Multivariate normal data with both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ unknown

When both sets of parameters are unknown, a conjugate family of priors is one in which

$$\boldsymbol{\Sigma} \sim \text{Inverse-Wishart}_{\nu}(\boldsymbol{\Lambda}^{-1})$$

and

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N(\boldsymbol{\eta}, \boldsymbol{\Sigma}/\kappa).$$

The inverse-Wishart distribution is defined on p. 575 of GCSR. The Wishart distribution is a multivariate analog of the gamma distribution.

If matrix $\boldsymbol{U}$ has the Wishart distribution, then $\boldsymbol{U}^{-1}$ has the inverse-Wishart distribution.

The quantity ν is a positive scalar, while $\boldsymbol{\Lambda}$ is a positive definite matrix. They play roles analogous to those played by α and β, respectively, in the Gamma(α, β) distribution.

The other parameters of the prior are the mean vector $\boldsymbol{\eta}$ and κ, the latter of which represents the a priori number of observations.

The prior has the form

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-[(\nu+d)/2+1]}$$

$$\times \exp\left(-\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}) - \tfrac{\kappa}{2}(\boldsymbol{\mu} - \boldsymbol{\eta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\eta})\right).$$

The posterior has this same form but with different parameters: ν, $\boldsymbol{\Lambda}$, $\boldsymbol{\eta}$ and κ become ν_n, $\boldsymbol{\Lambda}_n$, $\boldsymbol{\mu}_n$ and κ_n, respectively, where

$$\nu_n = \nu + n, \quad \kappa_n = \kappa + n,$$

$$\boldsymbol{\mu}_n = \left(\frac{\kappa}{\kappa + n}\right)\boldsymbol{\eta} + \left(\frac{n}{\kappa + n}\right)\bar{\boldsymbol{y}}$$

and

$$\boldsymbol{\Lambda}_n = \boldsymbol{\Lambda} + n\boldsymbol{S}^2 + \left(\frac{\kappa n}{\kappa + n}\right)(\bar{\boldsymbol{y}} - \boldsymbol{\eta})(\bar{\boldsymbol{y}} - \boldsymbol{\eta})^T,$$

where

$$\boldsymbol{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T.$$

Finally, we have the following properties of component distributions of the posterior:

- The conditional distribution of $\boldsymbol{\mu}$ given $\boldsymbol{\Sigma}$ and the data is $N(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}/\kappa_n)$.

- The marginal posterior of $\boldsymbol{\mu}$ is multivariate t.

- The marginal posterior of $\boldsymbol{\Sigma}$ is

$$\text{Inverse-Wishart}_{\nu_n}(\boldsymbol{\Lambda}_n^{-1}).$$

The Jeffreys noninformative prior when $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown is

$$\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(d+1)/2},$$

which is the limit of the conjugate prior as $\kappa \to 0$, $\nu \to -1$ and $|\boldsymbol{\Lambda}| \to 0$.

The resulting posterior is proper and given by

$$\boldsymbol{\Sigma}|\boldsymbol{y} \sim \text{Inverse-Wishart}_{n-1}\left((n\boldsymbol{S}^2)^{-1}\right)$$

and

$$\boldsymbol{\mu}|\boldsymbol{\Sigma}, \boldsymbol{y} \sim N(\bar{\boldsymbol{y}}, \boldsymbol{\Sigma}/n).$$

Example 14 *Structured $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$*

This example involves a case where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, but highly structured. By structured, we mean that there are fewer than d mean parameters and fewer than $d + d(d-1)/2$ covariance parameters that are unknown.

The data are $\boldsymbol{Y}^T = (Y_1, \ldots, Y_n)$, and our model is as follows:

$$(Y_i - \mu) = \rho(Y_{i-1} - \mu) + \epsilon_i, \quad i = 2, \ldots, n$$

The quantities $\epsilon_2, \ldots, \epsilon_n$ are unobserved random variables that are i.i.d. $N(0, \sigma^2)$ and independent of Y_1. We assume also that $Y_1 \sim N(\mu, \sigma^2/(1-\rho^2))$.

