# Bayesian Multivariate Normal

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

# Outline

Likelihood Distribution

Conjugate Estimation (Normal-Inverse Wishart)

MCMC Gibbs Sampler (Univariate and Multivatiate)

Non-Conjugate Estimation (Laplace-Inverse Wishart)

Regularization

Discussion

Homework

# Likelihood Distribution

We often observe a bivariate (multivariate) continuous random observation $\underset{2\times 1}{x}=(x_1,x_2)'$ that arises from a normal PDF with mean vector $\underset{2\times 1}{\mu}=(\mu_1,\mu_2)'$ and covariance matrix $\underset{2\times 2}{\Sigma}$.

We say $\underset{2\times 1}{x} \sim N(\underset{2\times 1}{\mu}, \underset{2\times 2}{\Sigma})$ and its PDF is

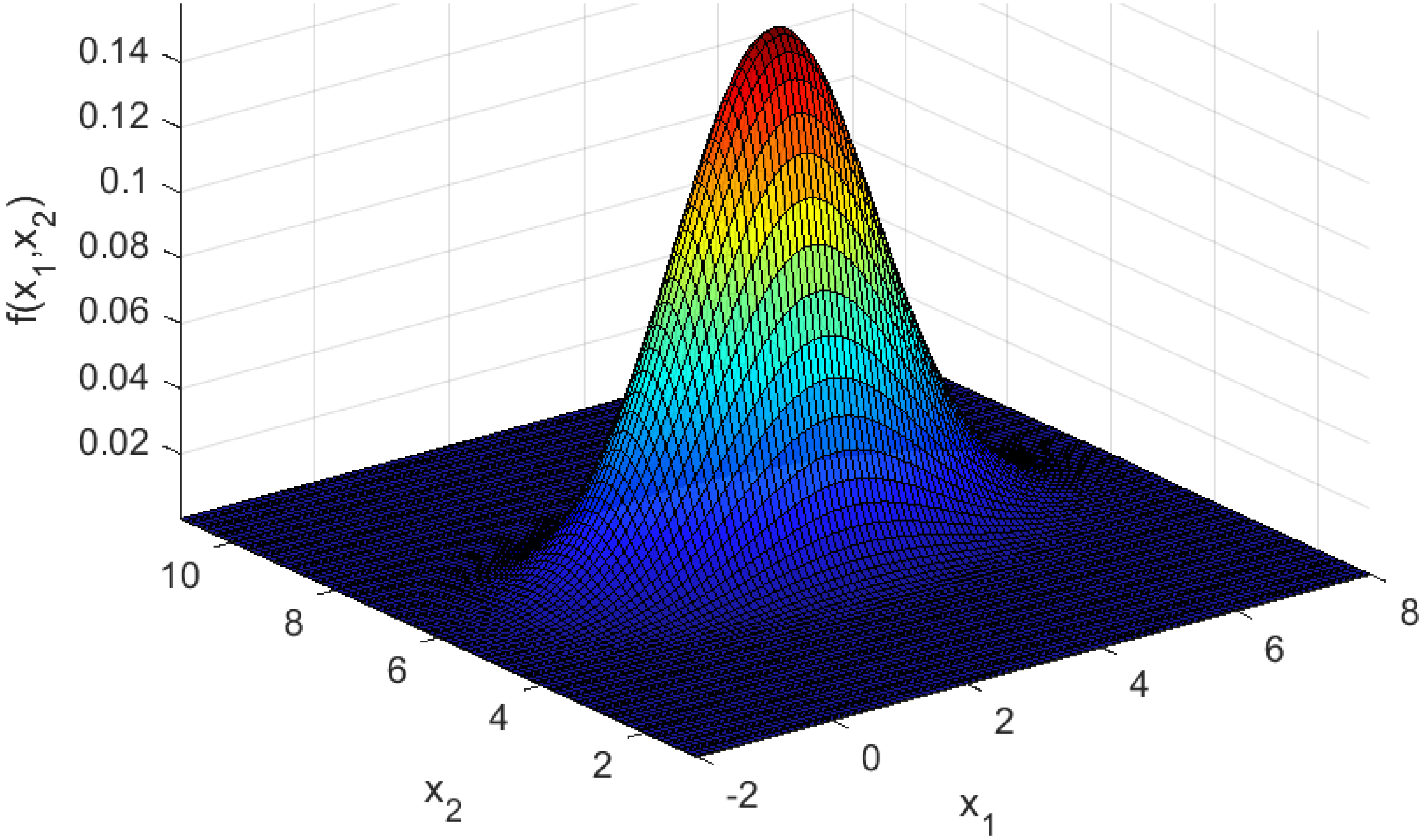
$$f(\underset{2\times 1}{x} \,|\, \underset{2\times 1}{\mu}, \underset{2\times 2}{\Sigma}) = (2\pi)^{-p/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

$$p = 2$$

$$x, \mu \in \mathbb{R}^p$$

$$\Sigma > 0$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$



## Likelihood Distribution

If we observe an iid sample $\underset{2\times 1}{x_1},\ldots,\underset{2\times 1}{x_n}$ from $f(\underset{2\times 1}{x}|\ \underset{2\times 1}{\mu},\ \underset{2\times 2}{\Sigma})$, then

the likelihood of the observations is

$$f(x_1,\ldots,x_n\,|\,\mu,\Sigma)=\prod_{i=1}^{n}(2\pi)^{-p/2}|\Sigma|^{-1/2}\,e^{-\frac{1}{2}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)}$$

which with some algebra becomes

$$f(x_1,\ldots,x_n\,|\,\mu,\Sigma)=(2\pi)^{-np/2}|\Sigma|^{-n/2}\,e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)}\ ,\qquad \begin{array}{c} p=2 \\ x,\mu\in\mathbb{R}^p \\ \Sigma>0 \end{array}\ .$$

Let's not forget that $f(x_1,\ldots,x_n|\mu,\Sigma)$ is the probability of observing particular data values given (assuming that we know) $\mu$ and $\Sigma$!

## Conjugate Estimation

When we expect normal observations $\underset{2\times 1}{x}$ from $N(\underset{2\times 1}{\mu}, \underset{2\times 2}{\Sigma})$, the joint conjugate prior distribution for $\underset{2\times 1}{\mu}$ and $\underset{2\times 2}{\Sigma}$ is

$$f(\underset{p\times 1}{\mu},\underset{p\times p}{\Sigma}) = f(\underset{p\times 1}{\mu}\,|\,\underset{p\times p}{\Sigma})\, f(\underset{p\times p}{\Sigma})$$

Think $p$=2.

Normal for $\mu|\Sigma$

$$f(\underset{p\times 1}{\mu}\,|\,\underset{p\times 1}{\mu_0},\underset{p\times p}{\Sigma},n_0) = (2\pi)^{-p/2}\,|\Sigma/n_0|^{-1/2}\, e^{-\frac{1}{2}(\mu-\mu_0)'(\Sigma/n_0)^{-1}(\mu-\mu_0)}$$

Inverse Wishart for $\Sigma$

$$f(\underset{p\times p}{\Sigma}\,|\,\underset{p\times p}{H},\nu) = k\,|H|^{\nu/2}|\Sigma|^{-(\nu+p+1)/2}\, e^{-\frac{1}{2}tr\Sigma^{-1}H}$$

# Conjugate Estimation

$$trAB = trBA$$

Normal observations $x$'s:

With likelihood

$$f(x_1,\ldots,x_n \mid \underset{p\times 1}{\mu}, \underset{p\times p}{\Sigma}) = (2\pi)^{-np/2} \, |\Sigma|^{-n/2} \, e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)}$$

and priors

$$f(\underset{p\times 1}{\mu} \mid \underset{p\times 1}{\mu_0}, \underset{p\times p}{\Sigma}, n_0) = (2\pi)^{-p/2} \, |\Sigma / n_0|^{-1/2} \, e^{-\frac{n_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)}$$

$$f(\underset{p\times p}{\Sigma} \mid \underset{p\times p}{H}, \nu) = kH^{\nu/2} \, |\Sigma|^{-(\nu+p+1)/2} \, e^{-\frac{1}{2}tr\Sigma^{-1}H} \quad ,$$

Proportional to product of priors and likelihood.

the posterior

$$f(\underset{p\times 1}{\mu}, \underset{p\times p}{\Sigma} \mid x_1,\ldots,x_n) \propto f(x_1,\ldots,x_n \mid \mu,\Sigma) \, f(\mu \mid \mu_0,\Sigma,n_0) \, f(\Sigma \mid H,\nu)$$

$$f(\underset{p\times 1}{\mu}, \underset{p\times p}{\Sigma} \mid x_1,\ldots,x_n) \propto |\Sigma|^{-(\nu_*+p+1)/2} \, e^{-\frac{1}{2}tr\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'+n_0(\mu-\mu_0)(\mu-\mu_0)'+H\right]} \quad .$$

$$\nu_* = \nu + n + 1$$

# Conjugate Estimation

Normal observations $x$'s:

Posterior

$$\underset{p\times 1}{f(\mu},\underset{p\times p}{\Sigma}|x_1,\ldots,x_n)\propto|\Sigma|^{-(\nu_*+p+1)/2}\,e^{-\frac{1}{2}tr\Sigma^{-1}[\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'+n_0(\mu-\mu_0)(\mu-\mu_0)'+H]}$$

utilize the fact that $\nu_*=n+\nu+1$

$$\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)=tr\Sigma^{-1}(X-1_n\mu')'(X-1_n\mu')$$

where $\underset{n\times 2}{X}=(x_1,\ldots,x_n)'$, $\underset{n\times 1}{1_n}=(1,\ldots,1)'$, and $\underset{2\times 1}{\mu}=(\mu_1,\mu_2)'$.

