

Six filling machines of the same make and model were examined to see whether they put the same amount of detergent into a box. 20 observations from each machine were taken. The nominal amount that should be in a box is 32 ounces.



Note that for this example, σ_j^2 is unknown, but can be estimated based the MSE from the 1-way ANOVA. We will proceed with this value ($\sigma_j^2 = \sigma^2 = 0.00244$) is assumed known.

Calculating $p(\tau|y)$:

$$\begin{aligned}
\hat{\mu} &= \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2}\bar{y}_{\cdot j}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2}} = \bar{y}_{\cdot\cdot} = 32.228 \\
V_\mu^{-1} &= \sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2} \\
V_\mu &= \frac{\sigma^2+\tau^2}{6} = \frac{0.00244+\tau^2}{6}
\end{aligned}$$

By plugging these values into

$$\begin{aligned} p(\tau|y) &\propto \frac{p(\tau)\prod_{j=1}^{J} N(\bar{y}_{.j}|\hat{\mu}, \sigma^2+\tau^2)}{N(\hat{\mu}|\hat{\mu}, V_\mu)} \\ &\propto p(\tau) V_\mu^{1/2} \prod_{j=1}^{J} \frac{1}{\sqrt{\sigma^2+\tau^2}} \exp\left(-\frac{(\bar{y}_{.j}-\hat{\mu})^2}{2(\sigma^2+\tau^2)}\right) \\ &= V_\mu^{1/2} \prod_{j=1}^{J} \frac{1}{\sqrt{\sigma^2+\tau^2}} \exp\left(-\frac{(\bar{y}_{.j}-\hat{\mu})^2}{2(\sigma^2+\tau^2)}\right) \end{aligned}$$

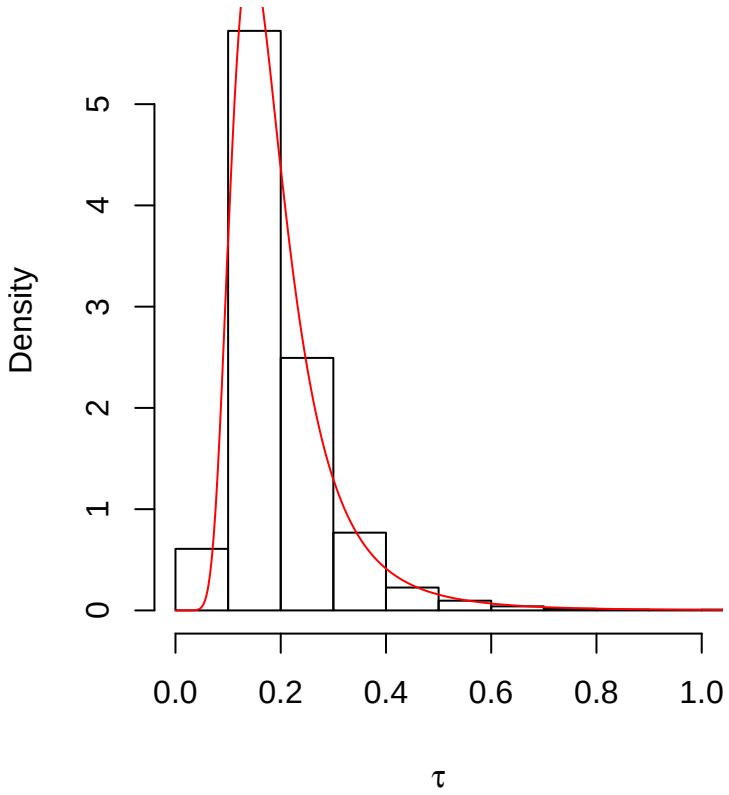
gives $p(\tau|y)$.