The vector of unknown parameters is $\boldsymbol{\theta} = (\rho, \mu, \sigma)$, and the parameter space is

$$\{(\rho, \mu, \sigma) : -1 < \rho < 1, -\infty < \mu < \infty, \sigma > 0\}.$$

This model is known as an *autoregressive process*, since the Y_i's are regressed on themselves. It is a common model in time series analysis.

In this context, the i in Y_i represents time, and $Y_1, \ldots, Y_n$ are observations made in chronological order, typically at evenly spaced time points.

It can be verified that $Y_1, \ldots, Y_n$ have a multivariate normal distribution such that $E(Y_i) = \mu$, $i = 1, \ldots, n$, and

$$\text{Cov}(Y_i, Y_j) = \left(\frac{\sigma^2}{1-\rho^2}\right)\rho^{|i-j|}, \quad i = 1, \ldots, n, \\ j = 1, \ldots, n.$$

Note that, for $k = 1, 2, \ldots$, ρ^k is the correlation between Y_i and Y_{i+k}, which does not depend on i.

So, in this autoregressive process the correlation between observations depends only upon how far apart in time they are made.

One can verify that the likelihood function is

$$
\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{\theta}) &\propto \sigma^{-n}\sqrt{1-\rho^2}\,\exp\left(-\frac{(y_1-\mu)^2}{2\sigma_\rho^2}\right) \\
&\quad \times \exp\left(-\frac{1}{2\sigma^2}\sum_{i=2}^{n}\epsilon_i^2\right),
\end{aligned}
$$

where

$$\sigma_\rho^2 = \frac{\sigma^2}{1-\rho^2}.$$

To a good approximation, the likelihood is

$$\widehat{f}(\boldsymbol{y}|\boldsymbol{\theta}) \propto \sigma^{-(n-1)}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=2}^{n}\epsilon_i^2\right).$$

Sufficient statistics in $\widehat{f}$ are

$$\bar{y}_1 = \frac{1}{n-1} \sum_{i=2}^{n} y_i, \quad \bar{y}_2 = \frac{1}{n-1} \sum_{i=2}^{n} y_{i-1},$$

$$s_1^2 = \frac{1}{n-1} \sum_{i=2}^{n} (y_i - \bar{y}_1)^2,$$

$$s_2^2 = \frac{1}{n-1} \sum_{i=2}^{n} (y_i - \bar{y}_2)^2$$

and

$$\widehat{\rho} = \frac{1}{n-1} \sum_{i=2}^{n} (y_i - \bar{y}_1)(y_{i-1} - \bar{y}_2)/s_2^2.$$

Use of the likelihood $\widehat{f}$ simplifies the computation of Jeffreys prior. One can verify that the information matrix is

$$I(\boldsymbol{\theta}) = (n-1)\begin{bmatrix} \frac{1}{(1-\rho^2)} & 0 & 0 \\ 0 & \frac{(1-\rho)^2}{\sigma^2} & 0 \\ 0 & 0 & \frac{2}{\sigma^2} \end{bmatrix},$$

and hence the Jeffreys prior is

$$\pi(\rho, \mu, \sigma) \propto \frac{1}{\sigma^2}\left(\frac{1-\rho}{1+\rho}\right)^{1/2} I_{(-1,1)}(\rho) I_{(0,\infty)}(\sigma).$$

According to this prior, the three parameters are a priori independent. The prior is improper as it is constant in μ. The marginal prior for σ is also improper, but the marginal prior for ρ is proper.