Thus the posterior becomes

$$\underset{p\times 1}{f(\mu},\underset{p\times p}{\Sigma}|x_1,\ldots,x_n)\propto|\Sigma|^{-(\nu_*+p+1)/2}\,e^{-\frac{1}{2}tr\Sigma^{-1}[(X-1_n\mu')'(X-1_n\mu')+n_0(\mu-\mu_0)(\mu-\mu_0)'+H]}\qquad \nu_*=\nu+n+1$$

# Conjugate Estimation

$$\hat{\mu} = (n\bar{x} + n_0\mu_0) / (n_0 + n)$$

Normal observations $x$'s:

With a little algebra, it can be shown that

$$(X - 1_n\mu')'(X - 1_n\mu') + n_0(\mu - \mu_0)(\mu - \mu_0)' + H = (\mu - \hat{\mu})(n_0 + n)(\mu - \hat{\mu})' + W$$

where

$$W = n_0\mu_0\mu_0' + H + X'X - (n_0\mu_0 + n\bar{x})(n_0\mu_0 + n\bar{x})' / (n_0 + n)$$

compare to $p$=1 $\longrightarrow$ $\omega = n_0\mu_0^2 + h + n\overline{x^2} - (n_0\mu_0 + n\bar{x})^2 / (n_0 + n)$

and the joint posterior for $\mu$ and $\Sigma$ becomes

$$f(\underset{p\times 1}{\mu}, \underset{p\times p}{\Sigma} | x_1, \ldots, x_n) \propto |\Sigma|^{-(\nu_* + p + 1)/2} \, e^{-\frac{1}{2} tr\Sigma^{-1}[(\mu - \hat{\mu})(n_0 + n)(\mu - \hat{\mu})' + W]}$$

$$\nu_* = \nu + n + 1$$

# Conjugate Estimation

$$f(\mu|\mu_0,\Sigma,n_0)=(2\pi)^{-p/2}|\Sigma/n_0|^{-1/2}e^{-\frac{n_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)}$$
$$f(\Sigma|H,\nu)=kH^{\nu/2}|\Sigma|^{-(\nu+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}H}$$
$$f(\mu,\Sigma|x's)\propto|\Sigma|^{-(\nu_*+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'+n_0(\mu-\mu_0)(\mu-\mu_0)'+H\right]}$$

Normal observations $x$'s:

The marginal PDF of $\mu$ can be found

$$f(\underset{p\times 1}{\mu}|x_1,\ldots,x_n)=\int_\Sigma f(\mu,\Sigma|x_1,\ldots,x_n)d\Sigma$$

$$f(\underset{p\times 1}{\mu}|x_1,\ldots,x_n)\propto\int_\Sigma|\Sigma|^{-(\nu_*+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}[(\mu-\hat{\mu})(n_0+n)(\mu-\hat{\mu})'+W]}d\Sigma$$

$$f(\underset{p\times 1}{\mu}|x_1,\ldots,x_n)\propto|H_*|^{-\nu_*/2}\underbrace{\int_\Sigma|H_*|^{\nu_*/2}|\Sigma|^{-(\nu_*+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}H_*}d\Sigma}_{=\text{constant}}$$

$$\nu_*=\nu+n+1$$
$$H_*=(\mu-\hat{\mu})(n_0+n)(\mu-\hat{\mu})'+W$$
$$W=n_0\mu_0\mu_0'+H+X'X-(n_0\mu_0+n\bar{x})(n_0\mu_0+n\bar{x})'/(n_0+n)$$

# Conjugate Estimation

$$f(\mu|\mu_0,\Sigma,n_0)=(2\pi)^{-p/2}|\Sigma/n_0|^{-1/2}e^{-\frac{n_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)}$$
$$f(\Sigma|H,\nu)=kH^{\nu/2}|\Sigma|^{-(\nu+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}H}$$
$$f(\mu,\Sigma|x's)\propto|\Sigma|^{-(\nu_*+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'+n_0(\mu-\mu_0)(\mu-\mu_0)'+H\right]}$$

Normal observations $x$'s:

The marginal PDF of $\mu$ can be found

$$f(\underset{p\times 1}{\mu}|x_1,...,x_n)\propto|H_*|^{-\nu_*/2}$$

$$f(\underset{p\times 1}{\mu}|x_1,...,x_n)\propto\frac{1}{|(\mu-\hat{\mu})(n_0+n)(\mu-\hat{\mu})'+W|^{\nu_*/2}}$$
$$f(\underset{p\times 1}{\mu}|x_1,...,x_n)\propto\frac{1}{[1+\frac{1}{\nu_\#}(\mu-\hat{\mu})'T^{-1}(\mu-\hat{\mu})]^{(\nu_\#+p)/2}}$$

Note: If $A$ $p\times q$ and $B$ is $q\times p$, then

$$|I_p+AB|=|I_q+BA|$$

Woodbury's Identity

$$\nu_*=\nu+n+1$$
$$\nu_\#=\nu+n-p+1$$

$$T=\frac{W}{\nu_\#(n_0+n)}$$

$$H_*=(\mu-\hat{\mu})(n_0+n)(\mu-\hat{\mu})'+W$$
$$W=n_0\mu_0\mu_0'+H+X'X-(n_0\mu_0+n\bar{x})(n_0\mu_0+n\bar{x})'/(n_0+n)$$

## Conjugate Estimation

$$f(\mu|\mu_0,\Sigma,n_0)=(2\pi)^{-p/2}|\Sigma/n_0|^{-1/2}e^{-\frac{n_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)}$$

$$f(\Sigma|H,\nu)=kH^{\nu/2}|\Sigma|^{-(\nu+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}H}$$

$$f(\mu,\Sigma|x's)\propto|\Sigma|^{-(\nu_*+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'+n_0(\mu-\mu_0)(\mu-\mu_0)'+H\right]}$$

Normal observations $x$'s:

From the marginal PDF of $\mu$, can be found to be Student-t

$$f(\underset{p\times 1}{\mu}|x_1,\ldots,x_n)=\frac{\Gamma(\frac{\nu_\#+p}{2})\left|T\right|^{-1/2}}{(\nu_\#\pi)^{p/2}\Gamma(\frac{\nu_\#}{2})}\frac{1}{[1+\frac{1}{\nu_\#}(\mu-\hat{\mu})'T^{-1}(\mu-\hat{\mu})]^{(\nu_\#+p)/2}}$$

$$\nu_\#=\nu+n-p+1$$

$$T=\frac{W}{\nu_\#(n_0+n)}$$

the expectation can be found to be

$$E(\underset{p\times 1}{\mu}|x_1,\ldots,x_n)=\hat{\mu}=\frac{n}{n_0+n}\bar{x}+\frac{n_0}{n_0+n}\mu_0$$

and covariance to be

$$\mathrm{cov}(\underset{p\times 1}{\mu}|x_1,\ldots,x_n)=\frac{\nu_\#}{\nu_\#-2}T$$

$$W=n_0\mu_0\mu_0'+H+X'X-(n_0\mu_0+n\bar{x})(n_0\mu_0+n\bar{x})'/(n_0+n)$$

# Conjugate Estimation

$$f(\mu|\mu_0,\Sigma,n_0)=(2\pi)^{-p/2}|\Sigma/n_0|^{-1/2}e^{-\frac{n_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)}$$
$$f(\Sigma|H,\nu)=kH^{\nu/2}|\Sigma|^{-(\nu+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}H}$$
$$f(\mu,\Sigma|x's)\propto|\Sigma|^{-(\nu_*+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'+n_0(\mu-\mu_0)(\mu-\mu_0)'+H\right]}$$

Normal observations $x$'s:

The marginal PDF of $\underset{p\times p}{\Sigma}$ can be found to be

$$f(\underset{p\times p}{\Sigma}|x_1,...,x_n)=\int_\mu f(\mu,\Sigma|x_1,...,x_n)d\mu$$

$$f(\underset{p\times p}{\Sigma}|x_1,...,x_n)\propto\int_\mu|\Sigma|^{-(\nu_*+p+1)/2}e^{-\frac{1}{2}tr\Sigma^{-1}[(\mu-\hat{\mu})(n_0+n)(\mu-\hat{\mu})'+W]}d\mu$$

$$f(\underset{p\times p}{\Sigma}|x_1,...,x_n)\propto|\Sigma|^{-(\nu_*+p)/2}e^{-\frac{1}{2}tr\Sigma^{-1}W}\underbrace{\int_\mu|\Sigma|^{-1/2}e^{-\frac{(n_0+n)}{2}(\mu-\hat{\mu})'\Sigma^{-1}(\mu-\hat{\mu})}d\mu}_{=\text{constant}}$$

$$\nu_*=\nu+n+1$$

$$W=n_0\mu_0\mu_0'+H+X'X-(n_0\mu_0+n\bar{x})(n_0\mu_0+n\bar{x})'/(n_0+n)$$

# Conjugate Estimation

$$f(\mu|\mu_0,\Sigma,n_0) = (2\pi)^{-p/2}|\Sigma/n_0|^{-1/2}\, e^{-\frac{n_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)}$$

$$f(\Sigma|H,\nu) = kH^{\nu/2}|\Sigma|^{-(\nu+p+1)/2}\, e^{-\frac{1}{2}tr\Sigma^{-1}H}$$

$$f(\mu,\Sigma|x's) \propto |\Sigma|^{-(\nu_*+p+1)/2}\, e^{-\frac{1}{2}tr\Sigma^{-1}\left[\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)'+n_0(\mu-\mu_0)(\mu-\mu_0)'+H\right]}$$

Normal observations $x$'s:

The marginal PDF of $\underset{p\times p}{\Sigma}$ can be found to be inverse Wishart

$$f(\underset{p\times p}{\Sigma}|x_1,\ldots,x_n) = k|W|^{(\nu_{**})/2}|\Sigma|^{-(\nu_{**}+p+1)/2}\, e^{-\frac{1}{2}tr\Sigma^{-1}W} \qquad \nu_{**} = \nu + n$$

with expectation, variances, and covariances

$$E(\underset{p\times p}{\Sigma}|x_1,\ldots,x_n) = \frac{W}{\nu_{**}-p-1}\ , \qquad \mathrm{var}(\Sigma_{ii}|x's) = \frac{2W_{ii}^2}{(\nu_{**}-p-1)^2(\nu_{**}-p-3)}$$

$$\mathrm{var}(\Sigma_{ij}|x's) = \frac{(\nu_{**}-p+1)W_{ij}^2+(\nu_{**}-p-1)W_{ii}W_{jj}}{(\nu_{**}-p)(\nu_{**}-p-1)^2(\nu_{**}-p-3)}\ , \qquad \mathrm{cov}(\Sigma_{ij}|x's) = \frac{2W_{ij}W_{kl}+(\nu_{**}-p-1)(W_{ik}W_{jl}+W_{il}W_{kj})}{(\nu_{**}-p)(\nu_{**}-p-1)^2(\nu_{**}-p-3)}\ .$$

$$W = n_0\mu_0\mu_0' + H + X'X - (n_0\mu_0+n\bar{x})(n_0\mu_0+n\bar{x})'/(n_0+n)$$

# MCMC Gibbs Sampler

Normal observations $x$'s:

Quite often we can't marginalize so an alternative is an MCMC Gibbs Sampler which will "generate" a sample joint posterior distribution from which we get sample marginal posterior distributions.

To implement a Gibbs Sampler, we need the posterior conditionals.

Let's examine with univariate observations, $\underset{1\times 1}{x}$, $\underset{1\times 1}{\mu}$, and $\underset{1\times 1}{\sigma^2}$, a simpler case, show that we get the same answer as the pencil and paper marginal, then implement for multivariate observations $\underset{p\times 1}{x}$, $\underset{p\times 1}{\mu}$, and $\underset{p\times p}{\Sigma}$.

# MCMC Gibbs Sampler

Normal observations $x$'s:

Example:

Priors:

$$f(\mu,\sigma^2|\cdot) \propto \frac{\Gamma(\frac{1}{2})h^{(\nu-2)/2}[(\sigma^2)^{(\nu+1)}/n_0]^{-1/2}}{2^{\nu/2}\Gamma(\frac{\nu-2}{2})}\exp\left\{-\frac{n_0(\mu-\mu_0)^2+h}{2\sigma^2}\right\}$$

Likelihood:

$$f(x_1,\ldots,x_n|\mu,\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\right\}$$

Posterior:

$$f(\mu,\sigma^2|x_1,\ldots,x_n,\cdot)=C(\sigma^2)^{-\frac{(\nu+n+1)}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n(x_i-\mu)^2+n_0(\mu-\mu_0)^2+h\right]\right\}$$

**But I'm not very good at math and can't integrate this wrt $\mu$ or $\sigma^2$!**

# MCMC Gibbs Sampler

$$f(\mu,\sigma^2 \mid x_1,\ldots,x_n,\cdot) \propto (\sigma^2)^{-(\nu+n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\mu)^2+(\mu-\mu_0)^2/n_0+h\right]\right\}$$

Normal observations $x$'s:

It is easily shown that the posterior conditionals are

From $f(\mu,\sigma^2 \mid x's)$ gather only terms with $\mu$ for $f(\mu \mid \sigma^2,x's)$ and only terms with σ² for $f(\sigma^2 \mid \mu,x's)$.
Reason out the normalizing constants.

$$f(\mu \mid \sigma^2,x_1,\ldots,x_n,\cdot)=\left[2\pi\frac{\sigma^2}{n_0+n}\right]^{-1/2}\exp\left\{-\frac{(\mu-\hat{\mu})^2}{2\sigma^2/(n_0+n)}\right\}$$

Normal

$$\hat{\mu}=\frac{n_0}{n_0+n}\mu_0+\frac{n}{n_0+n}\bar{x}$$

and

$$f(\sigma^2 \mid \mu,x_1,\ldots,x_n,\cdot)=\frac{\beta_*^{\alpha_*}}{\Gamma(\alpha_*)}(\sigma^2)^{-\alpha_*-1}e^{-\frac{\beta_*}{\sigma^2}}$$

Inverse Gamma

$$\alpha_*=(\nu+n-1)/2$$

$$\beta_*=[(n_0+n)(\mu-\hat{\mu})^2+n_0\mu_0^2+h+n\overline{x^2}-(n_0\mu_0+n\bar{x})^2/(n_0+n)]/2$$

$$E(\mu \mid \sigma^2,x's)=\hat{\mu}$$

$$E(\sigma^2 \mid \mu,x's)=\frac{\beta_*}{\alpha_*-1}$$

i.e. $\mu|\sigma^2,x's \sim N(\hat{\mu},\ \sigma^2/(n_0+n))$ and $\sigma^2|\mu,x's \sim \mathrm{IG}(\alpha_*,\beta_*)$.

# MCMC Gibbs Sampler

$\mu|\sigma^2, x\text{'s} \sim N(\hat{\mu}, \sigma^2/(n_0+n))$ and $\sigma^2|\mu, x\text{'s} \sim IG(\alpha_*, \beta_*)$

$$\alpha_* = (v+n-1)/2$$

$$\beta_* = [(n_0+n)(\mu-\hat{\mu})^2 + n_0\mu_0^2 + h + n\overline{x^2} - (n_0\mu_0 + n\bar{x})^2/(n_0+n)]/2$$

$$\hat{\mu} = \frac{n_0}{n_0+n}\mu_0 + \frac{n}{n_0+n}\bar{x}$$

Normal observations $x$'s:

For Gibbs Sampling we start with an initial value for one of the parameters and alternately generate random observations from the posterior conditionals.

Start with initial $\mu_{(0)}$,

generate a random $\sigma^2_{(1)}$ from $f(\sigma^2|\mu_{(0)}, x_1,\ldots,x_n)$,

generate a random $\mu_{(1)}$ from $f(\mu|\sigma^2_{(1)}, x_1,\ldots,x_n)$,

…

generated a random $\mu_{(L)}$ from $f(\mu|\sigma^2_{(L)},, x_1,\ldots,x_n)$

for a large number $L$ (like $10^6$).