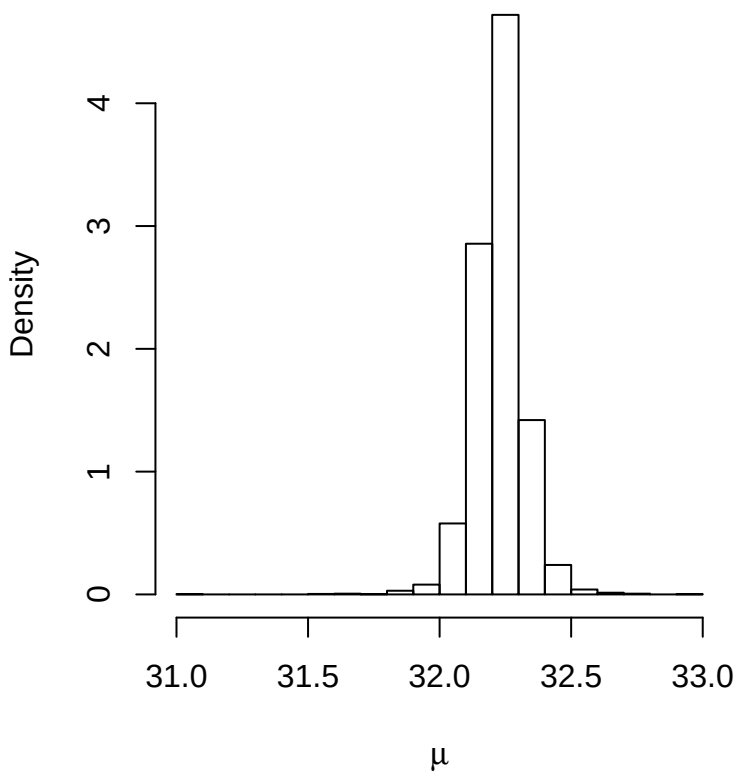
Lets simulate $\tau_1, \ldots, \tau_{5000}$ based on this unnormalized density.

Then $p(\mu_k|\tau_k, y) = N(\mu_k|\hat{\mu_k}, V_{\mu_k})$ is calculated by

$$\begin{aligned}
\hat{\mu_k} &= \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau_k^2}\bar{y}_{.j}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau_k^2}} = \bar{y}_{..} = 32.228 \\
V_\mu &= \frac{\sigma^2 + \tau_k^2}{6} = \frac{0.00244 + \tau_k^2}{6}
\end{aligned}$$

Now sample $\mu_1, \ldots, \mu_{5000}$ based on this conditional distributions.





