Normal model cont'd

Dr. Olanrewaju Michael Akande

Jan 31, 2020

Announcements

- Take "Participation Quiz II - Jan 31" on Sakai.
- Homework 3 now online.

Outline

- Inference for mean, conditional on variance (cont'd)
- Noninformative and improper priors
- Joint inference for mean and variance
- Back to the examples

The univariate normal model (cont'd)

Conditional inference for the mean (recap)

Normal data: $Y = (y_1, y_2, \ldots, y_n)$, where each

$$y_i \sim \mathcal{N}(\mu, \sigma^2); \quad \text{or} \quad y_i \sim \mathcal{N}(\mu, \tau^{-1}).$$

$+$ Normal Prior (when σ^2/ τ is known):

$$\mu|\sigma^2 \sim \mathcal{N}(\mu_0, \sigma_0^2); \quad \text{or} \quad \mu|\tau \sim \mathcal{N}(\mu_0, \tau_0^{-1}).$$

$\Rightarrow$ Normal posterior (in terms of precision):

$$\mu|Y, \tau \sim \mathcal{N}(\mu_n, \tau_n^{-1}).$$

where

- $\mu_n = \dfrac{\tau n \bar{y} + \tau_0 \mu_0}{\tau n + \tau_0}$; and
- $\tau_n = \tau n + \tau_0$.

Posterior with precision terms: combining information

- Posterior mean is weighted sum of prior information plus data information:

$$\begin{aligned}\mu_n &= \frac{n\tau\bar{y} + \tau_0\mu_0}{\tau n + \tau_0} \\ &= \frac{\tau_0}{\tau_0 + \tau n}\mu_0 + \frac{n\tau}{\tau_0 + \tau n}\bar{y}\end{aligned}$$

- Relatively easy to set μ_0 if we have a "prior" guess of the mean. What about τ_0?

- If we think of the prior mean as being based on κ_0 prior observations from a similar population as Y, then we might set $\sigma_0^2 = \frac{\sigma^2}{\kappa_0}$, which implies $\tau_0 = \kappa_0\tau$.

- Then the posterior mean is given by

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y}.$$

Posterior with variance terms

- In terms of variances, we have

$$\mu | Y, \sigma^2 \sim \mathcal{N}(\mu_n, \sigma_n^2)$$

where

$$\mu_n = \frac{\frac{n}{\sigma^2}\bar{y} + \frac{1}{\sigma_0^2}\mu_0}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}} \quad \text{and} \quad \sigma_n^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}.$$

- Easy to see that we can re-express the posterior information as a sum of the prior information and the information from the data.

Posterior with variance terms

- If we once again set $\sigma_0^2 = \dfrac{\sigma^2}{\kappa_0}$, the posterior mean is still given by

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y}.$$

- By the way, setting $\sigma_0^2 = \dfrac{\sigma^2}{\kappa_0} \Rightarrow$ prior dependence between μ and σ^2, whereas an arbitrary σ_0^2, independent on σ^2, $\Rightarrow$ prior independence between μ and σ^2.

Noninformative and improper priors

- Generally, we must specify both μ_0 and τ_0 to do inference.
- When prior distributions have no population basis, that is, there is no justification of the prior as "prior data", prior distributions can be difficult to construct.
- To that end, there is often the desire to construct noninformative priors, with the rationale being *"to let the data speak for themselves"*.
- For example, we could instead assume a uniform prior on μ that is constant over the real line, i.e., $\pi(\mu) \propto 1 \Rightarrow$ all values on the real line are equally likely apriori.
- Clearly, this is not a valid pdf since it will not integrate to 1 over the real line. Such priors are known as improper priors.
- An improper prior can still be very useful, we just need to ensure it results in a proper posterior.

Jeffreys' prior

- Question: is there a prior pdf (for a given model) that would be universally accepted as a noninformative prior?

- Laplace proposed the uniform distribution. This proposal lacks invariance under monotone transformations of the parameter.

- For example, a uniform prior on the binomial proportion parameter θ is not the same as a uniform prior on the odds parameter $\phi = \dfrac{\theta}{1-\theta}$.

- A more acceptable approach was introduced by Jeffreys. For single parameter models, the Jeffreys' prior defines a noninformative prior density of a parameter θ as

$$\pi(\theta) \propto \sqrt{\mathcal{I}(\theta)}$$

where $\mathcal{I}(\theta)$ is the Fisher information for θ.