$\cancel{\mu_{(0)}}, \sigma^2_{(1)}, \mu_{(1)}, \sigma^2_{(2)}, \ldots, \ldots, \sigma^2_{(L)}, \mu_{(L)}$

# MCMC Gibbs Sampler

$\mu|\sigma^2,x's \sim N(\hat{\mu}, \sigma^2/(n_0+n))$ and $\sigma^2|\mu,x's \sim IG(\alpha_*,\beta_*)$

$\alpha_* = (v+n-1)/2$

$\beta_* = (n_0+n)(\mu-\hat{\mu})^2 + n_0\mu_0^2 + h + n\overline{x^2} - (n_0\mu_0 + n\bar{x})^2/(n_0+n)$

Normal observations $x$'s:

Generated $\cancel{\mu_{(0)}}, \sigma^2_{(1)}, \mu_{(1)}, \sigma^2_{(2)}, \ldots, \sigma^2_{(L)}, \mu_{(L)}$, for large $L$.

We delete the first $B$ called the "burn-in" for equilibration, then the rest constitute random observations from the joint posterior PDF from which we get marginals $f(\mu|,x_1,\ldots,x_n)$ and $f(\sigma^2|x_1,\ldots,x_n)$.

$$\begin{array}{ccccccccc} \cancel{\mu_{(0)}}, & & ,\mu_{(1)}, & & ,\ldots, & & ,\mu_{(L)} \\ & \searrow & \nearrow & \searrow & \nearrow & \nearrow & \\ & \sigma^2_{(1)}, & & ,\sigma^2_{(2)}, & \ldots, & \sigma^2_{(L)}, & \end{array}$$

Average the $L$-$B$ for marginal expectations.

$$E(\mu|x_1,\ldots,x_n) \approx \frac{1}{L-B}\sum_{i=B+1}^{L}\mu_{(i)} \longrightarrow E(\mu|x's) = \hat{\mu}$$

$$E(\sigma^2|x_1,\ldots,x_n) \approx \frac{1}{L-B}\sum_{i=B+1}^{L}\sigma^2_{(i)} \longrightarrow E(\sigma^2|x's) = \frac{\beta_*}{\alpha_*-1}$$

# MCMC Gibbs Sampler

Normal observations $x$'s:

Example: Heights

Assume that we have assessed our conjugate prior hyperparameters to be

$$n_0 = 10$$

$$\mu_0 = 70$$

$$\sigma_0^2 = 5$$

$$\nu = 9$$

$$h = 45$$

and have a current data set $x_1,\ldots,x_n$ from $N(\mu,\sigma^2)$. $\longrightarrow$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 68.0753 \\ 70.6678 \\ 62.4823 \\ 68.7243 \\ 67.6375 \end{bmatrix}$$

# MCMC Gibbs Sampler

Normal observations $x$'s:

Example: Heights

$\mu|\sigma^2, x\text{'s} \sim N(\hat{\mu}, \sigma^2/(n_0+n))$ and $\sigma^2|\mu, x\text{'s} \sim IG(\alpha_*, \beta_*)$

Prior

$$n_0 = 10$$
$$\mu_0 = 70$$
$$\sigma_0^2 = 5$$
$$\nu = 9$$
$$h = 45$$

Likelihood

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 68.0753 \\ 70.6678 \\ 62.4823 \\ 68.7243 \\ 67.6375 \end{bmatrix}$$

$$n = 5$$
$$\bar{x} = 67.5175$$
$$s^2 = 9.2649$$
$$\overline{x^2} = 4566$$

Initial value: $\mu_{(0)}=68$

Generated, $\sigma^2_{(1)}, \mu_{(1)}, \sigma^2_{(2)}, \ldots, \sigma^2_{(L)}, \mu_{(L)}$

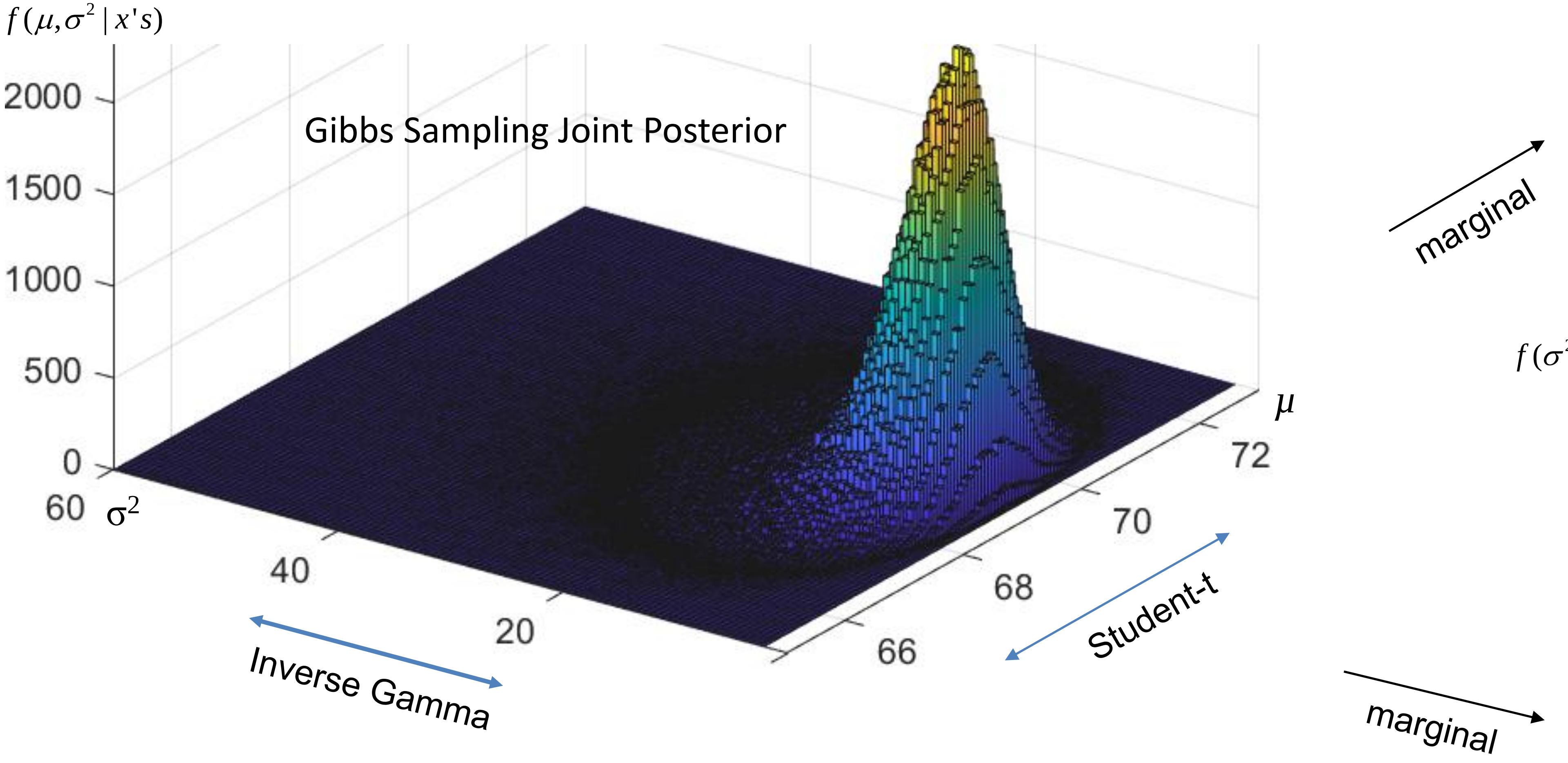
$Burn$=500

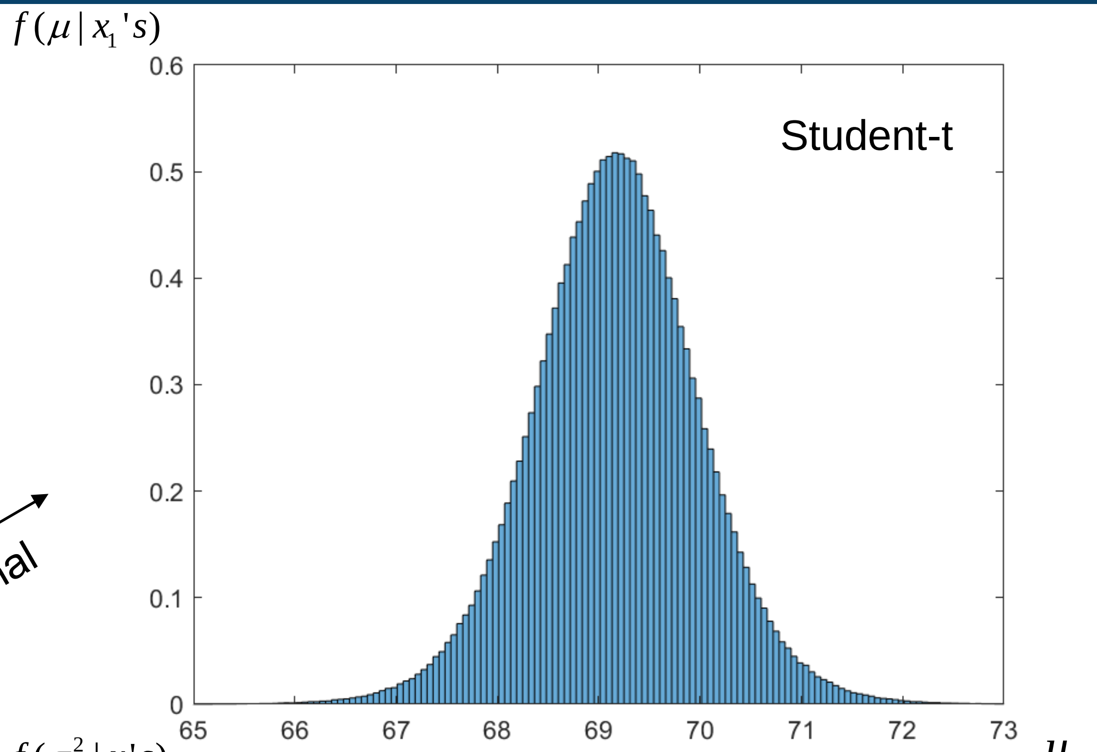
$$\hat{\mu} = \frac{n_0}{n_0+n}\mu_0 + \frac{n}{n_0+n}\bar{x}$$