$$E[\tau|y] = 0.2020 \qquad E[\tau^2|y] = 0.0518$$

$$\text{Mode}(\tau|y) = 0.143 \qquad \text{Mode}(\tau^2|y) = 0.0204$$

$$E[\mu|y] = 32.228 \qquad P[\mu > 32|y] = 0.9878$$

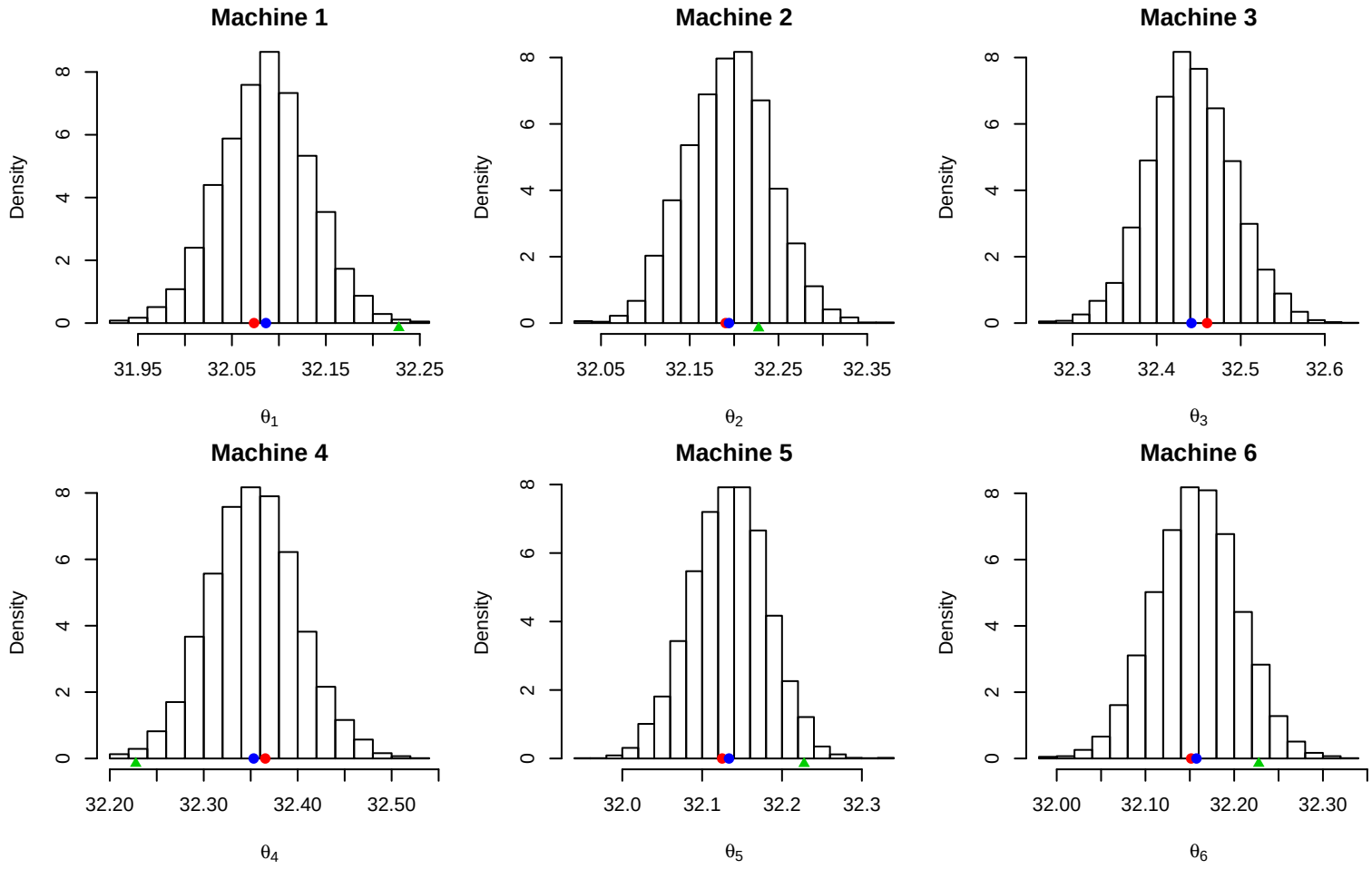
Now lets sample $\theta_{j,k}$ from

$$\theta_{j,k} \sim N(\hat{\theta}_{j,k}, V_{j,k})$$

where

$$\hat{\theta}_{j,k} = \frac{\frac{1}{\sigma_j^2}\bar{y}_{.j} + \frac{1}{\tau_k^2}\mu_k}{\frac{1}{\sigma_j^2} + \frac{1}{\tau_k^2}} \qquad V_{j,k} = \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{\tau_k^2}}$$

