Bayesian Statistics

Multivariate Normal Model

Nan Lin
Department of Mathematics
Washington University in St. Louis

Multivariate normal distribution

- A d-dimensional random vector $\boldsymbol{Y}^T = (Y_1, \dots, Y_d)$
  - $Y \sim N_d(\mu, \Sigma)$

- Parameters
  - Mean: $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)^T = E(\boldsymbol{Y}^T)$
  - Covariance matrix Σ:
    - $\Sigma_{ij} = Cov(Y_i, Y_j), i = 1, \dots, d, j = 1, \dots, d$
- p.d.f.

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-d/2} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2} (\boldsymbol{y} - \boldsymbol{\mu})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}))$$

Marginal and conditional distribution of a normal random vector

- Let $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ contain the first k and the last $d-k$ elements of $\boldsymbol{\mu}$, respectively.
- Similarly, define $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$.
- Then partition Σ as

$$\Sigma = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$$

- Then,
  - Marginally, $\boldsymbol{y}_1 \sim N(\boldsymbol{\mu}_1, \Sigma^{11})$ and $\boldsymbol{y}_2 \sim N(\boldsymbol{\mu}_2, \Sigma^{22})$
  - Conditionally,

$$\boldsymbol{y}_1 | \boldsymbol{y}_2 \sim N(\boldsymbol{\mu}_1 + \Sigma^{12}(\Sigma^{22})^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2), \Sigma^{11} - \Sigma^{12}(\Sigma^{22})^{-1}\Sigma^{21})$$

Conjugate prior when Σ is known

- Data:
  - a random sample of size n from a d-variate normal distribution
  - $\boldsymbol{Y}$: $n \times d$ matrix
    - The ith row: $\boldsymbol{Y}_i^T = (Y_{i1}, \ldots, Y_{id}), i = 1, \ldots, n.$
  - Likelihood

    $f(\boldsymbol{y}|\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-\frac{n}{2}} \exp(-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu}))$
- Assuming Σ is known
- Conjugate prior $\boldsymbol{\mu} \sim N(\boldsymbol{\eta}, \Lambda)$
- Posterior $\boldsymbol{\mu}|\boldsymbol{Y} \sim N(\boldsymbol{\mu}_n, \Lambda_n)$
  - $\boldsymbol{\mu}_n = (\Lambda^{-1} + n\Sigma^{-1})^{-1}(\Lambda^{-1}\boldsymbol{\eta} + n\Sigma^{-1}\overline{\boldsymbol{y}})$
  - $\Lambda_n^{-1} = \Lambda^{-1} + n\Sigma^{-1}$

Noninformative prior when Σ is known

- Let $|\Lambda^{-1}| \to 0$, we obtain a noninformative prior in the limit.
  - Uniform over R^d, so this is improper.
- Posterior
  - If $n \geq d$, $\boldsymbol{\mu}|\boldsymbol{Y} \sim N(\overline{\boldsymbol{y}}, \frac{\Sigma}{n})$

James-Stein estimator

- $Y_i \sim N(\mu_i, 1)$ independent, i.e. $\boldsymbol{Y} \sim N_d(\boldsymbol{\mu}, I)$
- James-Stein estimator
  - $\delta_{JS}(\boldsymbol{Y}) = \left(1 - \frac{d-2}{\sum_{i=1}^{d} Y_i^2}\right)\boldsymbol{Y}$
- Squared error loss
  - $L(\boldsymbol{\mu}, \boldsymbol{a}) = \sum_i^d (\mu_i - a_i)^2$
- Under the squared error loss, $\delta_{JS}(\boldsymbol{Y})$ has uniformly a smaller risk than $\boldsymbol{Y}$
  - The risk ratio is very close to 1 over most of the parameter space
  - Only near $\mu^T = (0,0, \dots, 0)$, the ratio is substantially smaller than 1.