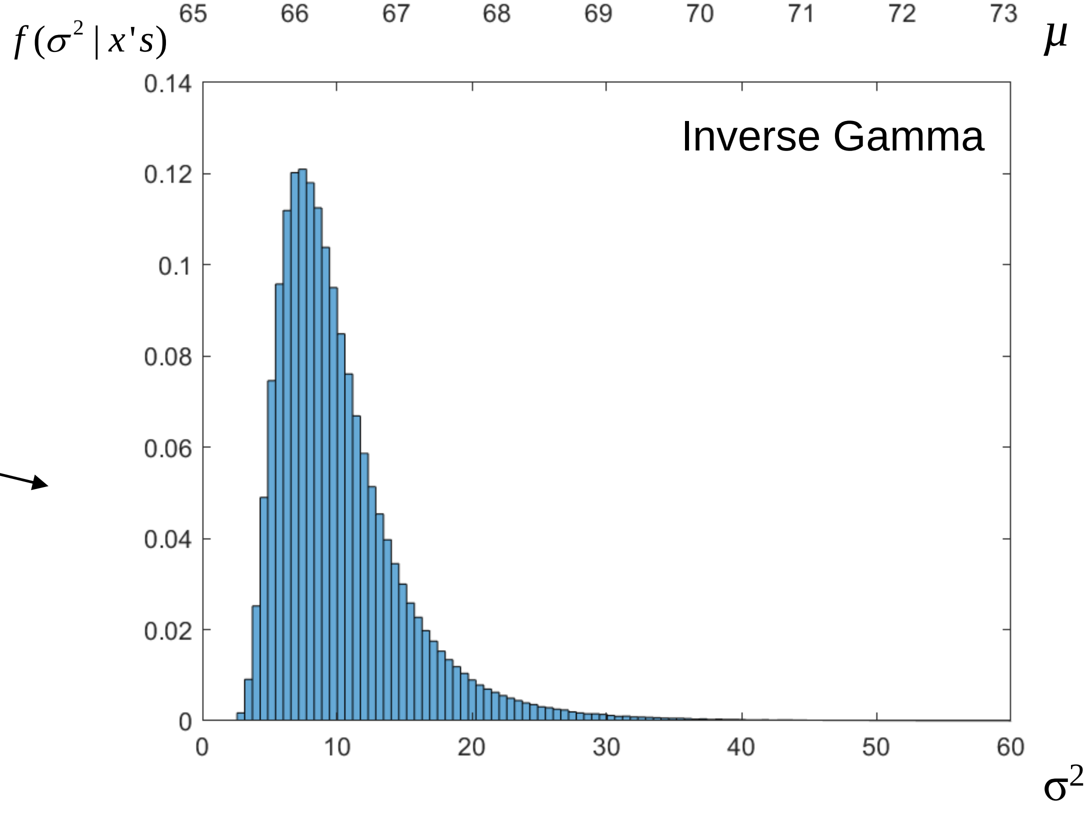
$$\alpha_* = (\nu + n - 1)/2$$
$$\beta_* = [(n_0+n)(\mu - \hat{\mu})^2 + n_0\mu_0^2 + h + n\overline{x^2} - (n_0\mu_0 + n\bar{x})^2/(n_0+n)]/2$$

# MCMC Gibbs Sampler

Normal observations $x$'s:



# MCMC Gibbs Sampler

$$E(\sigma^2 | \mu, x's) = \frac{\beta_*}{\alpha_* - 1}$$

Normal observations $x$'s:

Exact Theoretical:

$$E(\mu | x_1,\ldots,x_n) = 69.1725 \qquad E(\sigma^2 | x_1,\ldots,x_n) = 10.2603$$

$$\text{var}(\mu | x_1,\ldots,x_n) = 0.6840 \qquad \text{var}(\sigma^2 | x_1,\ldots,x_n) = 26.3184$$

Gibbs Sampling:

$$\bar{\mu} = \frac{1}{L-B}\sum_{i=B+1}^{L} \mu_{(i)} = 69.1730 \qquad \overline{\sigma^2} = \frac{1}{L-B}\sum_{i=B+1}^{L} \sigma^2_{(i)} = 10.2623$$

$$s_\mu^2 = \frac{1}{L-B-1}\sum_{i=B+1}^{L} (\mu_{(i)} - \bar{\mu})^2 = 0.6837 \qquad s_{\sigma^2}^2 = \frac{1}{L-B-1}\sum_{i=B+1}^{L} \left(\sigma^2_{(i)} - \overline{\sigma^2}\right)^2 = 26.3334$$

$$\overline{\mu\sigma^2} = \frac{1}{L-B}\sum_{i=B+1}^{L} \mu_{(i)}\sigma^2_{(i)} = 709.8846 \qquad r_{\mu\sigma^2} = 0.0026$$

# MCMC Gibbs Sampler

Normal observations *x*'s:

```
% DB Rowe Univariate
rng('default')
% Priors Hyperparameters
n0=10;
nu=n0-1;
mu0=70;
sigma02=5;
h=nu*sigma02;

% Likelihood data
mu=67;
sigma2=4;
n=5;
x=sqrt(sigma2)*randn(n,1)+
mu;
xbar=mean(x);
s2=var(x);
```

```
% Theoretical Marginal Posterior Means & Variances
muhat=(n0*mu0+n*xbar)/(n0+n);
Emu=(n0*mu0+n*xbar)/(n0+n);
nuast=n+nu-2;
kap =(n0*mu0^2+h+sum(x.^2)-(n0+n)*muhat^2);
tau2=kap/((n+n0)*(n+nu-2));
varmu=nuast/(nuast-2)*tau2;
Esigma2=kap/(nu+n-4);
varsigma2=2*kap^2/((nu+n-4)^2*(nu+n-6));
[Emu,varmu]
[Esigma2,varsigma2]
```

# MCMC Gibbs Sampler

Normal observations $x$'s:

```
%Gibbs Sampler
L=10^6; B=500;
aast=(nu+n-1)/2;
mus=zeros(L+1,1); sigma2s=zeros(L+1,1);
mus(1,1)=68;
for el=2:L+1
    bast=((n0+n)*(mus(el-1,1)-muhat)^2+n0*mu0^2+h+sum(x.^2)-(n0+n)*muhat^2)/2;
    sigma2s(el,1)=1/gamrnd(aast,1/bast);
    mus(el,1)=muhat+sqrt(sigma2s(el,1)/(n0+n))*randn(1,1);
end

% Gibbs Sampling Marginal Posterior Means and Variances
mubar=mean(mus(B+2:L+1-B));
muvar= var(mus(B+2:L+1-B));
sigma2bar=mean(sigma2s(B+2:L+1-B));
sigma2var= var(sigma2s(B+2:L+1-B));
[mubar,muvar]
[sigma2bar,sigma2var]
```

# MCMC Gibbs Sampler

Normal observations $x$'s:

```
% View Posterior Histograms
figure;
histogram(    mus(B+2:L+1,1),300,'Normalization','pdf')
xlim([65,73])
figure;
histogram(sigma2s(B+2:L+1,1),300,'Normalization','pdf')
xlim([0,60])
figure;
hist3([mus(B+2:L+1,1),sigma2s(B+2:L+1,1)],[300,300],'CDataMode','auto',...
    'FaceColor','interp','EdgeColor',[.1,.1,.1])
az=-55; el=35; view(az,el)
xlim([65,73]),ylim([0,60])
```

# MCMC Gibbs Sampler

Normal observations $x$'s:

We can also calculate the marginal expectations of multivariate $\underset{p\times 1}{\mu}$ and $\underset{p\times p}{\Sigma}$ via the Gibbs sampling MCMC technique.