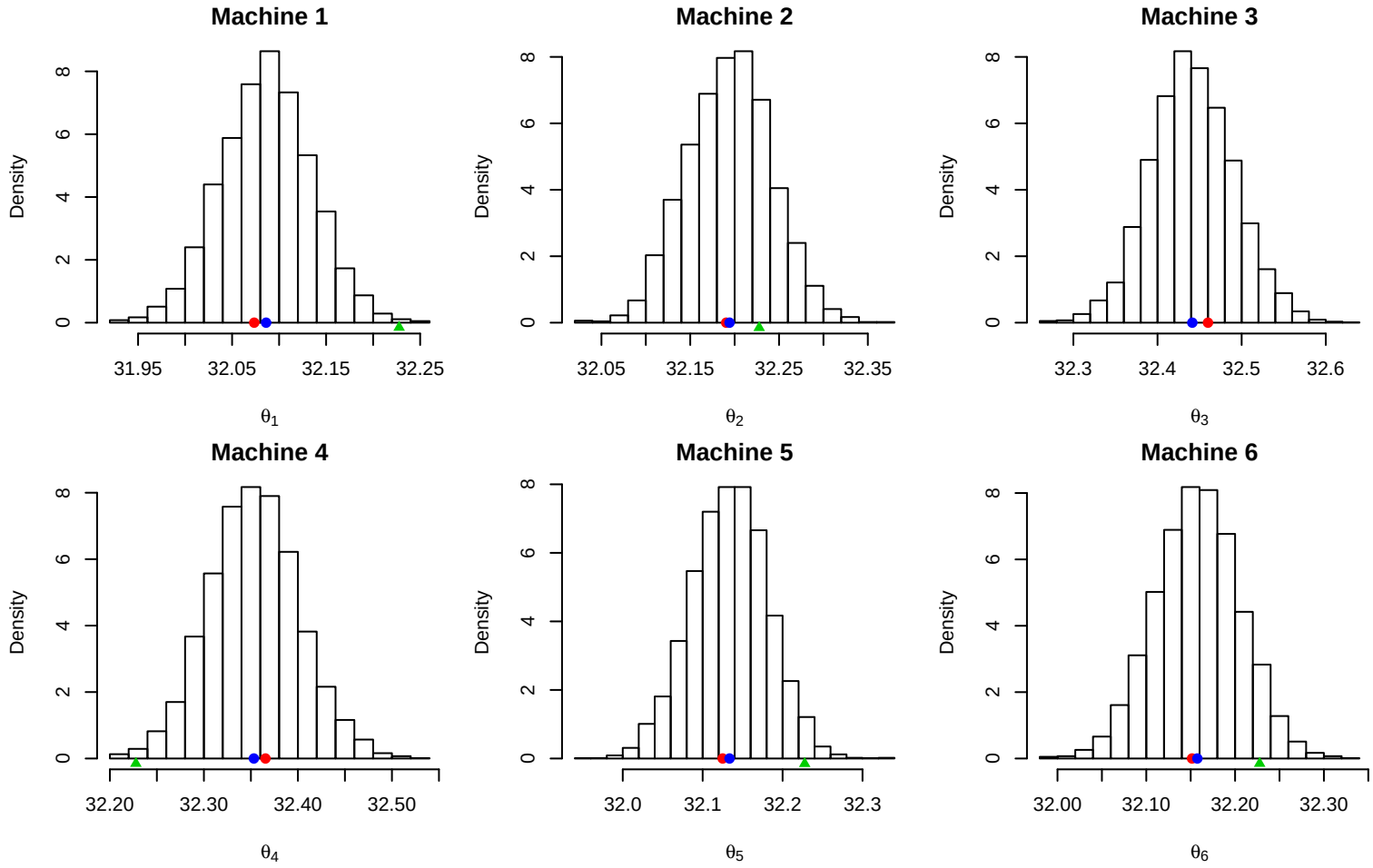
The histograms of these samples are



This plot suggests that we get some shrinkage in the estimate of the machine mean fills (posterior means are blue dots) from the sample averages (red dots). Note that the amount of shrinkage varies from machine to machine.