The posterior is

$$\pi(\rho, \mu, \sigma|\boldsymbol{y}) \;\propto\; \sigma^{-(n+1)} \left(\frac{1-\rho}{1+\rho}\right)^{1/2} \times \exp\left(-\frac{1}{2\sigma^2}\sum_{i=2}^{n} \epsilon_i^2\right).$$

Some component distributions are as follows:

1. The marginal of ρ is

$$\pi_1(\rho|\boldsymbol{y}) \propto (1-\rho^2)^{-1/2}\left[(\rho-\widehat{\rho})^2 + \left(\frac{s_1}{s_2}\right)^2 - \widehat{\rho}^2\right]^{-(n-1)/2}.$$

2. The conditional posterior of $1/\sigma^2$ given ρ is gamma with shape parameter $(n-1)/2$ and rate

$$\frac{1}{2}\sum_{i=2}^{n}\left[(y_i - \bar{y}_1) - \rho(y_{i-1} - \bar{y}_2)\right]^2.$$

3. The conditional posterior of μ given ρ and σ is

$$N\left(\frac{\bar{y}_1 - \rho\bar{y}_2}{(1-\rho)}, \frac{\sigma^2}{(n-1)(1-\rho)^2}\right).$$

One hundred observations were generated from the model on p. 140N with $\rho = .6$, $\mu = 0$ and $\sigma = 1$. The resulting sufficient statistics were

$$\bar{y}_1 = -0.08241449, \quad \bar{y}_2 = -0.08725065,$$

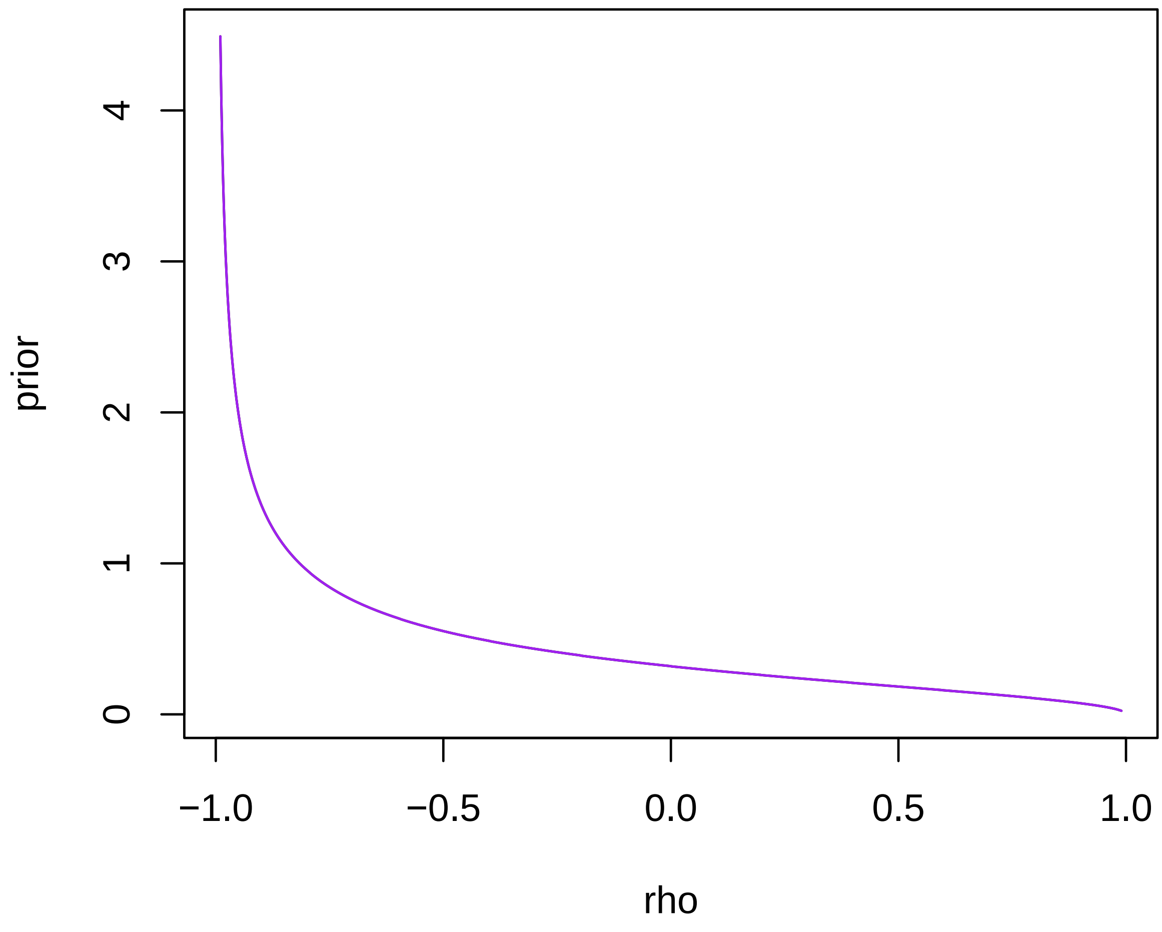
$$s_1^2 = 1.728523, \quad s_2^2 = 1.734809$$

