STAT 576 Bayesian Analysis

Lecture 3: Bayesian Inference II

Chencheng Cai

Washington State University

Recap: Single Parameter Bayesian Inference

- Bayesian Inference Procedure:
  - Name a prior
  - Get the posterior (proportional notation)
  - Point estimators: MAP, posterior mean, etc..
  - Credible interval: QBI, HDR.
  - Prediction for new observations.

Recap: Single Parameter Bayesian Inference

- Bayesian Inference Procedure:
  - Name a prior
  - Get the posterior (proportional notation)
  - Point estimators: MAP, posterior mean, etc..
  - Credible interval: QBI, HDR.
  - Prediction for new observations.
- Prior Elicitation:
  - Conjugate Prior
  - Uninformative Prior / Jeffreys Prior
  - (Improper Prior Distribution)

Recap: Single Parameter Bayesian Inference

- Bayesian Inference Procedure:
  - Name a prior
  - Get the posterior (proportional notation)
  - Point estimators: MAP, posterior mean, etc..
  - Credible interval: QBI, HDR.
  - Prediction for new observations.
- Prior Elicitation:
  - Conjugate Prior
  - Uninformative Prior / Jeffreys Prior
  - (Improper Prior Distribution)
- Important Examples:
  - Normal with known variance: $p(\theta) \propto 1$ (conj. prior: Normal)
  - Normal with known mean: $p(\sigma^2) \propto (\sigma^2)^{-1}$ (conj. prior: inv-Gamma)

Nuisance Parameter

- **Nuisance** parameters are parameters that are unknown and of no interest.

Nuisance Parameter

- **Nuisance** parameters are parameters that are unknown and of no interest.
- Suppose the unknown parameter is $\theta = (\theta_1, \theta_2)$.
- A well-defined observation model gives

$$y \mid \theta_1, \theta_2$$

- A Bayesian inference needs to define a prior for both θ_1 and θ_2: $p(\theta_1, \theta_2)$
- Then the **joint** posterior is obtained by

$$p(\theta_1, \theta_2 \mid y) \propto p(\theta_1, \theta_2) p(y \mid \theta_1, \theta_2)$$

Nuisance Parameter

- **Nuisance** parameters are parameters that are unknown and of no interest.
- Suppose the unknown parameter is $\theta = (\theta_1, \theta_2)$.
- A well-defined observation model gives

$$y \mid \theta_1, \theta_2$$

- A Bayesian inference needs to define a prior for both θ_1 and θ_2: $p(\theta_1, \theta_2)$
- Then the **joint** posterior is obtained by

$$p(\theta_1, \theta_2 \mid y) \propto p(\theta_1, \theta_2) p(y \mid \theta_1, \theta_2)$$

- If we are only interested in θ_1, we need to get the **marginal** posterior for θ_1:

$$p(\theta_1 \mid y) = \int p(\theta_1, \theta_2 \mid y) d\mu(\theta_2)$$

Nuisance Parameter

- An important observation for the marginal posterior is

$$p(\theta_1 \mid y) \propto \int p(\theta_1 \mid \theta_2, y) p(\theta_2 \mid y) d\mu(\theta_2)$$

Nuisance Parameter

- An important observation for the marginal posterior is

$$p(\theta_1 \mid y) \propto \int p(\theta_1 \mid \theta_2, y) p(\theta_2 \mid y) d\mu(\theta_2)$$

- First observation:
  - In order to draw samples from $p(\theta_1 \mid y)$
  - We may first draw θ_2 from $p(\theta_2 \mid y)$ (if it is much easier)
  - Then draw θ_1 from $p(\theta_1 \mid \theta_2, y)$ with θ_2 drawn in the first step.