For Gibbs sampling, we need the posterior conditionals

$$f(\underset{p\times 1}{\mu}|\underset{p\times p}{\Sigma},x_1,\ldots,x_n)=\frac{f(\mu,\Sigma,x_1,\ldots,x_n)}{f(\Sigma|x_1,\ldots,x_n)}$$

$$f(\underset{p\times p}{\Sigma}|\underset{p\times 1}{\mu},x_1,\ldots,x_n)=\frac{f(\mu,\Sigma,x_1,\ldots,x_n)}{f(\mu|x_1,\ldots,x_n)}$$

## MCMC Gibbs Sampler

Normal observations $x$'s:

For Gibbs sampling we start with an initial value for one of the parameters and sequentially generate random observations from the posterior conditionals.

Start with $\mu_{(0)}$, generate a random $\Sigma_{(1)}$ from $f(\Sigma|\mu,x_1,\ldots,x_n)$, given $\Sigma_{(1)}$, generate a random $\mu_{(1)}$ from $f(\mu|\Sigma,x_1,\ldots,x_n)$, ... continue until we have generated a random $\mu_{(L)}$ from $f(\mu|\Sigma,x_1,\ldots,x_n)$ for a large number $L$ (like $10^6$).

## MCMC Gibbs Sampler

$$W = n_0\mu_0\mu_0' + H + X'X - (n_0\mu_0 + n\bar{x})(n_0\mu_0 + n\bar{x})'/(n_0 + n)$$

$$\hat{\mu} = \frac{n_0}{n_0 + n}\mu_0 + \frac{n}{n_0 + n}\bar{x}$$

Normal observations $x$'s:

Generated $\mu_{(0)},\Sigma_{(1)},\mu_{(1)},\Sigma_{(2)},\ldots,\Sigma_{(L)},\mu_{(L)}$, for large $L$.

We delete the first $B$ (say 500) called the burn-in in order to ensure equilibration, then average the rest for marginal expectations.

$$\underset{p\times 1}{\bar{\mu}} = \frac{1}{L-B}\sum_{i=B+1}^{L}\mu_{(i)} \longrightarrow E(\mu \mid x's) = \hat{\mu}$$

$$\underset{p\times p}{\bar{\Sigma}} = \frac{1}{L-B}\sum_{i=B+1}^{L}\Sigma_{(i)} \longrightarrow E(\Sigma \mid x_1,\ldots,x_n) = \frac{W}{\nu_{**} - p - 1}$$

$$\overline{\mu_j\Sigma_k} = \frac{1}{L-B}\sum_{i=B+1}^{L}\mu_{j(i)}\Sigma_{k(i)} \longrightarrow E(\mu_j,\Sigma_k \mid x's)$$

# MCMC Gibbs Sampler

From $f(\mu,\Sigma|x's)$ gather only terms with $\mu$ for $f(\mu|\Sigma,x's)$ and only terms with $\Sigma$ for $f(\Sigma|\mu,x's)$ .
Find normalizing constants.

Normal observations $x$'s:

The posterior conditionals can be shown to be

$$f(\mu|\Sigma,x_1,...,x_n) \propto f(\mu,\Sigma|x_1,...,x_n)$$

$$f(\mu|\Sigma,x_1,...,x_n) \propto |\Sigma/(n_0+n)|^{-1/2}\, e^{-\frac{(n_0+n)}{2}(\mu-\hat{\mu})'\Sigma^{-1}(\mu-\hat{\mu})}$$

$$\hat{\mu} = (n\bar{x} + n_0\mu_0)/(n_0+n)$$

$$H_* = (\mu-\hat{\mu})(n_0+n)(\mu-\hat{\mu})' + W$$

$$W = n_0\mu_0\mu_0' + H + X'X$$
$$\quad - (n\bar{x} + n_0\mu_0)(n\bar{x} + n_0\mu_0)'/(n_0+n)$$

and

$$f(\Sigma|\mu,x_1,...,x_n) \propto f(\mu,\Sigma|x_1,...,x_n)$$

$$f(\Sigma|\mu,x_1,...,x_n) \propto |\Sigma|^{-(\nu_*+p+1)/2}\, e^{-\frac{1}{2}tr\Sigma^{-1}H_*}$$

$\Sigma|\mu,x_1,\ldots,x_n$, is inverse Wishart and $\mu|\Sigma,x_1,\ldots,x_n$ is normal.

# MCMC Gibbs Sampler

$$W = n_0\mu_0\mu_0' + H + X'X - (n\bar{x} + n_0\mu_0)(n\bar{x} + n_0\mu_0)'/(n_0 + n)$$

Normal observations $x$'s:

Generated $\cancel{\mu_{(0)}}, \Sigma_{(1)}, \mu_{(1)}, \Sigma_{(2)}, \ldots, \Sigma_{(L)}, \mu_{(L)}$, for large $L$.

We delete the first $B$ called the "burn-in" for equilibration, then the rest constitute random observations from the joint posterior PDF from which we get marginals $f(\mu|, x_1, \ldots, x_n)$ and $f(\Sigma\,|x_1, \ldots, x_n)$.

$\cancel{\mu_{(0)}}, \quad \Sigma_{(1)}, \quad ,\mu_{(1)}, \quad ,\Sigma_{(2)}, \ldots, \quad ,\ldots, \quad \Sigma_{(L)}, \quad ,\mu_{(L)}$

Average the $L$-$B$ for marginal expectations.

$$E(\mu\,|\,x_1,\ldots,x_n) \approx \frac{1}{L-B}\sum_{i=B+1}^{L}\mu_{(i)} \longrightarrow E(\mu\,|\,x's) = \hat{\mu}$$

$$E(\Sigma\,|\,x_1,\ldots,x_n) \approx \frac{1}{L-B}\sum_{i=B+1}^{L}\Sigma_{(i)} \longrightarrow E(\Sigma\,|\,x's) = \frac{W}{n+\nu-p-1}$$

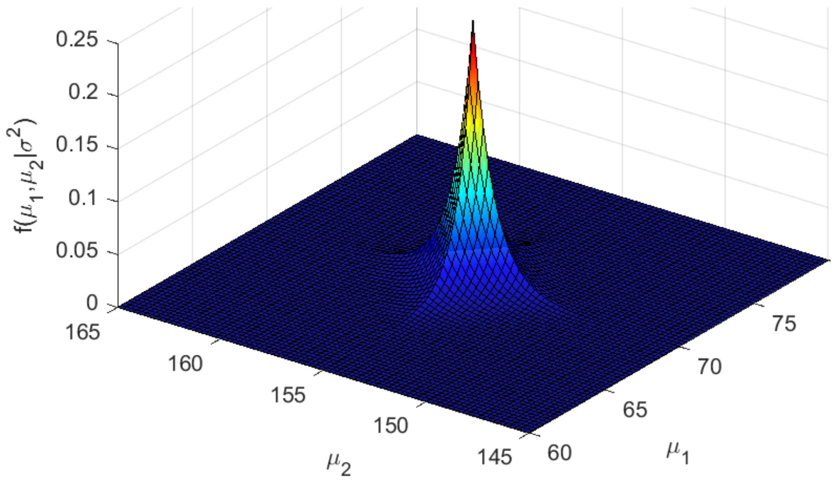
# Non-Conjugate Estimation

Non-conjugate prior distributions for bivariate (multivariate) normal observations are not simple. One common strategy is to assume independent priors for the $\mu$'s and common variance with no covariance, $\Sigma=\sigma^2 I_p$. (Will not capture correlation!)

$$E(\mu_j \mid \sigma^2) = \mu_{0j}$$
$$\mathrm{var}(\mu_j \mid \sigma^2) = \frac{2\sigma^2}{n_j}$$
$$j = 1,2$$

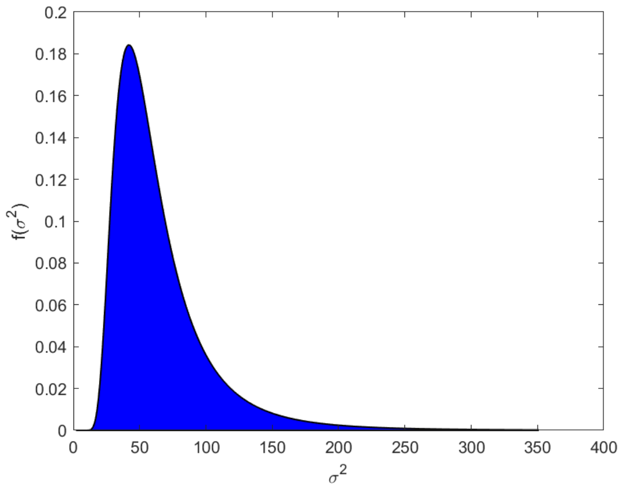
Laplace (independent)

$$f(\mu_1 \mid \sigma^2, n_1) f(\mu_2 \mid \sigma^2, n_2) = \frac{1}{4\sigma^2/n_1} e^{-\frac{|\mu_1-\mu_{01}|}{2\sigma^2/n_1}} \frac{1}{4\sigma^2/n_2} e^{-\frac{|\mu_2-\mu_{02}|}{2\sigma^2/n_2}}$$