Jeffreys' prior

- The Fisher information gives a way to measure the amount of information a random variable Y carries about an unknown parameter θ of a distribution that describes Y.

- Formally, $\mathcal{I}(\theta)$ is defined as

$$\mathcal{I}(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(y;\theta)\right)^2 \Big| \theta\right] = \int_{\mathcal{Y}} \left(\frac{\partial}{\partial\theta}\log f(y;\theta)\right)^2 f(y;\theta)dy.$$

- Alternatively,

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial^2\theta}\log f(y;\theta)\Big|\theta\right] = -\int_{\mathcal{Y}} \left(\frac{\partial^2}{\partial^2\theta}\log f(y;\theta)\right) f(y;\theta)dy.$$

- Turns out that the Jeffreys' prior for μ under the normal model, when σ^2 is known, is

$$\pi(\mu) \propto 1,$$

the uniform prior over the real line. You should try to derive this.

Inference for mean, conditional on variance using Jeffreys' prior

- Recall that for σ^2 known, the normal likelihood simplifies to

$$\propto \ \exp\left\{-\frac{1}{2}\tau n(\mu - \bar{y})^2\right\},$$

ignoring everything else that does not depend on μ.

- With the Jeffreys' prior $\pi(\mu) \propto 1$, can we derive the posterior distribution?

Inference for mean, conditional on variance using Jeffreys' prior

- Posterior:

$$\begin{aligned}\pi(\mu|Y,\sigma^2) &\propto \exp\left\{-\frac{1}{2}\tau n(\mu-\bar{y})^2\right\}\pi(\mu)\\ &\propto \exp\left\{-\frac{1}{2}\tau n(\mu-\bar{y})^2\right\}.\end{aligned}$$

- This is the kernel of a normal distribution with
  - mean $\bar{y}$, and
  - precision $n\tau$ or variance $\frac{1}{n\tau}=\frac{\sigma^2}{n}$.
- Written differently, we have $\mu|Y,\sigma^2 \sim \mathcal{N}(\bar{y},\frac{\sigma^2}{n})$
- This should look familiar to you. Does it?

Joint inference for mean and variance

- What happens when $\sigma/\ \tau$ is unknown? We need a joint prior $\pi(\mu, \sigma^2)$ for μ and σ^2.

- Write the joint prior distribution for the mean and variance as the product of a conditional and a marginal distribution, so we can take advantage of our work so far.

- That is,

$$\pi(\mu, \sigma^2) \ = \ \pi(\mu|\sigma^2)\pi(\sigma^2).$$

- For $\pi(\sigma^2)$, we need a distribution with support on $(0, \infty)$. One such family is the gamma family, but this is NOT conjugate for the variance of a normal distribution.

- The gamma distribution is, however, conjugate for the precision τ, and in that case, we say that σ^2 has an inverse-gamma distribution.

Conditional specification of prior

- Once again, suppose $Y = (y_1, y_2, \ldots, y_n)$, where each

$$y_i \sim \mathcal{N}(\mu, \sigma^2); \quad \text{or} \quad y_i \sim \mathcal{N}(\mu, \tau^{-1}).$$

- A conjugate joint prior is given by

$$\tau = \frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$
$$\mu|\sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right) \quad \text{or} \quad \mu|\tau \sim \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0 \tau}\right).$$

- This is often called a normal-gamma prior distribution.
- σ_0^2 is the prior guess for σ^2, while ν_0 is often referred to as the "prior degrees of freedom", our degree of confidence in σ_0^2.
- We do not have conjugacy if we replace $\frac{\sigma^2}{\kappa_0}$ in the normal prior with an arbitrary prior variance independent of σ^2. To do inference in that scenario, we need Gibbs sampling (to come next week!).