Nuisance Parameter

- An important observation for the marginal posterior is

$$p(\theta_1 \mid y) \propto \int p(\theta_1 \mid \theta_2, y) p(\theta_2 \mid y) d\mu(\theta_2)$$

- First observation:
  - In order to draw samples from $p(\theta_1 \mid y)$
  - We may first draw θ_2 from $p(\theta_2 \mid y)$ (if it is much easier)
  - Then draw θ_1 from $p(\theta_1 \mid \theta_2, y)$ with θ_2 drawn in the first step.
- Second observation:
  - In order to construct a conjugate joint prior
  - We may find a conjugate prior for the conditional observation model:

    $$p(y \mid \theta_1, \theta_2)$$

    with fixed θ_2
  - Then find a conjugate prior for the marginal observation model:

$$p(y \mid \theta_2) = \int p(y \mid \theta_1, \theta_2) p(\theta_1 \mid \theta_2) d\mu(\theta_1)$$

Normal with Unkonwn Mean and Variance

- Suppose we observe
$$y_1, \ldots, y_n \sim \mathcal{N}(\mu, \sigma^2), \quad i.i.d.$$
with unknown μ and σ^2.

Normal with Unkonwn Mean and Variance

- Suppose we observe
$$y_1, \ldots, y_n \sim \mathcal{N}(\mu, \sigma^2), \quad i.i.d.$$
with unknown μ and σ^2.
- The observation model is
$$p(y_1, \ldots, y_n \mid \mu, \sigma^2) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma^2}} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\} = (\sigma^2)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2}\right\}$$

Normal with Unkonwn Mean and Variance

- Suppose we observe
$$y_1, \ldots, y_n \sim \mathcal{N}(\mu, \sigma^2), \quad i.i.d.$$
with unknown μ and σ^2.
- The observation model is
$$p(y_1, \ldots, y_n \mid \mu, \sigma^2) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma^2}} \exp\left\{-\frac{(y_i - \mu)^2}{2\sigma^2}\right\} = (\sigma^2)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2\sigma^2}\right\}$$
- Notice that
$$\sum_{i=1}^{n}(y_i - \mu)^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2$$

Normal with Unkonwn Mean and Variance

- Suppose we observe

$$y_1, \dots, y_n \sim \mathcal{N}(\mu, \sigma^2), \quad i.i.d.$$

  with unknown μ and σ^2.

- The observation model is

$$p(y_1, \dots, y_n \mid \mu, \sigma^2) \propto \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma^2}} \exp\left\{-\frac{(y_i - \mu)^2}{2\sigma^2}\right\} = (\sigma^2)^{-n/2} \exp\left\{-\frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2\sigma^2}\right\}$$

- Notice that

$$\sum_{i=1}^{n}(y_i - \mu)^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2$$

- Therefore, we write (with $s^2 = (n-1)^{-1}\sum_i (y_i - \bar{y})^2$ the sample variance)

$$p(y_1, \dots, y_n \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y} - \mu)^2}{2\sigma^2}\right\}$$

Normal with Unkonwn Mean and Variance

- The score function is

$$\nabla \ell(\mu, \sigma^2) = \begin{pmatrix} -\frac{n(\mu-\bar{y})}{\sigma^2} \\ \frac{(n-1)s^2+n(\bar{y}-\mu)^2}{2(\sigma^2)^2} - \frac{n}{2\sigma^2} \end{pmatrix}$$

- The Fisher's information (2×2 **matrix**) is

$$\mathcal{I}(\mu, \sigma^2) = -\mathbb{E}_{\mu,\sigma^2}[\Delta \ell(\mu, \sigma)] = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

- The estimations of μ and of σ^2 are independent.

Normal — Uninformative Prior

- Attemp 1:
  - Since estimating μ and σ^2 are independent, recall the Uninformative prior:

$$\text{Normal with Known Variance :} p(\mu) \propto 1$$
$$\text{Normal with Known Mean :} p(\sigma^2) \propto 1/\sigma^2$$

  - By independence, we construct the following joint prior:

$$p(\mu, \sigma^2) \propto 1/\sigma^2$$

  - The above prior is uniform in $(\mu, \log \sigma^2)$.

Normal — Uninformative Prior

- Attemp 1:
  - Since estimating μ and σ^2 are independent, recall the Uninformative prior:

$$\text{Normal with Known Variance}: p(\mu) \propto 1$$
$$\text{Normal with Known Mean}: p(\sigma^2) \propto 1/\sigma^2$$

  - By independence, we construct the following joint prior:

$$p(\mu, \sigma^2) \propto 1/\sigma^2$$

  - The above prior is uniform in $(\mu, \log \sigma^2)$.
- Attemp 2:
  - With Jeffreys prior, we define the prior using the Fisher's information by

$$p(\mu, \sigma^2) \propto \sqrt{|\mathcal{I}|} \propto 1/\sigma^3$$

  - The prior is uniform in $(\mu/\sigma, \log \sigma^2)$.