When both μ and Σ are unknown

- Conjugate prior
  - $\Sigma^{-1} \sim Wishart\,(\Lambda^{-1}, \nu)$
    - i.e. $\Sigma \sim Inv - Wishart\,(\Lambda^{-1}, \nu)$
  - $\mu | \Sigma \sim N(\eta, \frac{\Sigma}{\kappa})$

Wishart distribution

- Suppose $X_i = (x_{i1}, \dots, x_{id}) \sim N_d(0, V), i = 1, \dots, d,$
  - Independent
  - $X: \nu \times d$ matrix
- Then $S = X^T X \sim Wishart\ (V_{d\times d}, \nu)$
  - Symmetric
  - Positive definite
  - Same requirement for V
- p.d.f.
  - $f(S|V,\nu) \propto \frac{|S|^{\frac{\nu-d-1}{2}}}{|V|^{\frac{\nu}{2}}} \exp[-\frac{1}{2} tr(V^{-1}S)]$
- A multivariate generalization of Gamma (or χ^2) distribution

Some properties of $Wishart\ (V, \nu)$

- $E(S_{ij}) = \nu V_{ij}$
- $Var(S_{ij}) = \nu(V_{ij}^2 + V_{ii}V_{jj})$
- $Cov(S_{ij}, S_{kl}) = \nu(V_{ik}V_{jl} + V_{il}V_{jk})$

- If $S \sim Wishart\ (V_{d\times d}, \nu)$, then $\Omega = S^{-1} \sim Inv - Wishart\ (V, \nu)$
  - p.d.f.

$$f(\Omega) \propto \frac{|\Omega|^{\frac{-(\nu+d+1)}{2}}}{|V|^{\frac{\nu}{2}}} \exp[-\frac{1}{2} tr(V^{-1}\Omega^{-1})]$$

When both μ and Σ are unknown - Conjugate prior

- Conjugate prior
  - $\Sigma^{-1} \sim Wishart\,(\Lambda^{-1}, \nu)$
    - i.e. $\Sigma \sim Inv - Wishart\,(\Lambda^{-1}, \nu)$
  - $\mu|\Sigma \sim N(\eta, \frac{\Sigma}{\kappa})$
- Joint prior distribution
  - $\pi(\mu, \Sigma) =$
    $|\Sigma|^{-(\frac{\nu+d}{2}+1)} \exp(-\frac{1}{2} tr(\Lambda\Sigma^{-1}) - \frac{\kappa}{2}(\mu-\eta)^T\Sigma^{-1}(\mu-\eta))$

When both μ and Σ are unknown - Conjugate prior

- Posterior
  - $\Sigma^{-1}|y \sim Wishart\,(\Lambda_n^{-1}, \nu_n)$
  - $\mu|\Sigma, y \sim N(\mu_n, \frac{\Sigma}{\kappa_n})$
  - where
    - $\nu_n = \nu + n\,, \kappa_n = \kappa + n$
    - $\mu_n = \left(\frac{\kappa}{\kappa+n}\right)\eta + \left(\frac{n}{\kappa+n}\right)\bar{y}$
    - $\Lambda_n = \Lambda + nS^2 + \left(\frac{\kappa n}{\kappa+n}\right)(\bar{y}-\eta)(\bar{y}-\eta)^T$
    - $S^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})^T$
- Marginal posterior of $\mu|y$ is multivariate t-distribution.

When both μ and Σ are unknown – Jeffrey's prior

- In the conjugate prior, let $\kappa \to 0, \nu \to -1$, and $|\Lambda| \to 0$,
  - $\pi(\mu, \Sigma) \propto |\Sigma|^{-(d+1)/2}$
- Posterior
  - $\Sigma^{-1}|y \sim Wishart\left((nS^2)^{-1}, n-1\right)$
  - $\mu|\Sigma, y \sim N(\bar{y}, \frac{\Sigma}{n})$

When both μ and Σ are unknown but structured – AR(1)

- Data: $\boldsymbol{Y}^T = (Y_1, \dots, Y_n)$
- Autoregressive process

$$Y_i - \mu = \rho(Y_{i-1} - \mu) + \epsilon_i, i = 2, \dots, n,$$
$$\epsilon_i \sim N(0, \sigma^2) \text{ i.i.d. and independent of } Y_1$$

- Assume that $Y_1 \sim N(\mu, \frac{\sigma^2}{1-\rho^2})$
- Unknown parameters: $\boldsymbol{\theta} = \{(\rho, \mu, \sigma): -1 < \rho < 1, -\infty < \mu < \infty, \sigma > 0\}$
- Model: $\boldsymbol{Y}|\boldsymbol{\theta}$ is multivariate normal with
  - $E(Y_i) = \mu$ for all i
  - $Cov(Y_i, Y_j) = \left(\frac{\sigma^2}{1-\rho^2}\right)\rho^{|i-j|}, i = 1, \dots, n, j = 1, \dots, n$
  - $Cor(Y_i, Y_{i+k}) = \rho^k$, for all k