Inverse Gamma

$$f(\sigma^2 \mid h, \nu) = \frac{h^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} (\sigma^2)^{-(\nu+2)/2} e^{-\frac{h}{2\sigma^2}}$$



# Non-Conjugate Estimation

Normal observations $x$'s:

Likelihood

$$f(x_1,\ldots,x_n \mid \mu,\sigma^2) = (2\pi\sigma^2)^{-np/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_{i1}-\mu_1)^2+(x_{i2}-\mu_2)^2]}$$

Priors

$$f(\mu_1 \mid \sigma^2, n_1)\, f(\mu_2 \mid \sigma^2, n_2) = \frac{n_1 n_2}{16(\sigma^2)^{4/2}} e^{-\frac{1}{2\sigma^2}[n_1|\mu_1-\mu_{01}|+n_2|\mu_2-\mu_{02}|]}$$

$$f(\sigma^2 \mid h,\nu) = \frac{h^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}(\sigma^2)^{-(\nu+2)/2} e^{-\frac{h}{2\sigma^2}}$$

Proportional to product of priors and likelihood.

Posterior

$$f(\mu,\sigma^2 \mid x_1,\ldots,x_n) = \frac{f(x_1,\ldots,x_n \mid \mu,\sigma^2)\, f(\mu_1 \mid \sigma^2)\, f(\mu_2 \mid \sigma^2)\, f(\sigma^2)}{f(x_1,\ldots,x_n)}$$

$$f(\mu,\sigma^2 \mid x_1,\ldots,x_n) = \frac{C}{(\sigma^2)^{(np+\nu+6)/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_{i1}-\mu_1)^2+(x_{i2}-\mu_2)^2]+n_1|\mu_1-\mu_{01}|+n_2|\mu_2-\mu_{02}|+h}$$

# Non-Conjugate Estimation

Normal observations $x$'s:

Posterior

$$f(\mu,\sigma^2 \mid x_1,\ldots,x_n) = \frac{C}{(\sigma^2)^{(np+\nu+6)/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_{i1}-\mu_1)^2+(x_{i2}-\mu_2)^2]+n_1|\mu_1-\mu_{01}|+n_2|\mu_2-\mu_{02}|+h}$$

$$C = \frac{n_1 n_2}{16(\sigma^2)^{4/2}} \frac{h^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \frac{(2\pi\sigma^2)^{-np/2}}{f(x_1,\ldots,x_n)}$$

$$f(x_1,\ldots,x_n) = \int f(x_1,\ldots,x_n \mid \mu,\sigma^2) f(\mu_1 \mid \sigma^2) f(\mu_2 \mid \sigma^2) f(\sigma^2) d\mu_1 d\mu_2 d\sigma^2$$

$$f(x_1,\ldots,x_n) = \int \frac{n_1 n_2 h^{\frac{\nu}{2}}(2\pi)^{-np/2}}{16\cdot 2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} (\sigma^2)^{(np+\nu+6)/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_{i1}-\mu_1)^2+(x_{i2}-\mu_2)^2]+n_1|\mu_1-\mu_{01}|+n_2|\mu_2-\mu_{02}|+h} d\mu_1 d\mu_2 d\sigma^2$$

# Non-Conjugate Estimation

Proportional to product of priors and likelihood.

## Normal observations $x$'s:

Posterior

$$f(\mu,\sigma^2 \mid x_1,\ldots,x_n) = \frac{C}{(\sigma^2)^{(np+\nu+6)/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_{i1}-\mu_1)^2+(x_{i2}-\mu_2)^2]+n_1|\mu_1-\mu_{01}|+n_2|\mu_2-\mu_{02}|+h}$$

some algebra leads to

$$f(\mu_1,\mu_2,\sigma^2 \mid x_1,\ldots,x_n,\cdot) = \frac{C}{(\sigma^2)^{(np+\nu+6)/2}} e^{-\frac{1}{2\sigma^2}S(\mu_1,\mu_2)}$$

$$\begin{aligned} S(\mu_1,\mu_2) = {} & n(\mu_1-\bar{x}_1)^2 + n_1|\mu_1-\mu_{01}| \\ & + n(\mu_2-\bar{x}_2)^2 + n_2|\mu_2-\mu_{02}| \\ & + \sum_{i=1}^{n}(x_{i1}^2+x_{i2}^2) - n^2(\bar{x}_1^2+\bar{x}_2^2) + h \end{aligned}$$

Common to assume

$$|\mu_1-\mu_{01}| = \sqrt{(\mu_1-\mu_{01})^2+\delta}$$

where $\delta$ is very small. Differentiable.

# Non-Conjugate Estimation

Normal observations $x$'s:

Then the most common thing to do is to find MAP estimates.

$$S(\mu_1,\mu_2)=n(\mu_1-\bar{x}_1)^2+n_1|\mu_1-\mu_{01}|+n(\mu_2-\bar{x}_2)^2+n_2|\mu_2-\mu_{02}|$$
$$+\sum_{i=1}^{n}(x_{i1}^2+x_{i2}^2)-n(\bar{x}_1^2+\bar{x}_2^2)+h$$

Find $\mu_1$ that minimizes

$$S(\mu_1)=\frac{n}{n+n_1}(\mu_1-\bar{x}_1)^2+\frac{n_1}{n+n_1}|\mu_1-\mu_{01}|$$

and $\mu_2$ that minimizes

$$S(\mu_2)=\frac{n}{n+n_2}(\mu_2-\bar{x}_2)^2+\frac{n_2}{n+n_2}|\mu_2-\mu_{02}|$$

Common to assume

$$|\mu_j-\mu_{0j}|=\sqrt{(\mu_j-\mu_{0j})^2+\delta}$$

where $\delta$ is very small. Differentiable.

**Example:**

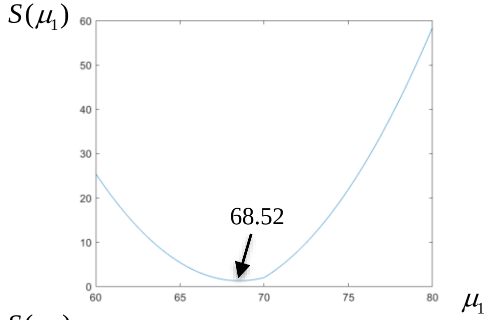
$n_1=n_2=10=n_0$

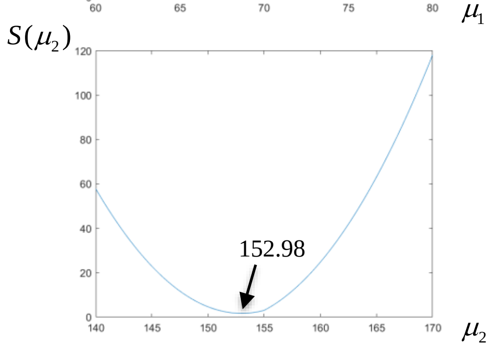
$\mu_1=70,\mu_2=155$

$n=10,\nu=9$

$\sigma_0^2=5,h=45$

$\bar{x}=67.5175$




But one can computationally calculate expectations!

$$E(\mu_j\,|\,x_1,\ldots,x_n)\qquad E(\sigma^2\,|\,x_1,\ldots,x_n)$$

## Regularization

You will see, especially in engineering that a “score” function is formed of the form

$$S(\mu_1,\mu_2)=\underbrace{\sum_{i=1}^{n}[(x_{i1}-\mu_1)^2+(x_{i2}-\mu_2)^2]}_{\text{likelihood term(s)}}+\underbrace{R(\mu_1,\mu_2)}_{\text{prior term(s)}}$$

where $R(\mu_1,\mu_2)$ is called a “regularizer” with no Bayesian context. The regularizer is used to obtain identifiability. The sole objective is to minimize the score function with respect to $\mu_1$ and $\mu_2$. This yields the MAP but not entire PDF so no CIs. Often the regularizer assumes a prior mean of zero!

$\mu_0=0$

$|\mu_1-\mu_0|$

$(\mu_1-\mu_0)^2$

When you see a regularizer, think “What prior are we assuming?”