Posterior for the mean given variance, under normal-gamma prior

- Based on the normal-gamma prior, we need $\pi(\mu|Y, \sigma^2)$ and $\pi(\tau|Y)$.
- For $\pi(\mu|Y, \sigma^2)$, we can leverage our previous results. We have

$$\mu|Y, \sigma^2 \sim \mathcal{N}(\mu_n, \frac{\sigma^2}{\kappa_n}) \quad \text{or} \quad \mu|Y, \tau \sim \mathcal{N}(\mu_n, \frac{1}{\kappa_n \tau})$$

where

$$\mu_n = \frac{\frac{n}{\sigma^2}\bar{y} + \frac{\kappa_0}{\sigma^2}\mu_0}{\frac{n}{\sigma^2} + \frac{\kappa_0}{\sigma^2}} = \frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_n} = \frac{\kappa_0}{\kappa_n}\mu_0 + \frac{n}{\kappa_n}\bar{y} \quad \text{and} \quad \kappa_n = \kappa_0 + n.$$

- μ_n is simply the sample mean of the current and prior observations, and posterior variance of μ given σ^2 is σ^2 divided by the total number of observations (prior and current).

Posterior derivation

- Some algebra is required to get the marginal posterior of τ. Let's write the full joint posterior and go from there. We must keep some of the terms we discarded in the last lecture.

- Recall the likelihood

$$L(Y; \mu, \tau) \propto \tau^{\frac{n}{2}} \exp\left\{-\frac{1}{2}\tau s^2(n-1)\right\} \exp\left\{-\frac{1}{2}\tau n(\mu - \bar{y})^2\right\},$$

- Now, $\mu|\tau \sim \mathcal{N}\left(\mu_0, \frac{1}{\kappa_0 \tau}\right) \Rightarrow$

$$\pi(\mu|\tau) \propto \exp\left\{-\frac{1}{2}\kappa_0\tau(\mu - \mu_0)^2\right\}.$$

- and $\tau \sim \mathrm{Ga}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right) \Rightarrow$

$$\pi(\tau) \propto \tau^{\frac{\nu_0}{2}-1} \exp\left\{-\frac{\tau\nu_0\sigma_0^2}{2}\right\}.$$

Posterior derivation

$$
\begin{aligned}
\Rightarrow \pi(\mu,\tau|Y) \;\propto\; & \pi(\mu|\sigma^2)\times\pi(\tau)\times L(Y;\mu,\sigma^2) \\
\propto\; & \underbrace{\exp\left\{-\frac{1}{2}\kappa_0\tau(\mu-\mu_0)^2\right\}}_{\propto\,\pi(\mu|\sigma^2)} \times \underbrace{\tau^{\frac{\nu_0}{2}-1}\exp\left\{-\frac{\tau\nu_0\sigma_0^2}{2}\right\}}_{\propto\,\pi(\tau)} \\
& \times \underbrace{\tau^{\frac{n}{2}}\exp\left\{-\frac{1}{2}\tau s^2(n-1)\right\}\exp\left\{-\frac{1}{2}\tau n(\mu-\bar{y})^2\right\}}_{\propto\,L(Y;\mu,\sigma^2)} \\
=\; & \underbrace{\exp\left\{-\frac{1}{2}\kappa_0\tau(\mu-\mu_0)^2\right\}\exp\left\{-\frac{1}{2}\tau n(\mu-\bar{y})^2\right\}}_{\text{Terms involving }\mu} \\
& \times \underbrace{\tau^{\frac{\nu_0}{2}-1}\exp\left\{-\frac{\tau\nu_0\sigma_0^2}{2}\right\}\tau^{\frac{n}{2}}\exp\left\{-\frac{1}{2}\tau s^2(n-1)\right\}}_{\text{Terms NOT involving }\mu} \\
=\; & \exp\left\{-\frac{1}{2}\kappa_0\tau(\mu^2-2\mu\mu_0+\mu_0^2)\right\}\exp\left\{-\frac{1}{2}\tau n(\mu^2-2\mu\bar{y}+\bar{y}^2)\right\} \\
& \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}
\end{aligned}
$$

Posterior derivation

$$\begin{aligned}\pi(\mu,\tau|Y) &\propto \exp\left\{-\frac{1}{2}\left[\kappa_0\tau(\mu^2-2\mu\mu_0)+\tau n(\mu^2-2\mu\bar{y})\right]\right\}\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2\right]\right\}\\ &\qquad\times\tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}\\ &= \exp\left\{-\frac{1}{2}\left[\mu^2(n\tau+\kappa_0\tau)-2\mu(n\tau\bar{y}+\kappa_0\tau\mu_0)\right]\right\}\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2\right]\right\}\\ &\qquad\times\tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}\end{aligned}$$

- Set $a^\star=(n\tau+\kappa_0\tau)$ and $b^\star=(n\tau\bar{y}+\kappa_0\tau\mu_0)$, then complete the square for the first part like we did in the last lecture.