Normal — Uninformative Prior

- Attemp 1:
  - Since estimating μ and σ^2 are independent, recall the Uninformative prior:

$$\text{Normal with Known Variance}: p(\mu) \propto 1$$
$$\text{Normal with Known Mean}: p(\sigma^2) \propto 1/\sigma^2$$

  - By independence, we construct the following joint prior:

$$p(\mu, \sigma^2) \propto 1/\sigma^2$$

  - The above prior is uniform in $(\mu, \log \sigma^2)$.
- Attemp 2:
  - With Jeffreys prior, we define the prior using the Fisher's information by

$$p(\mu, \sigma^2) \propto \sqrt{|\mathcal{I}|} \propto 1/\sigma^3$$

  - The prior is uniform in $(\mu/\sigma, \log \sigma^2)$.
- **Only the second one is uninformative.**

Uninformative Prior

- Jeffreys prior for multiparameter case:

$$p(\theta_1, \ldots, \theta_k) \propto \sqrt{|\mathcal{I}(\theta_1, \ldots, \theta_k)|}$$

Uninformative Prior

- Jeffreys prior for multiparameter case:

$$p(\theta_1, \ldots, \theta_k) \propto \sqrt{|\mathcal{I}(\theta_1, \ldots, \theta_k)|}$$

- Reasoning:
  - We assign uniform prior $p(\theta) \propto 1$ for the case that

$$\mathcal{I}(\theta) \propto \boldsymbol{I}$$

  - For any bijective continous mapping $\lambda = g(\theta)$, we have

$$\mathcal{I}(\lambda) = \left(\frac{\partial\theta}{\partial\lambda}\right)^T \mathcal{I}(\theta) \left(\frac{\partial\theta}{\partial\lambda}\right)$$

  - This corresponds to the change-of-variable of $p(\theta)$ to λ:

$$p(\lambda) = p(\theta)\left|\frac{\partial\theta}{\partial\lambda}\right| \propto \sqrt{|\mathcal{I}(\lambda)|}$$

Normal — Uninformative Prior

- Recall the observation model:

$$p(y_1, \ldots, y_n \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

- Now we choose the Jeffreys prior as $p(\mu, \sigma^2) \propto 1/\sigma^3$.

Normal — Uninformative Prior

- Recall the observation model:

$$p(y_1, \ldots, y_n \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

- Now we choose the Jeffreys prior as $p(\mu, \sigma^2) \propto 1/\sigma^3$.
- The joint posterior is

$$p(\mu, \sigma^2 \mid y_1, \ldots, y_n) \propto (\sigma^2)^{-(n+3)/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

Normal — Uninformative Prior

- Recall the observation model:

$$p(y_1, \ldots, y_n \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

- Now we choose the Jeffreys prior as $p(\mu, \sigma^2) \propto 1/\sigma^3$.
- The joint posterior is

$$p(\mu, \sigma^2 \mid y_1, \ldots, y_n) \propto (\sigma^2)^{-(n+3)/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

- The conditional posterior for μ is

$$p(\mu \mid \sigma^2, y_1, \ldots, y_n) \sim \mathcal{N}(\bar{y}, \sigma^2/n)$$

Normal — Uninformative Prior

- Recall the observation model:

$$p(y_1, \ldots, y_n \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

- Now we choose the Jeffreys prior as $p(\mu, \sigma^2) \propto 1/\sigma^3$.
- The joint posterior is

$$p(\mu, \sigma^2 \mid y_1, \ldots, y_n) \propto (\sigma^2)^{-(n+3)/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

- The conditional posterior for μ is

$$p(\mu \mid \sigma^2, y_1, \ldots, y_n) \sim \mathcal{N}(\bar{y}, \sigma^2/n)$$

- The conditional posterior for σ^2 is

$$p(\sigma^2 \mid \mu, y_1, \ldots, y_n) \sim \text{Inv-Gamma}((n+1)/2, [(n-1)s^2 + n(\bar{y}-\mu)^2]/2)$$