## Discussion

# Questions?

Likelihood

$$f(x|\mu,\Sigma)=(2\pi)^{-p/2}|\Sigma|^{-1/2}\,e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

Priors

$$f(\mu|\mu_0,\Sigma,n_0)=(2\pi)^{-p/2}|\Sigma/n_0|^{-1/2}\,e^{-\frac{1}{2}(\mu-\mu_0)'(\Sigma/n_0)^{-1}(\mu-\mu_0)}$$

$$f(\Sigma|H,\nu)=kH^{\nu/2}|\Sigma|^{-(\nu+p+1)/2}\,e^{-\frac{1}{2}tr\Sigma^{-1}H}$$

$$f(\mu|\sigma^2)=\frac{n_1n_2}{16(\sigma^2)^{4/2}}e^{-\frac{1}{2\sigma^2}[n_1|\mu_1-\mu_{01}|+n_2|\mu_2-\mu_{02}|]}$$

$$\underset{\beta}{\text{ArgMax}}\ f(\beta,\sigma^2|x's)$$

$$\underset{\sigma^2}{\text{ArgMax}}\ f(\sigma^2|\beta,x's)$$

Marginal Posteriors

$$f(\mu|x_1,\ldots,x_n)\propto\frac{1}{[1+\frac{1}{\nu_\#}(\mu-\hat{\mu})'T^{-1}(\mu-\hat{\mu})]^{(\nu_\#+p)/2}}$$

$$f(\Sigma|x_1,\ldots,x_n)=k|W|^{(\nu_{**})/2}|\Sigma|^{-(\nu_{**}+p+1)/2}\,e^{-\frac{1}{2}tr\Sigma^{-1}W}$$

Gibbs Sampler

$$\cancel{\mu_{(0)}},\ \Sigma_{(1)},\ \mu_{(1)},\ \Sigma_{(2)},\ldots,\ \ldots,\Sigma_{(L)},\ \mu_{(L)}\ \longrightarrow\ \bar{\mu}=\frac{1}{L-B}\sum_{i=B+1}^{L}\mu_{(i)}\qquad \bar{\Sigma}=\frac{1}{L-B}\sum_{i=B+1}^{L}\Sigma_{(i)}$$

# Homework 10

1. Show that the exponent in the posterior distribution is

$$(X-1_n\mu')'(X-1_n\mu')+n_0(\mu-\mu_0)(\mu-\mu_0)'+H$$

$$=(\mu-\hat{\mu})(n_0+n)(\mu-\hat{\mu})'+W$$

$$W=n_0\mu_0\mu_0'+H+X'X-(n\bar{x}+n_0\mu_0)(n\bar{x}+n_0\mu_0)'/(n_0+n)$$

# Homework 10

2. Derive the marginal posterior PDF for $\mu|x$'s with conjugate priors

$$f(\mu \mid x_1,\ldots,x_n) \propto \frac{1}{[1+\frac{1}{\nu_\#}(\mu-\hat{\mu})T^{-1}(\mu-\hat{\mu})']^{\nu_*/2}}$$

Hint: This is an inverse Wishart PDF integral.

, where $\nu_*=n+\nu+1$,

$$\nu_\# = n+\nu-p+1$$

$$\hat{\mu} = \frac{n}{n_0+n}\bar{x} + \frac{n_0}{n_0+n}\mu_0$$ , and $$T = \frac{W}{\nu_\#(n_0+n)}$$ .

3. Derive the marginal posterior PDF for $\Sigma|x$'s with conjugate priors

$$f(\Sigma \mid x_1,\ldots,x_n) = k\,|W|^{(n+\nu)/2}|\Sigma|^{-(n+\nu+p+1)/2}\, e^{-\frac{1}{2}tr\Sigma^{-1}W}.$$

Hint: This is a Normal PDF integral.

$$W = n_0\mu_0\mu_0' + H + X'X - (n\bar{x}+n_0\mu_0)(n\bar{x}+n_0\mu_0)'/(n_0+n)$$

# Homework 10

4**.You plan to see bivariate normal observations, $\underset{2\times1}{x}\sim N(\underset{2\times1}{\mu}, \underset{2\times2}{\Sigma})$.

You assess conjugate normal-inverse Wishart priors

with hyperparameters $n_0$=10, $\mu_0$=(69,145)', $\nu$=$n_0$-1,

and $H$=$n_0\times$[7,0;0,21].

$$\nu_\# = \nu + n - p + 1$$

Generate (observe) $n$=10 bivariate HW's from the normal PDF

with $\mu$=(67,150)' and $\Sigma$=[4,6;6,16]. Calculate posterior estimates

$$\hat{\mu} = \frac{n\bar{x} + n_0\mu_0}{n_0 + n} \qquad \hat{\Sigma}_\mu = \frac{\nu_\#}{\nu_\# - 2}\frac{W}{\nu_\#(n_0 + n)}$$

$$W = n_0\mu_0\mu_0' + H + X'X - (n\bar{x} + n_0\mu_0)(n\bar{x} + n_0\mu_0)'/(n_0 + n)$$

** For students that have had 6010 and 6020.

## Homework 10

5**.Repeat problem 4. a total of $m$ times to obtain $\hat{\mu}_1,\ldots,\hat{\mu}_m$ and $\hat{\Sigma}_{\mu 1},\ldots,\hat{\Sigma}_{\mu m}$. Make histograms of 11 and 12 elements of $\mu$'s. Make histograms of the 11, 22, and 12 elements of $\Sigma$'s.

Calculate means, variances, and covariances of the elements of $\mu$'s and $\Sigma$'s.
Compare to the theoretical values and comment.

** For students that have had 6010 and 6020.

# Homework 10

6**.Implement an MCMC Gibbs sampler to generate $\mu_{(0)},\Sigma_{(1)},\mu_{(1)},\Sigma_{(2)},\ldots,\Sigma_{(L)},\mu_{(L)}$, for large $L$ to estimate

$$E(\mu \mid x_1,\ldots,x_n) \approx \frac{1}{L-B}\sum_{i=B+1}^{L}\mu_{(i)}$$

$$E(\Sigma \mid x_1,\ldots,x_n) \approx \frac{1}{L-B}\sum_{i=B+1}^{L}\Sigma_{(i)}$$

** For students that have had 6010 and 6020.

Compare your results to the exact results in 4.

$$E(\mu \mid x_1,\ldots,x_n) = \frac{n\bar{x}+n_0\mu_0}{n_0+n} \qquad E(\Sigma \mid x_1,\ldots,x_n) = \frac{W}{\nu_* - p - 1}$$

$$\nu_* = \nu + n + 1$$

$$W = n_0\mu_0\mu_0' + H + X'X - (n\bar{x}+n_0\mu_0)(n\bar{x}+n_0\mu_0)'/(n_0+n)$$

## Homework 10

***Show off Question

7***.Assume that you have independent Laplace priors and an independent likelihood. Assume that you have assessed hyperparameters $n_1$=20, $\mu_0$=69, $n_2$=10, $\mu_{02}$=145. Using the same random sample $x_1,\ldots,x_n$ in problem 4, $n$=10 from $N(\mu,\Sigma)$. Numerically find MAPs $\mu_1$ and $\mu_2$.

$$S(\mu_1)=\frac{n}{n+n_1}(\mu_1-\bar{x}_1)^2+\frac{n_1}{n+n_1}|\mu_1-\mu_{01}|$$

$$S(\mu_2)=\frac{n}{n+n_2}(\mu_2-\bar{x}_2)^2+\frac{n_2}{n+n_2}|\mu_2-\mu_{02}|$$

$\mu$=(67,150)'
$\Sigma$=[4,6;6,16]

Compare to 4. What if you repeated $10^6$ times?

Variances and CIs not possible because MAPs.

# Homework 10

***Show off Question

8***.Assume that you quantified your available prior information for each mean as $f(\mu_j|\sigma^2)=Cf_N(\mu_j|\sigma^2)f_L(\mu_j|\sigma^2)$. For $j=1,2$

$$f(\mu_j \mid \mu_{0j},\sigma^2,n_{1j},n_{2j}) = C\frac{1}{\sqrt{2\pi\sigma^2/n_{1j}}}e^{-\frac{(\mu_j-\mu_{0j})^2}{2\sigma^2/n_{1j}}}\frac{1}{4\sigma^2/n_{2j}}e^{-\frac{|\mu_j-\mu_{0j}|}{2\sigma^2/n_{2j}}}$$

with normal likelihood and inverse gamma prior for $\sigma^2$.

The score functions are:

$$S(\mu_j) = \frac{n}{n+n_{1j}+n_{2j}}(\mu_1-\bar{x}_1)^2 + \frac{n_{1j}}{n+n_{1j}+n_{2j}}(\mu_1-\mu_{0j})^2 + \frac{n_{2j}}{n+n_{1j}+n_{2j}}|\mu_1-\mu_{0j}|$$

Using same data in 4, estimate with new score functions.

Compare to problems 4 and 7. What if repeated $10^6$ times?