$$\begin{aligned}\Rightarrow\pi(\mu,\tau|Y) &\propto \exp\left\{-\frac{1}{2}\left[\mu^2a^\star-2\mu b^\star\right]\right\}\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2\right]\right\}\\ &\qquad\times\tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}\end{aligned}$$

Posterior derivation

$$
\begin{aligned}
\Rightarrow \pi(\mu,\tau|Y) \;\propto\;& \exp\left\{-\frac{1}{2}a^{\star}\left[\mu-\frac{b^{\star}}{a^{\star}}\right]^2+\frac{(b^{\star})^2}{2a^{\star}}\right\}\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2\right]\right\}\\
&\times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}\\
=\;& \exp\left\{-\frac{1}{2}a^{\star}\left[\mu-\frac{b^{\star}}{a^{\star}}\right]^2\right\}\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2-\frac{(b^{\star})^2}{a^{\star}}\right]\right\}\\
&\times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}\\
=\;& \exp\left\{-\frac{1}{2}a^{\star}\left[\mu-\frac{b^{\star}}{a^{\star}}\right]^2\right\}\underbrace{\exp\left\{-\frac{1}{2}\left[\kappa_0\tau\mu_0^2+\tau n\bar{y}^2-\frac{(n\tau\bar{y}+\kappa_0\tau\mu_0)^2}{(n\tau+\kappa_0\tau)}\right]\right\}}_{\text{Expand terms and recombine}}\\
&\times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}\\
=\;& \exp\left\{-\frac{1}{2}a^{\star}\left[\mu-\frac{b^{\star}}{a^{\star}}\right]^2\right\}\exp\left\{-\frac{1}{2}\left[\frac{n\kappa_0\tau^2(\mu_0^2-2\mu_0\bar{y}+\bar{y}^2)}{\tau(n+\kappa_0)}\right]\right\}\\
&\times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}
\end{aligned}
$$

Posterior derivation

$$
\begin{aligned}
\Rightarrow \pi(\mu,\tau|Y) &\propto \exp\left\{-\frac{1}{2}a^\star\left[\mu-\frac{b^\star}{a^\star}\right]^2\right\}\exp\left\{-\frac{\tau}{2}\left[\frac{n\kappa_0(\bar{y}-\mu_0)^2}{(n+\kappa_0)}\right]\right\} \\
&\qquad \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\} \\
&= \exp \underbrace{\left\{-\frac{1}{2}a^\star\left[\mu-\frac{b^\star}{a^\star}\right]^2\right\}}_{\text{Substitute the values for } a^\star \text{ and } b^\star \text{ back}} \\
&\qquad \times \tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau\left[\nu_0\sigma_0^2+s^2(n-1)\right]}{2}\right\}\exp\left\{-\frac{\tau}{2}\left[\frac{n\kappa_0(\bar{y}-\mu_0)^2}{(n+\kappa_0)}\right]\right\} \\
&= \underbrace{\exp\left\{-\frac{1}{2}(n\tau+\kappa_0\tau)\left[\mu^2-\frac{(n\tau\bar{y}+\kappa_0\tau\mu_0)}{(n\tau+\kappa_0\tau)}\right]^2\right\}}_{\text{Normal Kernel}} \\
&\qquad \times \underbrace{\tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau}{2}\left[\nu_0\sigma_0^2+s^2(n-1)+\frac{n\kappa_0}{(n+\kappa_0)}(\bar{y}-\mu_0)^2\right]\right\}}_{\text{Gamma Kernel}}
\end{aligned}
$$

Posterior derivation

$$
\begin{aligned}
\Rightarrow \pi(\mu,\tau|Y) \propto\ & \underbrace{\exp\left\{-\frac{1}{2}\tau(n+\kappa_0)\left[\mu^2-\frac{(n\bar{y}+\kappa_0\mu_0)}{(n+\kappa_0)}\right]^2\right\}}_{\text{Normal Kernel}} \\
& \times \underbrace{\tau^{\frac{\nu_0+n}{2}-1}\exp\left\{-\frac{\tau}{2}\left[\nu_0\sigma_0^2+s^2(n-1)+\frac{n\kappa_0}{(n+\kappa_0)}(\bar{y}-\mu_0)^2\right]\right\}}_{\text{Gamma Kernel}} \\
= \ & \mathcal{N}\left(\mu_n, \frac{1}{\kappa_n\tau}\right)\times \text{Gamma}\left(\frac{\nu_n}{2},\frac{\nu_n\sigma_n^2}{2}\right) = \pi(\mu|Y,\tau)\pi(\tau|Y),
\end{aligned}
$$

where

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

- Turns out that the marginal posterior of μ, that is, $\pi(\mu|Y) = \int_0^\infty \pi(\mu,\tau|Y)d\tau$ is a **t-distribution**. You can derive that distribution if you are interested, we won't spend time on it in class.