Normal — Uninformative Prior

- The marginal posterior for σ^2:

$$p(\sigma^2 \mid y_1, \ldots, y_n) \propto \int p(\mu, \sigma^2 \mid y_1, \ldots, y_n) d\mu \propto (\sigma^2)^{-(n+2)/2} \exp\left\{ -\frac{(n-1)s^2}{2\sigma^2} \right\}$$

Normal — Uninformative Prior

- The marginal posterior for σ^2:

$$p(\sigma^2 \mid y_1, \ldots, y_n) \propto \int p(\mu, \sigma^2 \mid y_1, \ldots, y_n) d\mu \propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}$$

- Or we can take

$$p(\sigma^2 \mid y_1, \ldots, y_n) \propto \frac{p(\mu, \sigma^2 \mid y_1, \ldots, y_n)}{p(\mu \mid \sigma^2 y_1, \ldots, y_n)} \propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}$$

Normal — Uninformative Prior

- The marginal posterior for σ^2:

$$p(\sigma^2 \mid y_1, \ldots, y_n) \propto \int p(\mu, \sigma^2 \mid y_1, \ldots, y_n) d\mu \propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}$$

- Or we can take

$$p(\sigma^2 \mid y_1, \ldots, y_n) \propto \frac{p(\mu, \sigma^2 \mid y_1, \ldots, y_n)}{p(\mu \mid \sigma^2 y_1, \ldots, y_n)} \propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}$$

- Therefore,
$p(\sigma^2 \mid y_1, \ldots, y_n) \sim \text{InvGamma}(n/2, (n-1)s^2/2) \sim \text{Scaled-Inv-}\chi^2(n, s^2)$

Normal — Uninformative Prior

- The marginal posterior for σ^2:

$$p(\sigma^2 \mid y_1, \ldots, y_n) \propto \int p(\mu, \sigma^2 \mid y_1, \ldots, y_n) d\mu \propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}$$

- Or we can take

$$p(\sigma^2 \mid y_1, \ldots, y_n) \propto \frac{p(\mu, \sigma^2 \mid y_1, \ldots, y_n)}{p(\mu \mid \sigma^2 y_1, \ldots, y_n)} \propto (\sigma^2)^{-(n+2)/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}$$

- Therefore,
  $p(\sigma^2 \mid y_1, \ldots, y_n) \sim \text{InvGamma}(n/2, (n-1)s^2/2) \sim \text{Scaled-Inv-}\chi^2(n, s^2)$
- The densities:

$$\text{InvGamma}(\alpha, \beta) \propto x^{-\alpha-1} e^{-\beta/x}, \quad \text{Scaled-Inv-}\chi^2(\nu, \tau^2) \propto x^{-\nu/2-1} e^{-\nu\tau^2/(2x)}$$

Normal — Uninformative Prior

- The marginal posterior for μ is:

$$p(\mu \mid y_1, \ldots, y_n) \propto \frac{p(\mu, \sigma^2 \mid y_1, \ldots, y_n)}{p(\sigma^2 \mid \mu^2, y_1, \ldots, y_n)}$$
$$\propto \left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]^{-(n+1)/2}$$
$$\propto \left[1 + \frac{n(\bar{y} - \mu)^2}{(n-1)s^2}\right]^{-(n+1)/2}$$

Normal — Uninformative Prior

- The marginal posterior for μ is:

$$
\begin{aligned}
p(\mu \mid y_1, \ldots, y_n) &\propto \frac{p(\mu, \sigma^2 \mid y_1, \ldots, y_n)}{p(\sigma^2 \mid \mu^2, y_1, \ldots, y_n)} \\
&\propto \left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]^{-(n+1)/2} \\
&\propto \left[1 + \frac{n(\bar{y} - \mu)^2}{(n-1)s^2}\right]^{-(n+1)/2}
\end{aligned}
$$

- It follows a noncentral scaled t distribution $t_n(\bar{y}, (n-1)s^2/n^2)$.