and

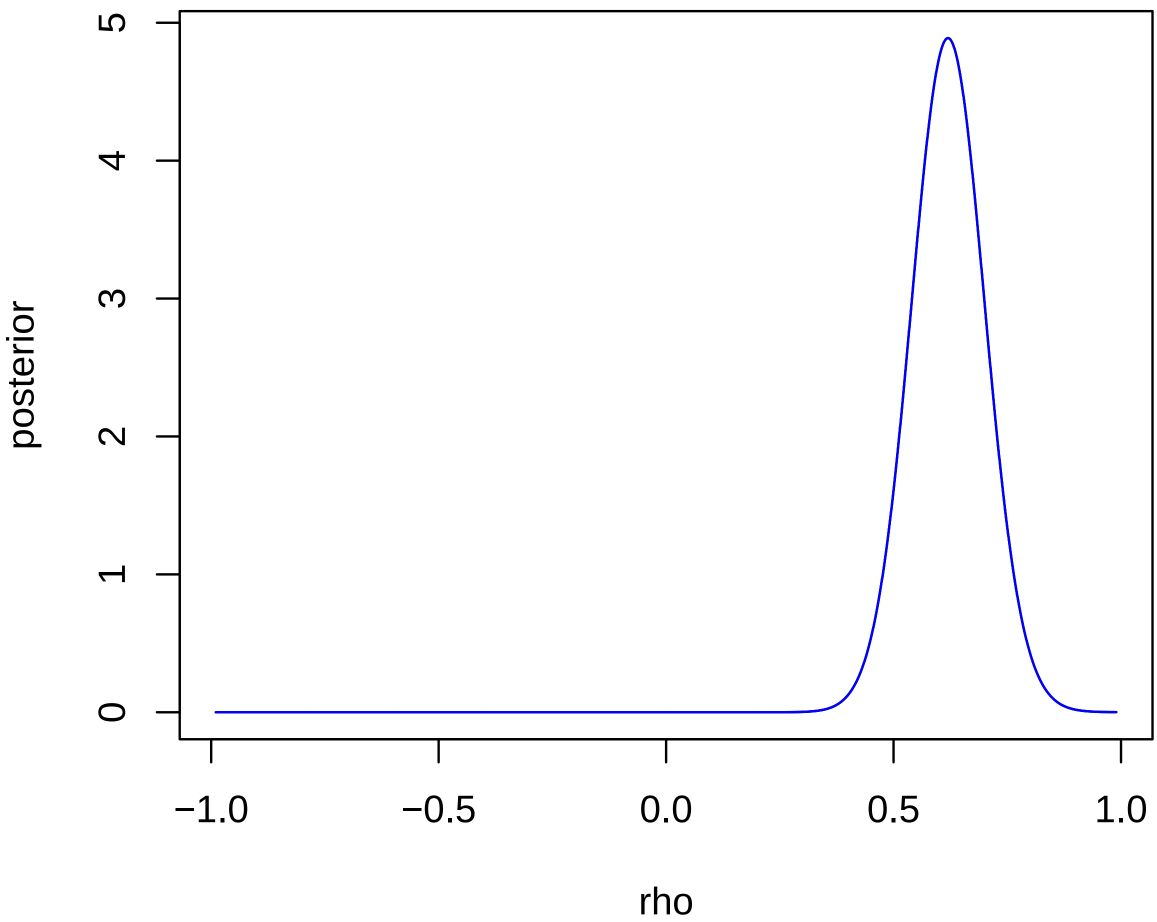
$$\hat{\rho} = 0.6134537$$

Noninformative prior for ρ

$$\pi(\rho) = \frac{1}{\pi}\left(\frac{1-\rho}{1+\rho}\right)^{1/2} I_{(-1,1)}(\rho)$$