Back to our examples

- Pygmalion: questions of interest
  - Is the average improvement for the accelerated group larger than that for the no growth group?
    - What is $\Pr[\mu_A > \mu_N | Y_A, Y_N)$?
  - Is the variance of improvement scores for the accelerated group larger than that for the no growth group?
    - What is $\Pr[\sigma_A^2 > \sigma_N^2 | Y_A, Y_N)$?
- Job training: questions of interest
  - Is the average change in annual earnings for the training group larger than that for the no training group?
    - What is $\Pr[\mu_T > \mu_N | Y_T, Y_N)$?
  - Is the variance of change in annual earnings for the training group larger than that for the no training group?
    - What is $\Pr[\sigma_T^2 > \sigma_N^2 | Y_T, Y_N)$?

Mildly informative priors

- We will focus on the Pygmalion study. Follow the same approach for the job training data.
- Suppose you have no idea whether students would improve IQ on average. Set $\mu_{0A} = \mu_{0N} = 0$.
- Suppose you don't have any faith in this belief, and think it is the equivalent of having only 1 prior observation in each group. Set $\kappa_{0A} = \kappa_{0N} = 1$.
- Based on the literature, SD of change scores should be around 10 in each group, but still you don't have a lot of faith in this belief. Set $\nu_{0A} = \nu_{0N} = 1$ and $\sigma^2_{0A} = \sigma^2_{0N} = 100$.
- Graph priors to see if they accord with your beliefs. Sampling new values of Y from the priors offers a good check.

Recall the Pygmalion data

- Data:
  - Accelerated group (A): 20, 10, 19, 15, 9, 18.
  - No growth group (N): 3, 2, 6, 10, 11, 5.
- Summary statistics:
  - $\bar{y}_A = 15.2$; $s_A = 4.71$.
  - $\bar{y}_N = 6.2$; $s_N = 3.65$.

Analysis with mildly informative priors

$$
\begin{aligned}
\kappa_{nA} &= \kappa_{0A} + n_A = 1 + 6 = 7 \\
\kappa_{nN} &= \kappa_{0N} + n_N = 1 + 6 = 7 \\
\nu_{nA} &= \nu_{0A} + n_A = 1 + 6 = 7 \\
\nu_{nN} &= \nu_{0N} + n_N = 1 + 6 = 7 \\
\mu_{nA} &= \frac{\kappa_{0A}\mu_{0A} + n_A\bar{y}_A}{\kappa_{nA}} = \frac{(1)(0) + (6)(15.2)}{7} \approx 13.03 \\
\mu_{nN} &= \frac{\kappa_{0N}\mu_{0N} + n_N\bar{y}_N}{\kappa_{nN}} = \frac{(1)(0) + (6)(6.2)}{7} \approx 5.31 \\
\sigma^2_{nA} &= \frac{1}{\nu_{nA}}\left[\nu_{0A}\sigma^2_{0A} + s^2_A(n_A - 1) + \frac{n_A\kappa_{0A}}{\kappa_{nA}}(\bar{y}_A - \mu_{0A})^2\right] \\
&= \frac{1}{7}\left[(1)(100) + (22.17)(5) + \frac{(6)(1)}{(7)}(15.2 - 0)^2\right] \approx 58.41 \\
\sigma^2_{nN} &= \frac{1}{\nu_{nN}}\left[\nu_{0N}\sigma^2_{0N} + s^2_N(n_N - 1) + \frac{n_N\kappa_{0N}}{\kappa_{nN}}(\bar{y}_N - \mu_{0N})^2\right] \\
&= \frac{1}{7}\left[(1)(100) + (13.37)(5) + \frac{(6)(1)}{(7)}(6.2 - 0)^2\right] \approx 28.54
\end{aligned}
$$