When both μ and Σ are unknown but structured – AR(1)

- Likelihood

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = f(y_n|y_{n-1},\boldsymbol{\theta})f(y_{n-1}|y_{n-2},\boldsymbol{\theta})\cdots f(y_2|y_1,\boldsymbol{\theta})f(y_1|\boldsymbol{\theta})$$

$$\propto \sigma^{-n}\sqrt{1-\rho^2}\exp\left(-\frac{(y_1-\mu)^2}{2\sigma_\rho^2}\right)\exp(-\frac{1}{2\sigma^2}\sum_{i=2}^{n}\epsilon_i^2)$$

where $\sigma_\rho^2 = \frac{\sigma^2}{1-\rho^2}, \epsilon_i = y_i - \mu - \rho(y_{i-1}-\mu)$

- Use an approximate likelihood $f(\boldsymbol{y}|y_1,\boldsymbol{\theta})$: conditional on initial value

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

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

When both μ and Σ are unknown but structured – AR(1)

- Jeffery's prior

$$\pi(\rho, \mu, \sigma) \propto \frac{1}{\sigma^2}\left(\frac{1-\rho}{1+\rho}\right)^{\frac{1}{2}}, -1 < \rho < 1, \sigma > 0$$

- Joint posterior

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

When both μ and Σ are unknown but structured – AR(1)

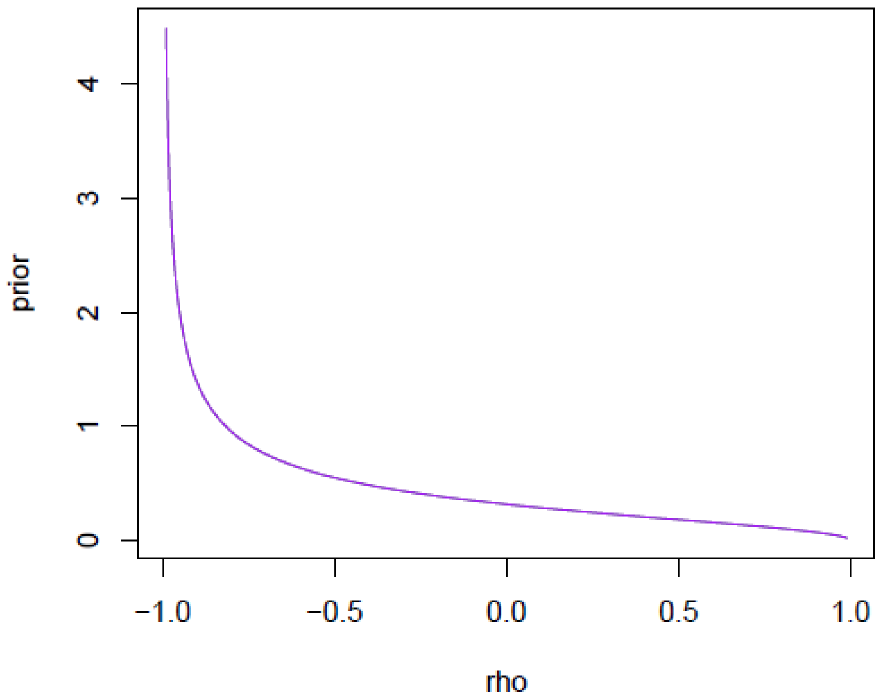
- Marginal posterior
  - $f(\rho|\boldsymbol{y}) \propto \frac{1}{\sqrt{1-\rho^2}}\left[(\rho-\hat{\rho})^2+\left(\frac{s_1}{s_2}\right)^2-\hat{\rho}^2\right]^{-(n-1)/2}$
  - $\sigma^{-2}|\rho,\boldsymbol{y}\sim Gamma(\frac{n-1}{2},\frac{1}{2}\sum_{i=2}^{n}[(y_i-\bar{y}_1)-\rho(y_{i-1}-\bar{y}_2)]^2)$
  - $\mu|\rho,\sigma,\boldsymbol{y}\sim N(\frac{\bar{y}_1-\rho\bar{y}_2}{1-\rho},\frac{\sigma^2}{(n-1)(1-\rho)^2})$