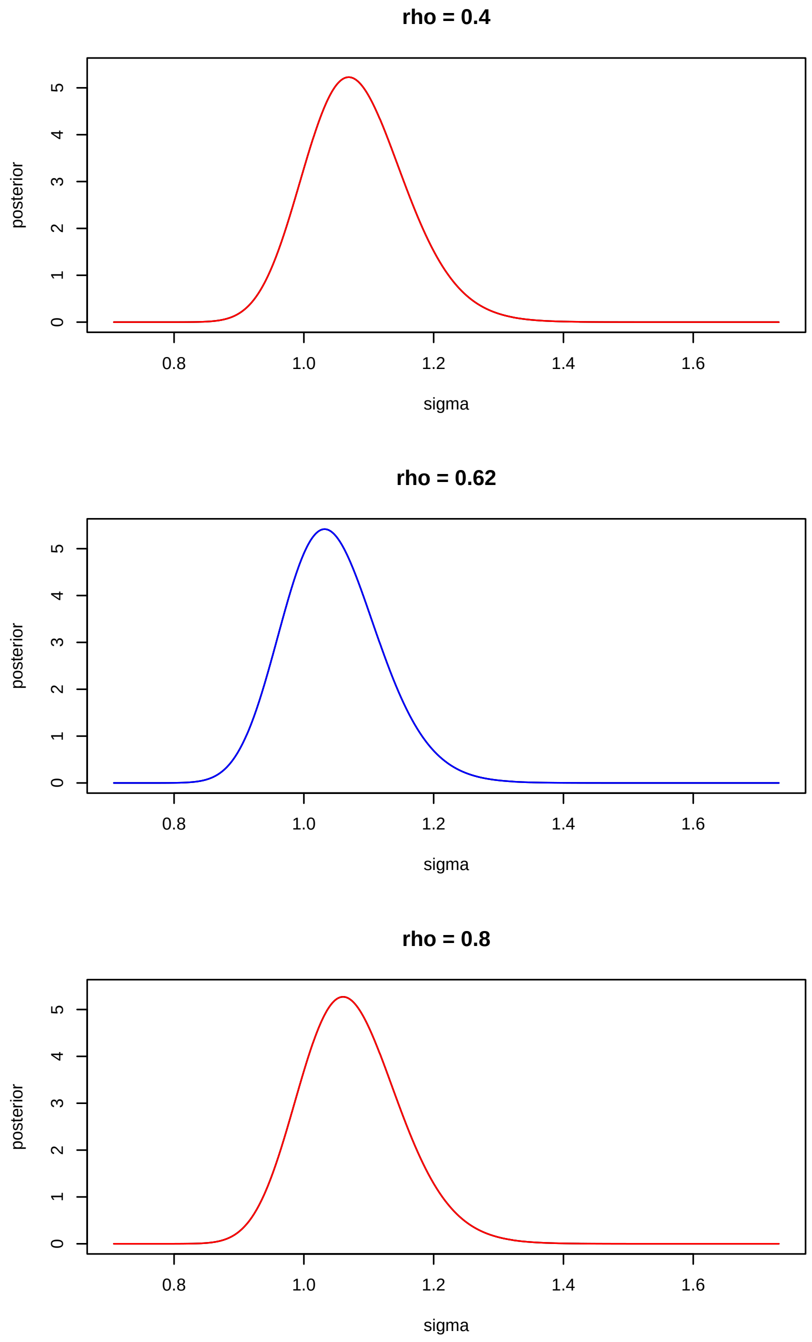


Marginal posterior of ρ
for generated data