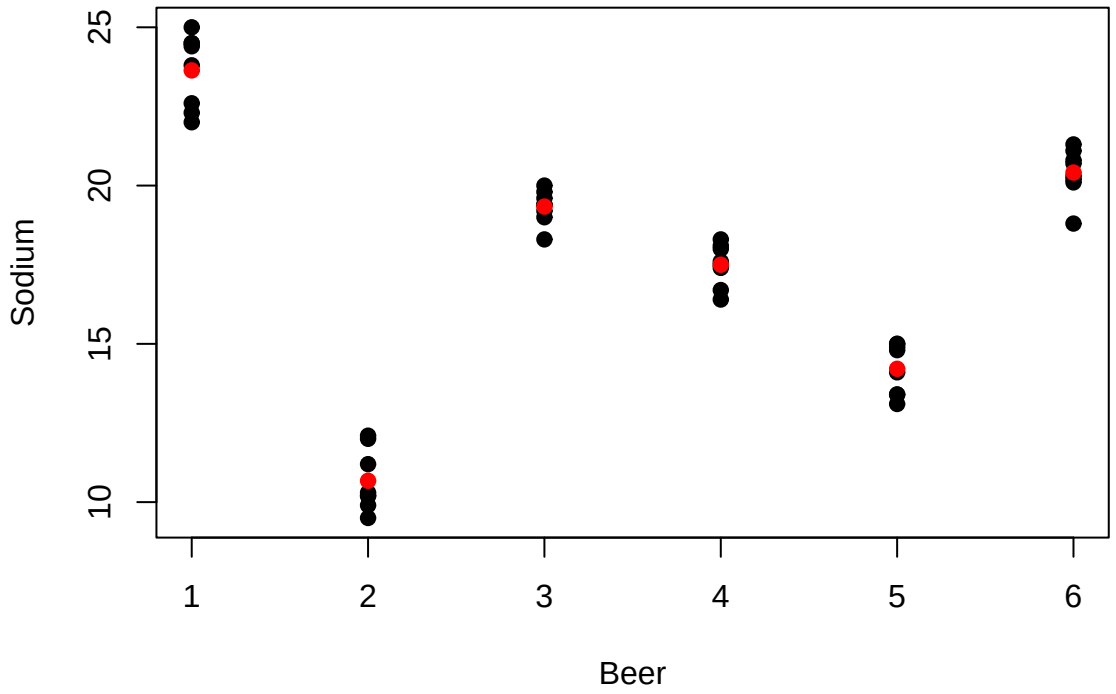
Also of interest is which machines have different fill levels. We can answer this by looking at $P[\theta_i < \theta_j | y]$ for different pairs of machines.

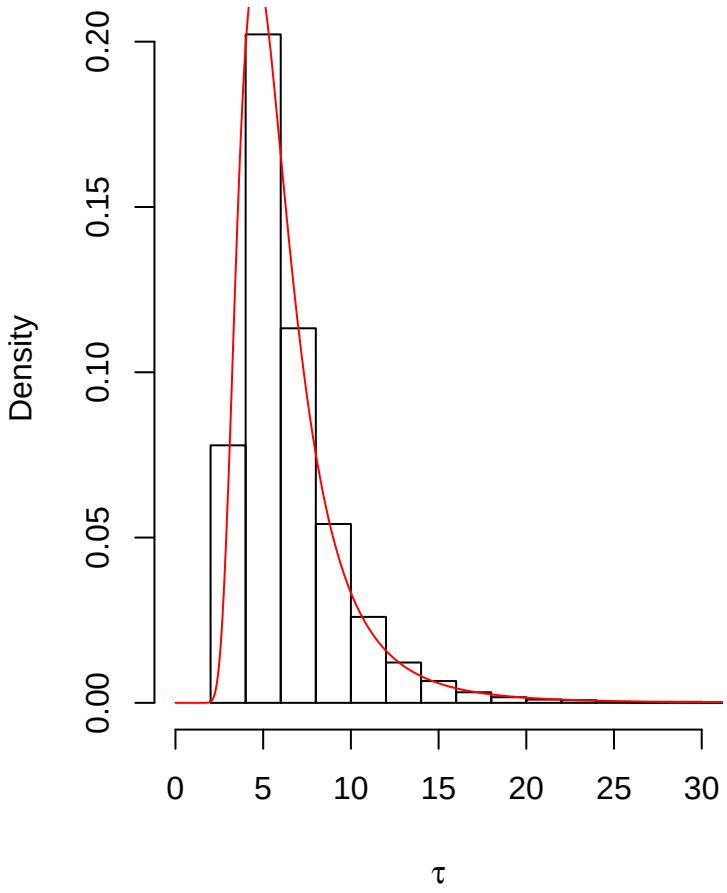
For example $[\theta_1 < \theta_3 | y] = 1$, whereas $[\theta_1 < \theta_5 | y] = 0.7508$