  where

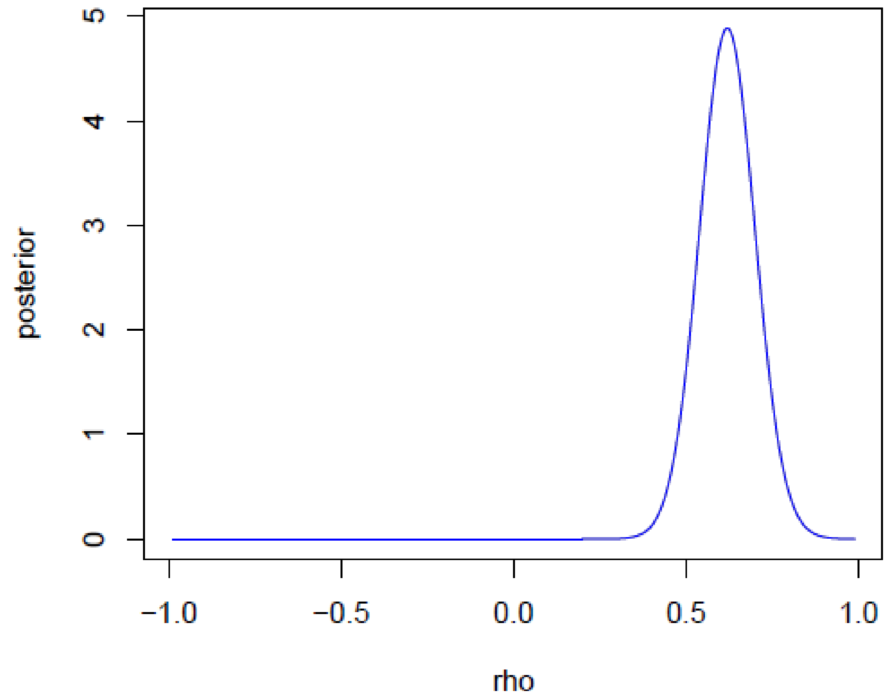
  - $\bar{y}_1=\frac{1}{n-1}\sum_{i=2}^{n}y_i\,,\bar{y}_2=\frac{1}{n-1}\sum_{i=2}^{n}y_{i-1}$
  - $s_1^2=\frac{1}{n-1}\sum_{i=2}^{n}(y_i-\bar{y}_1)^2\,,s_2^2=\frac{1}{n-1}\sum_{i=2}^{n}(y_i-\bar{y}_2)^2$
  - $\hat{\rho}=\frac{1}{n-1}\sum_{i}^{n}(y_i-\bar{y}_1)(y_{i-1}-\bar{y}_2)/s_2^2$

When both μ and Σ are unknown but structured – AR(1)