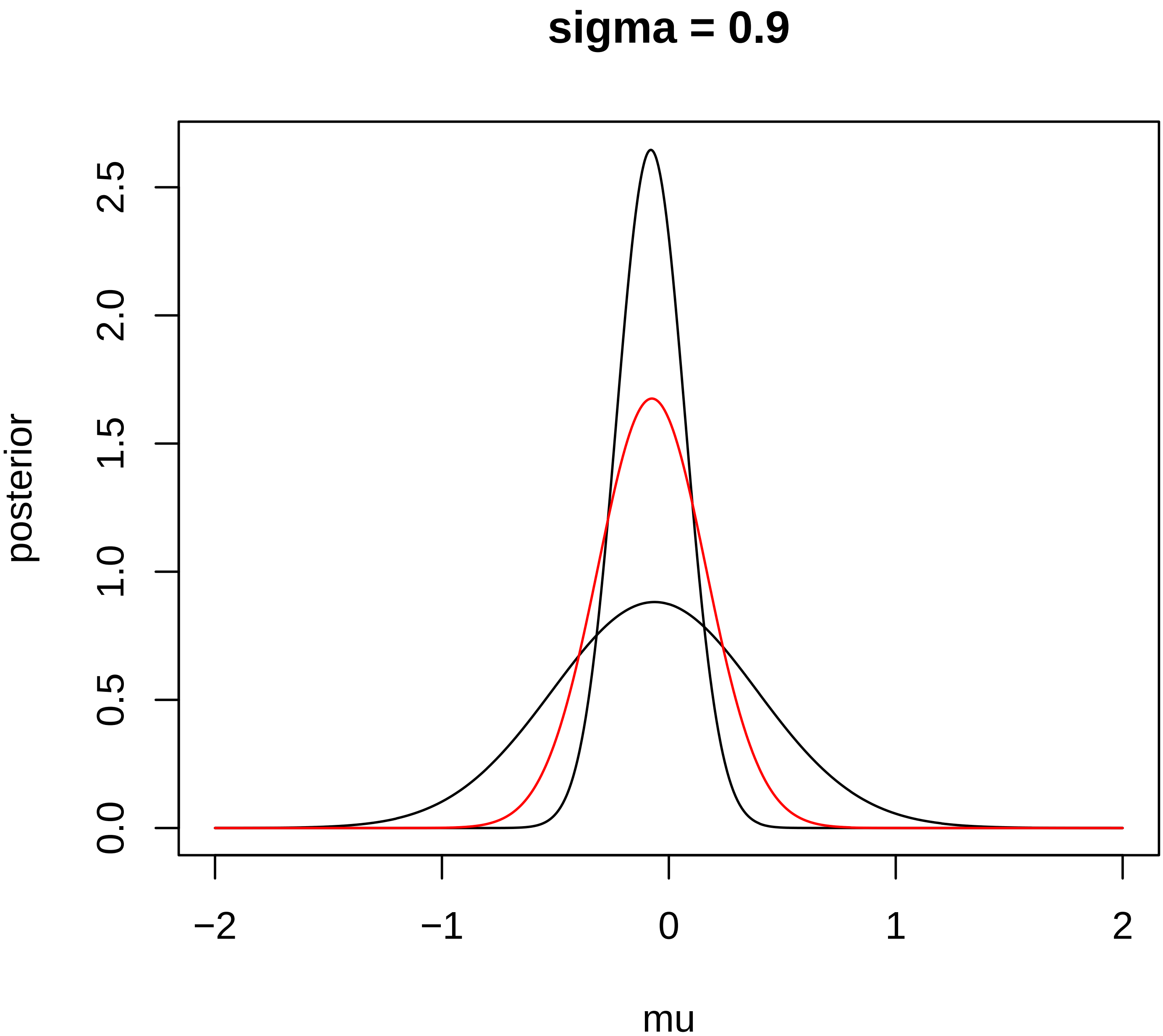
The mode of this posterior is at 0.620.