Normal — Uninformative Prior

- The marginal posterior for μ is:

$$\begin{aligned} p(\mu \mid y_1, \dots, y_n) &\propto \frac{p(\mu, \sigma^2 \mid y_1, \dots, y_n)}{p(\sigma^2 \mid \mu^2, y_1, \dots, y_n)} \\ &\propto \left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]^{-(n+1)/2} \\ &\propto \left[1 + \frac{n(\bar{y} - \mu)^2}{(n-1)s^2}\right]^{-(n+1)/2} \end{aligned}$$

- It follows a noncentral scaled t distribution $t_n(\bar{y}, (n-1)s^2/n^2)$.
- The kernel:

$$t_\nu(\mu, \tau^2) \propto \left[1 + \frac{(x-\mu)^2}{\nu\tau^2}\right]^{-(\nu+1)/2}$$

Normal — Conjugate Prior

- Recall the observation model:

$$p(y_1, \ldots, y_n \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

Normal — Conjugate Prior

- Recall the observation model:

$$p(y_1, \dots, y_n \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{-\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right\}$$

- We need some prior is the following form:

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-\alpha} \exp\left\{-\frac{\beta + \gamma(\mu-\delta)^2}{2\sigma^2}\right\}$$

for some hyperparameters $(\alpha, \beta, \gamma, \delta)$.

Normal — Conjugate Prior

- Recall the observation model:

$$p(y_1, \ldots, y_n \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left\{ -\frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2\sigma^2} \right\}$$

- We need some prior is the following form:

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-\alpha} \exp\left\{ -\frac{\beta + \gamma(\mu-\delta)^2}{2\sigma^2} \right\}$$

  for some hyperparameters $(\alpha, \beta, \gamma, \delta)$.
- We observe:
  - $\mu \mid \sigma^2 \sim \mathcal{N}(\delta, \sigma^2/\gamma)$
  - $\sigma^2 \mid \mu \sim \text{InvGamma}(\alpha - 1, (\beta + \gamma(\mu-\delta)^2)/2)$
  - $\sigma^2 \sim \text{InvGamma}(\alpha - 3/2, \beta/2)$
  - $\mu \sim t_{2\alpha-3}(\delta, \beta/(\gamma(2\alpha-3)))$

Normal — Conjugate Prior

- We found the following combination most convenient:

$$\sigma^2 \sim \text{InvGamma}, \quad \mu \mid \sigma^2 \sim \text{Normal}$$

Normal — Conjugate Prior

- We found the following combination most convenient:

$$\sigma^2 \sim \text{InvGamma}, \quad \mu \mid \sigma^2 \sim \text{Normal}$$

- With a bit change of notation, we define the prior as

$$\sigma^2 \sim \text{InvGamma}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right) \sim \text{Scaled-Inv-}\chi^2(\nu_0, \sigma_0^2), \quad \mu \mid \sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)$$

Normal — Conjugate Prior

- We found the following combination most convenient:

$$\sigma^2 \sim \text{InvGamma}, \quad \mu \mid \sigma^2 \sim \text{Normal}$$

- With a bit change of notation, we define the prior as

$$\sigma^2 \sim \text{InvGamma}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right) \sim \text{Scaled-Inv-}\chi^2(\nu_0, \sigma_0^2), \quad \mu \mid \sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)$$

- This prior is called **Normal-Inverse-Gamma** distribution or **Normal-Inverse-χ^2** distribution with density:

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-(\nu_0+3)/2} \exp\left\{-\frac{\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2}{2\sigma^2}\right\}$$

- N-Inv-Gamma $\left(\mu_0, \kappa_0, \frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right)$ or N-Inv-χ^2 $\left(\mu_0, \kappa_0, \nu_0, \sigma_0^2\right)$

Normal — Conjugate Prior

- We found the following combination most convenient:

$$\sigma^2 \sim \text{InvGamma}, \quad \mu \mid \sigma^2 \sim \text{Normal}$$

- With a bit change of notation, we define the prior as

$$\sigma^2 \sim \text{InvGamma}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right) \sim \text{Scaled-Inv-}\chi^2(\nu_0, \sigma_0^2), \quad \mu \mid \sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)$$