- Simulate data from $\rho = 0.6, \mu = 0, \sigma = 1$

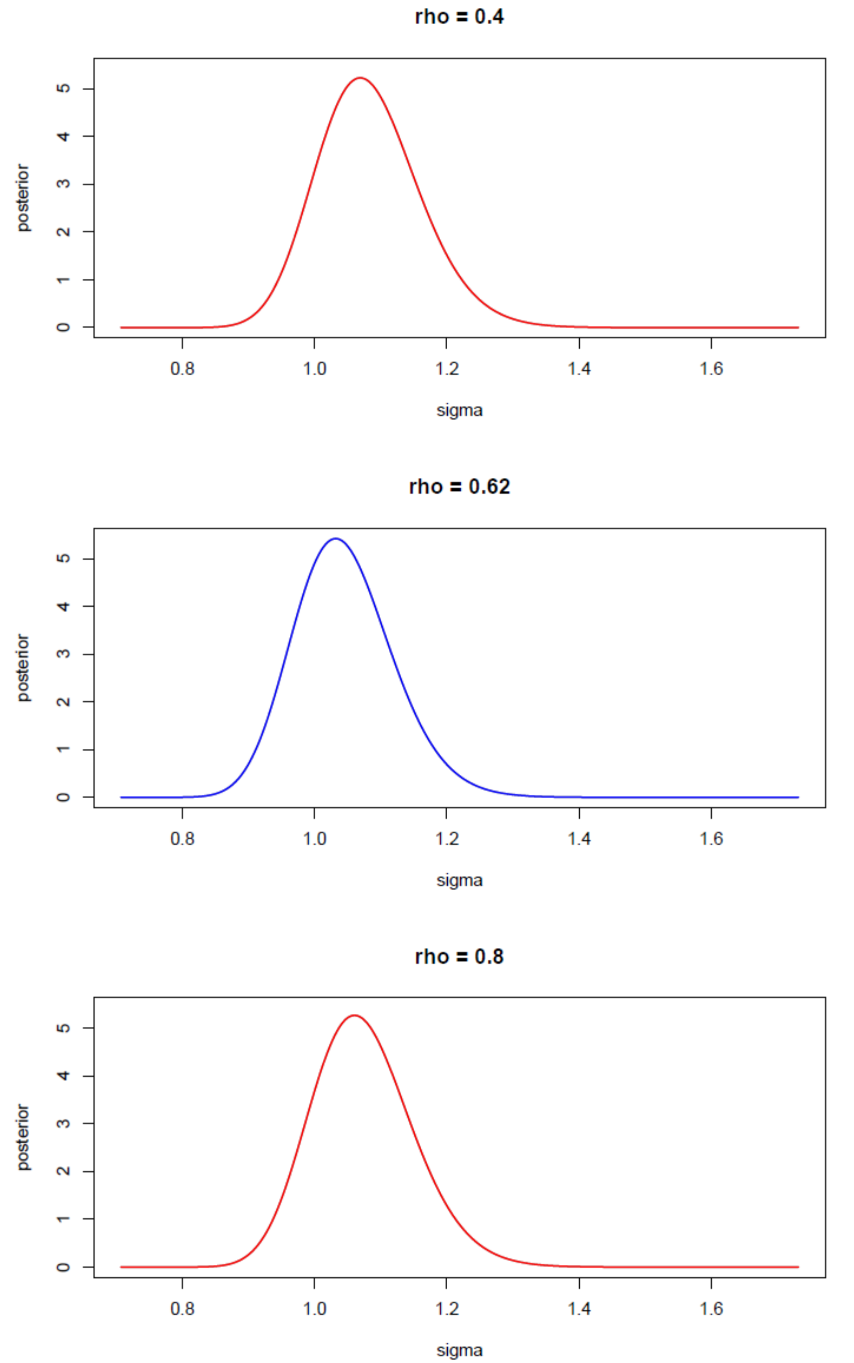


Marginal prior of ρ

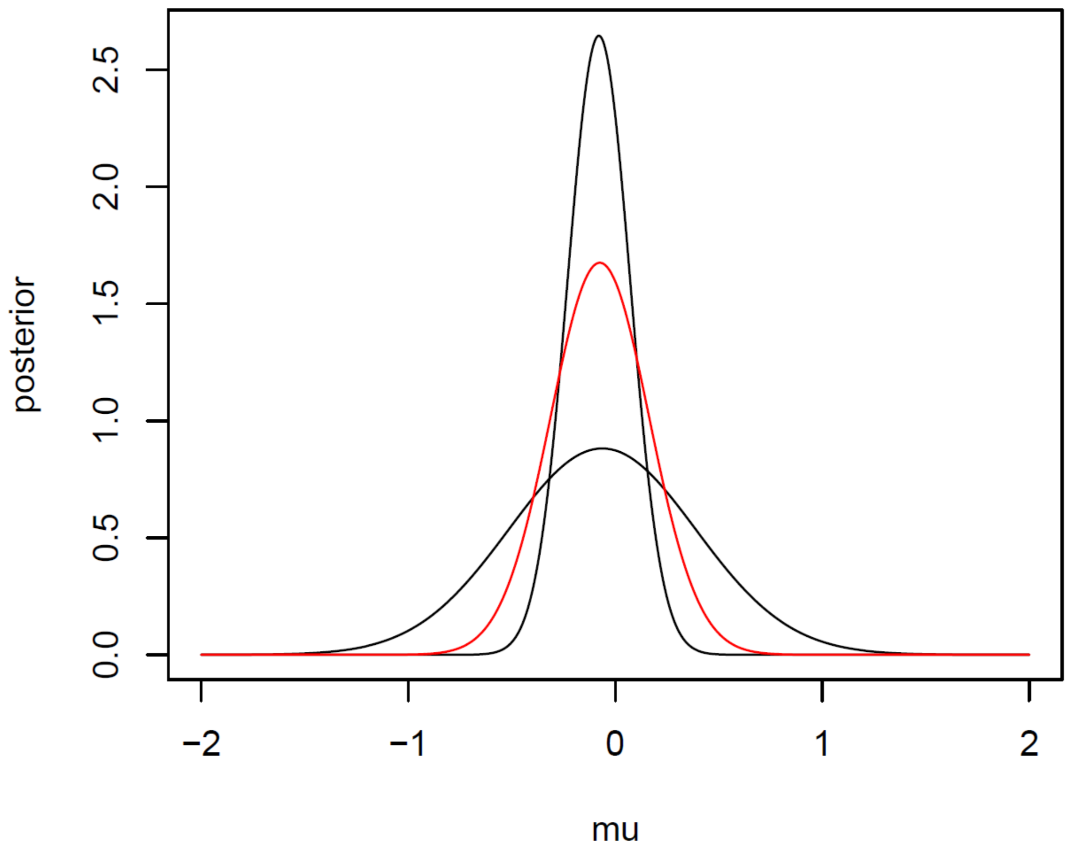


Marginal posterior of ρ

- Posterior of σ conditional on ρ

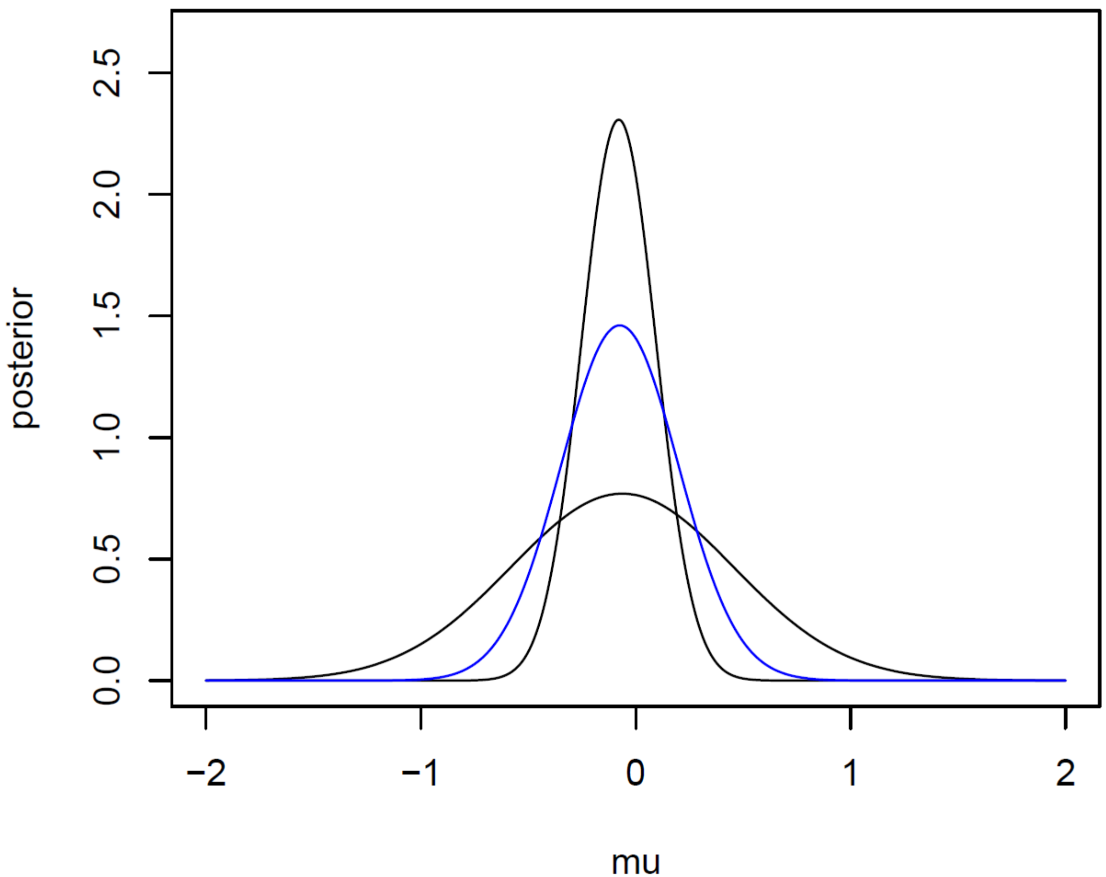


Posterior of μ conditional on $(\rho, \sigma = 0.9)$



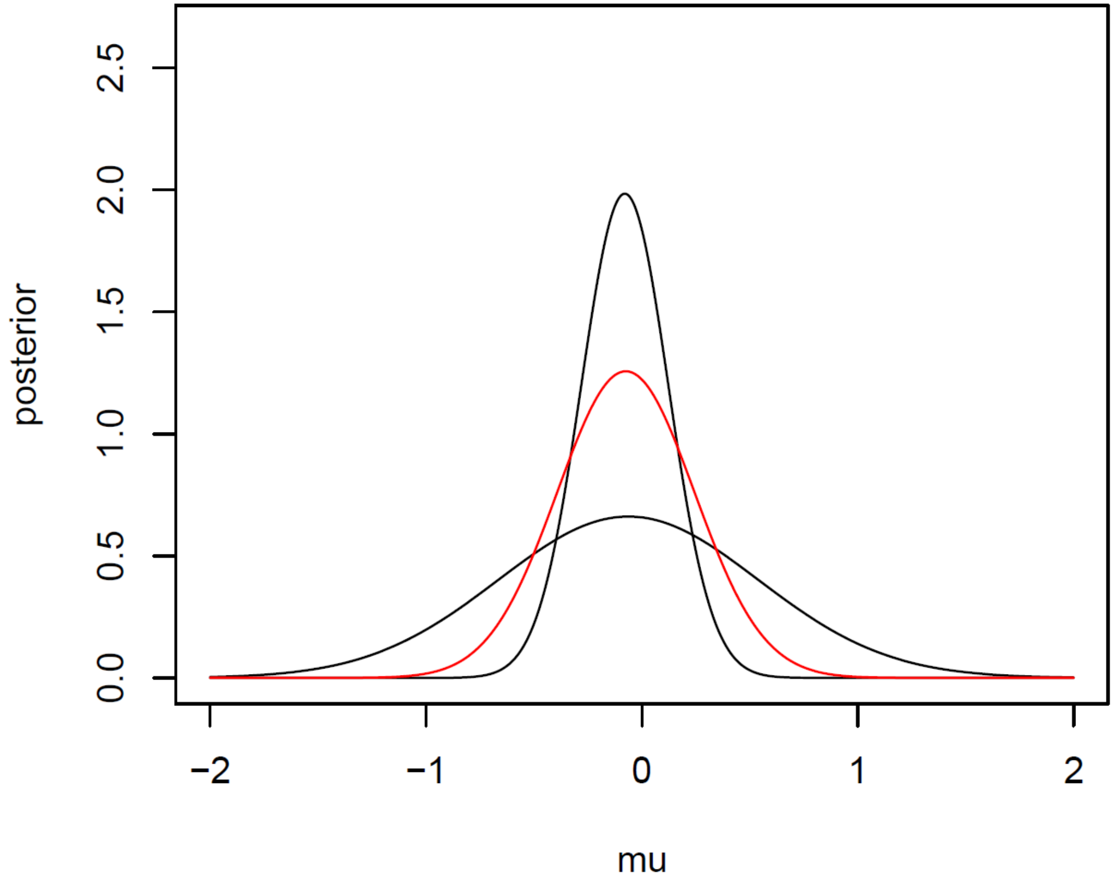
The red curve corresponds to $\rho = 0.62$ and the other two to $\rho = 0.4$ and 0.8.

Posterior of μ conditional on $(\rho, \sigma = 1.032)$



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

Posterior of μ conditional on $(\rho, \sigma = 1.2)$



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

Forecasting

- $f(y_{n+1}, \dots, y_{n+m} | y_1, \dots y_n) = \int f(y_{n+1}, \dots, y_{n+m} | \boldsymbol{\theta}, y_1, \dots y_n) f(\boldsymbol{\theta} | y_1, \dots y_n) d\theta$
- Simulation
  - Generate from the posterior $\boldsymbol{\theta}^* \sim \boldsymbol{\theta} | y_1, \dots y_n$
  - Generate from $y_{n+1}, \dots, y_{n+m} | \boldsymbol{\theta}^*, y_1, \dots y_n$