Posterior of σ conditional on ρ



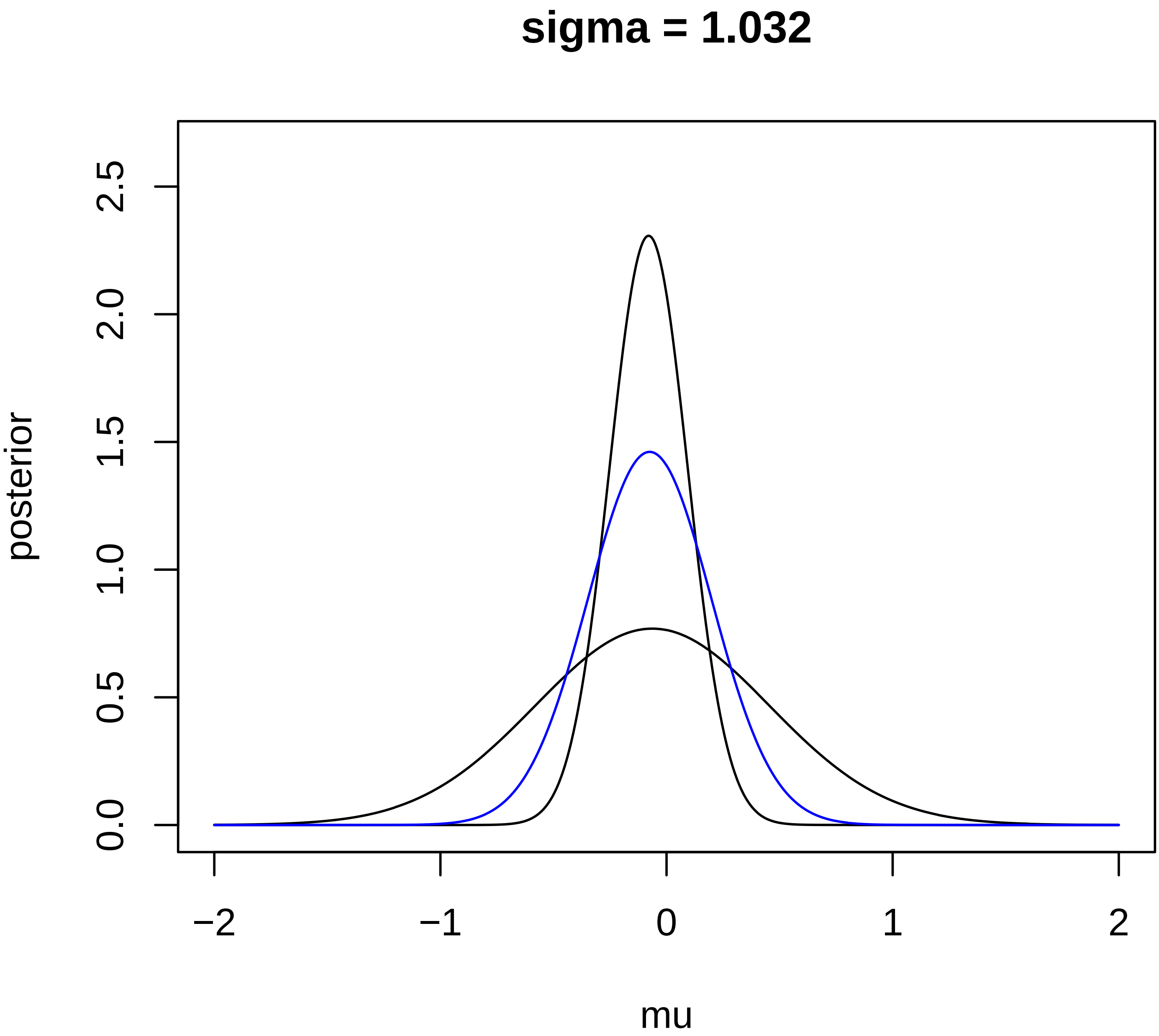
At $\rho = 0.62$, the mode of the conditional is 1.032.