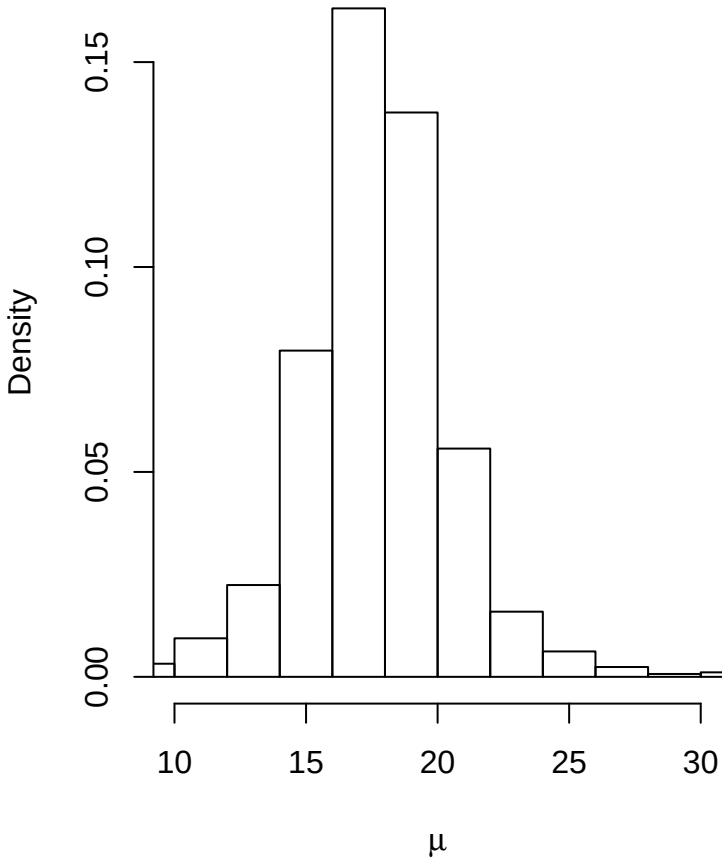


Sodium Content in Beer

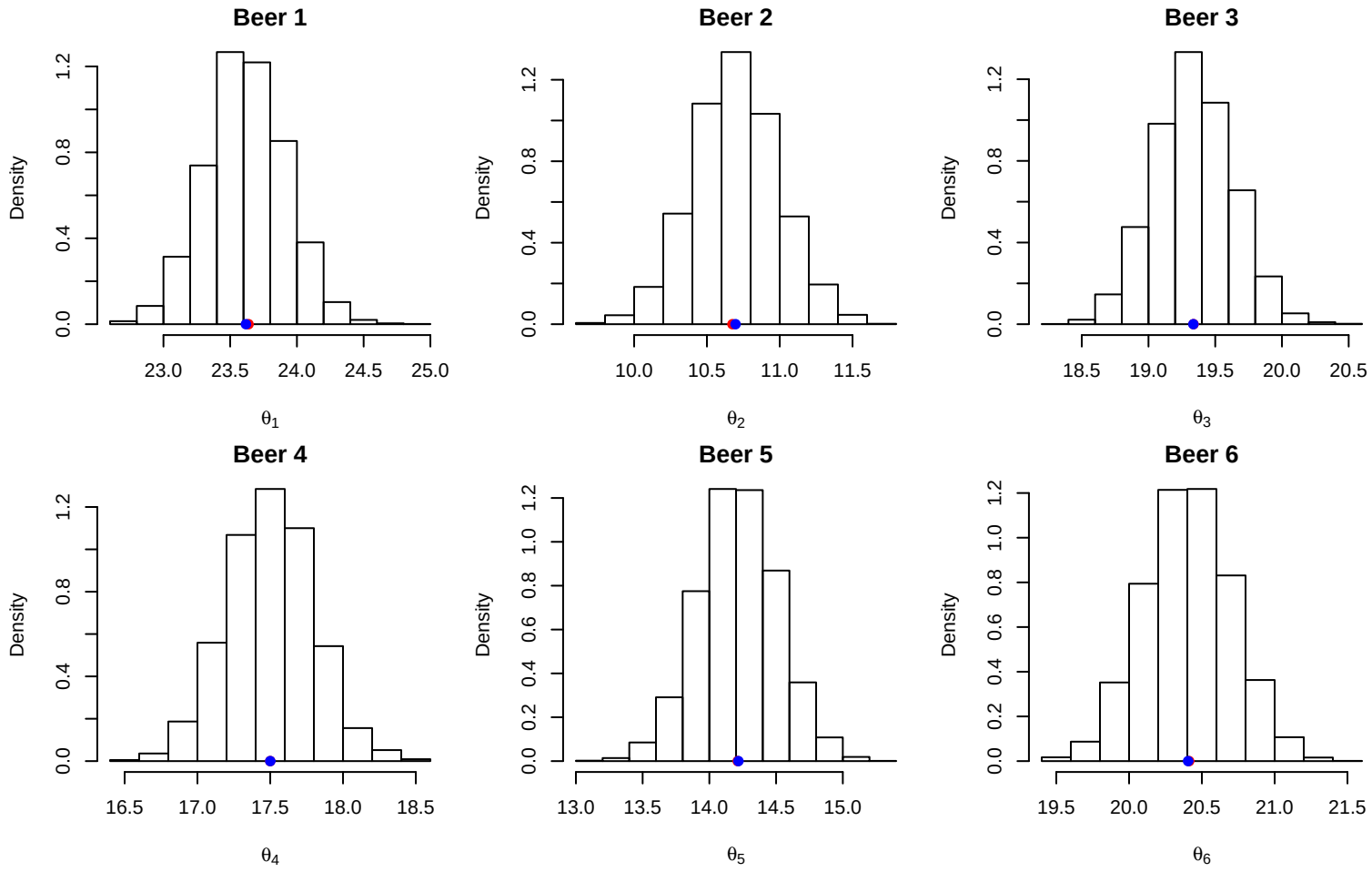
A study was done to investigate the sodium content of 6 randomly chosen brands of U.S. and Canadian beer. For each brand, 8 randomly chosen bottles or cans were analyzed to measure the sodium content (in mg) of each bottle or can. For this analysis, $\sigma_j^2 = 0.0895$, which again is based on the MSE from the 1-way ANOVA.