- This prior is called **Normal-Inverse-Gamma** distribution or **Normal-Inverse-**χ^2 distribution with density:

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-(\nu_0+3)/2} \exp\left\{-\frac{\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2}{2\sigma^2}\right\}$$

- N-Inv-Gamma $\left(\mu_0, \kappa_0, \frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right)$ or N-Inv-χ^2 $(\mu_0, \kappa_0, \nu_0, \sigma_0^2)$
- The Jeffreys prior corresponds to $\mu_0 = 0 = \kappa = 0 = \nu = 0 = \sigma_0 = 0$

Normal — Conjugate Prior

The posterior is

$$
\begin{aligned}
&p(\mu, \sigma^2 \mid y) \\
&\propto (\sigma^2)^{-(\nu_0+n+3)/2} \exp\left\{ -\frac{\nu_0\sigma_0^2 + (n-1)s^2 + \kappa_0(\mu-\mu_0)^2 + n(\mu-\bar{y})^2}{2\sigma^2} \right\} \\
&\propto (\sigma^2)^{-(\nu_0+n+3)/2} \exp\left\{ -\frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0-\bar{y})^2 + (\kappa_0+n)\left(\mu - \frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\right)^2}{2\sigma^2} \right\}
\end{aligned}
$$

which is N-Inv-Gamma $\left(\mu_0, \kappa_0, \frac{\nu_0}{2}, \frac{\nu_0\sigma^2}{2}\right)$ with

$$
\begin{aligned}
\mu_n &= \frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0-\bar{y})^2
\end{aligned}
$$

Normal — Conjugate Prior

$$p(\mu, \sigma^2 \mid y) \sim \text{N-Inv-}\Gamma\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \kappa_0 + n, \frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0 - \bar{y})^2}{2}\right)$$

- Now recall our previous discussion on the marginal/conditional distributions.

Normal — Conjugate Prior

$$p(\mu, \sigma^2 \mid y) \sim \text{N-Inv-}\Gamma\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \kappa_0 + n, \frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0 - \bar{y})^2}{2}\right)$$

- Now recall our previous discussion on the marginal/conditional distributions.
- conditional posterior of μ:

$$p(\mu \mid \sigma^2, y) \sim \mathcal{N}\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \frac{\sigma^2}{\kappa_0 + n}\right)$$

Normal — Conjugate Prior

$$p(\mu, \sigma^2 \mid y) \sim \text{N-Inv-}\Gamma\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \kappa_0 + n, \frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0 - \bar{y})^2}{2}\right)$$

- Now recall our previous discussion on the marginal/conditional distributions.
- conditional posterior of μ:

$$p(\mu \mid \sigma^2, y) \sim \mathcal{N}\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \frac{\sigma^2}{\kappa_0 + n}\right)$$

- conditional posterior of σ^2:

$$p(\sigma^2 \mid \mu, y) \sim \text{InvGamma}\left(\frac{\nu_0 + n + 1}{2}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \kappa_0(\mu - \mu_0)^2 + n(\mu - \bar{y})^2}{2}\right)$$

Normal — Conjugate Prior

$$p(\mu, \sigma^2 \mid y) \sim \text{N-Inv-}\Gamma\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \kappa_0 + n, \frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0 - \bar{y})^2}{2}\right)$$

- marginal posterior of σ^2:

$$p(\sigma^2 \mid y) \sim \text{InvGamma}\left(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0 - \bar{y})^2}{2}\right)$$

Normal — Conjugate Prior

$$p(\mu, \sigma^2 \mid y) \sim \text{N-Inv-}\Gamma\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \kappa_0 + n, \frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0 - \bar{y})^2}{2}\right)$$

- marginal posterior of σ^2:

$$p(\sigma^2 \mid y) \sim \text{InvGamma}\left(\frac{\nu_0 + n}{2}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0 - \bar{y})^2}{2}\right)$$

- marginal posterior of μ:

$$p(\mu \mid y) \sim t_{\nu_0+n}\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{n+\kappa_0}(\mu_0 - \bar{y})^2}{(\nu_0 + n)(\kappa_0 + n)}\right)$$