Posterior of μ conditional on ρ and σ



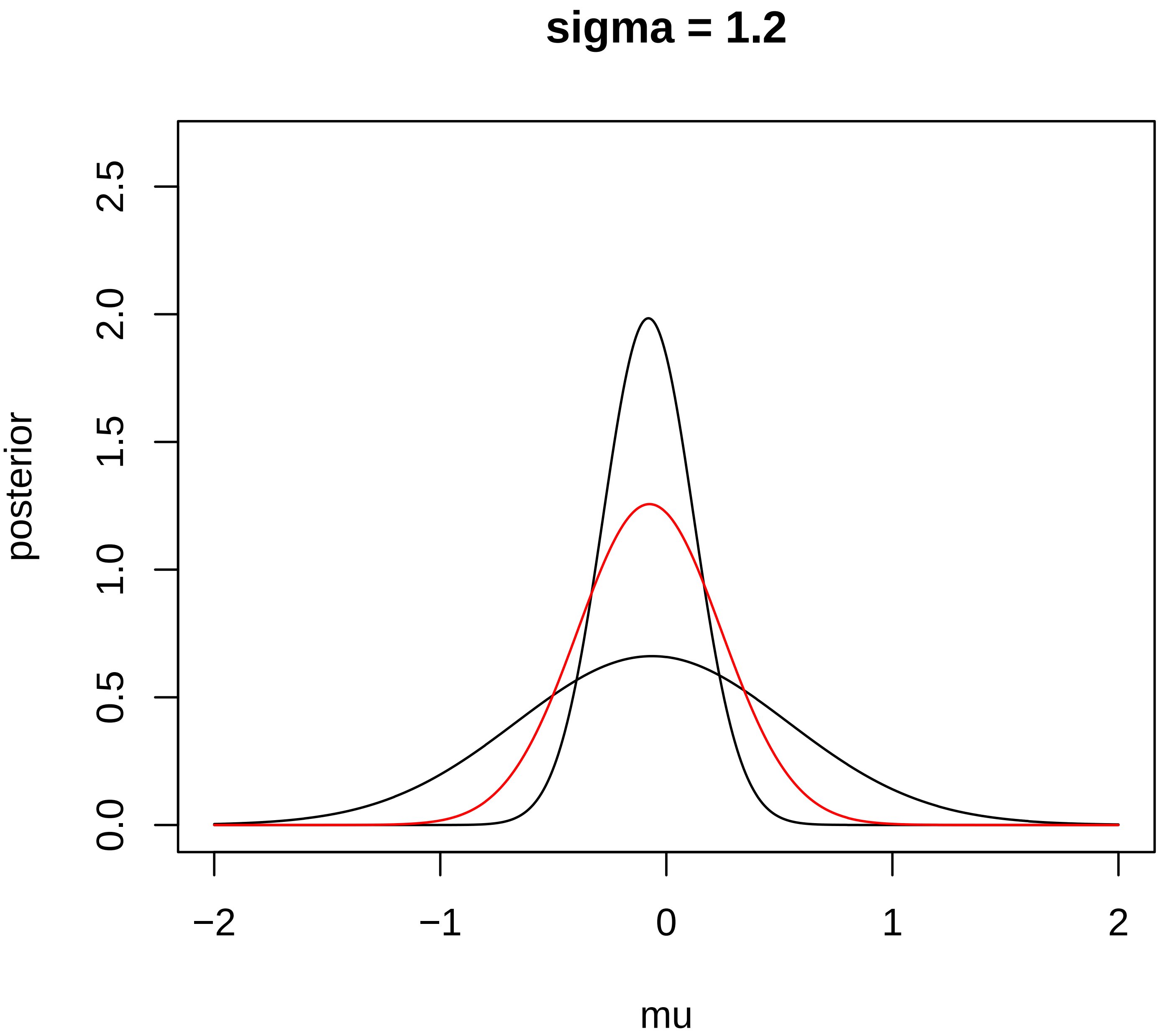
The red curve corresponds to $\rho = 0.62$ and the other two to $\rho = 0.4$ and 0.8.

Posterior of μ conditional on ρ and σ



The blue curve corresponds to $\rho = 0.62$ and the other two to $\rho = 0.4$ and 0.8.

Posterior of μ conditional on ρ and σ



The red curve corresponds to $\rho = 0.62$ and the other two to $\rho = 0.4$ and 0.8.