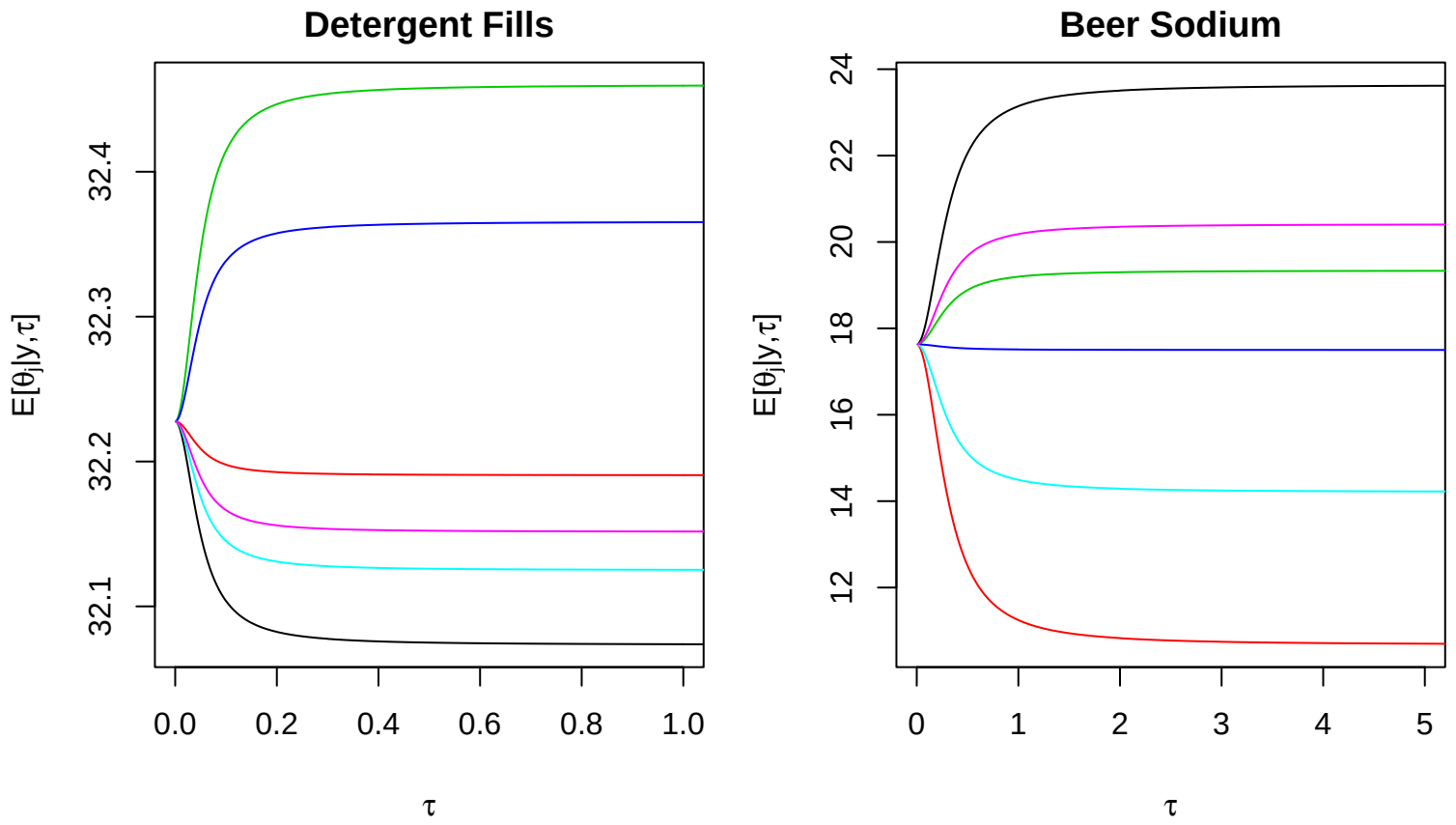




$$E[\tau|y] = 6.448 \qquad E[\tau^2|y] = 50.847$$
$$\text{Mode}(\tau|y) = 4.61 \qquad \text{Mode}(\tau^2|y) = 21.25$$
$$E[\mu|y] = 17.67 \qquad SD(\mu|y) = 2.928$$



There is much less shrinkage in this example. This is not surprising since τ appears to be much bigger relative to σ_j^2 in this example.



The relationship between the amount of shrinkage and σ_j^2 and τ^2 can be seen by

$$E[\theta_i|\mu,\tau,y] = \frac{\tau^2}{\sigma_j^2+\tau^2}\bar{y}_{.j} + \frac{\sigma_j^2}{\sigma_j^2+\tau^2}\mu$$