Analysis with mildly informative priors

- So our joint posterior is

$$\mu_A|Y_A,\tau_A \sim \mathcal{N}\left(\mu_{nA},\frac{1}{\kappa_{nA}\tau_A}\right) = \mathcal{N}\left(13.03,\frac{1}{7\tau_A}\right)$$

$$\tau_A|Y_A \sim \text{Gamma}\left(\frac{\nu_{nA}}{2},\frac{\nu_{nA}\sigma^2_{nA}}{2}\right) = \text{Gamma}\left(\frac{7}{2},\frac{7(58.41)}{2}\right)$$

$$\mu_N|Y_N,\tau_N \sim \mathcal{N}\left(\mu_{nN},\frac{1}{\kappa_{nN}\tau_N}\right) = \mathcal{N}\left(5.31,\frac{1}{7\tau_N}\right)$$

$$\tau_N|Y_N \sim \text{Gamma}\left(\frac{\nu_{nN}}{2},\frac{\nu_{nN}\sigma^2_{nN}}{2}\right) = \text{Gamma}\left(\frac{7}{2},\frac{7(28.54)}{2}\right)$$

Monte Carlo sampling

- To evaluate whether the accelerated group has larger IQ gains than the normal group, we would like to estimate quantities like $\Pr[\mu_A > \mu_N | Y_A, Y_N)$ which are based on the **marginal posterior** of μ rather than the **conditional distribution**.
- Fortunately, this is easy to do by generating samples of μ and σ^2 from their joint posterior.

Monte Carlo sampling

- Suppose we simulate values using the following Monte Carlo procedure:

$$
\begin{aligned}
\tau^{(1)} &\sim \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}\right) \\
\mu^{(1)} &\sim \mathcal{N}\left(\mu_n, \frac{1}{\kappa_n \tau^{(1)}}\right) \\
\tau^{(2)} &\sim \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}\right) \\
\mu^{(2)} &\sim \mathcal{N}\left(\mu_n, \frac{1}{\kappa_n \tau^{(2)}}\right) \\
&\vdots \\
&\vdots \\
&\vdots \\
\tau^{(m)} &\sim \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}\right) \\
\mu^{(m)} &\sim \mathcal{N}\left(\mu_n, \frac{1}{\kappa_n \tau^{(m)}}\right)
\end{aligned}
$$

Monte Carlo sampling

- Note that we are sampling each $\mu^{(j)}$, $j = 1, \ldots, m$, from its conditional distribution, not from the marginal.
- The sequence of pairs $\{(\tau, \mu)^{(1)}, \ldots, (\tau, \mu)^{(m)}\}$ simulated using this method are independent samples from the joint posterior $\pi(\mu, \tau | Y)$.
- Additionally, the simulated sequence $\{\mu^{(1)}, \ldots, \mu^{(m)}\}$ are independent samples from the **marginal posterior distribution**.
- While this may seem odd, keep in mind that while we drew the μ's as conditional samples, each was conditional on a different value of τ.
- Thus, together they constitute marginal samples of μ.

Monte Carlo sampling

It is easy to sample from these posteriors:

```
aA <- 7/2
aN <- 7/2
bA <- (7/2)*58.41
bN <- (7/2)*28.54
muA <- 13.03
muN <- 5.31
kappaA <- 7
kappaN <- 7
tauA_postsample <- rgamma(10000,aA,bA)
thetaA_postsample <- rnorm(10000,muA,sqrt(1/(kappaA*tauA_postsample)))
tauN_postsample <- rgamma(10000,aN,bN)
thetaN_postsample <- rnorm(10000,muN,sqrt(1/(kappaN*tauN_postsample)))
sigma2A_postsample <- 1/tauA_postsample
sigma2N_postsample <- 1/tauN_postsample
```

Monte Carlo sampling

- Is the average improvement for the accelerated group larger than that for the no growth group?
    - What is $\Pr[\mu_A > \mu_N | Y_A, Y_N)$?

```
mean(thetaA_postsample > thetaN_postsample)
```

```
## [1] 0.9721
```

- Is the variance of improvement scores for the accelerated group larger than that for the no growth group?
    - What is $\Pr[\sigma^2_A > \sigma^2_N | Y_A, Y_N)$?

```
mean(sigma2A_postsample > sigma2N_postsample)
```

```
## [1] 0.8185
```

- What can we conclude from this?

Improper prior

- Let's be very objective with the prior selection. In fact, let's be extreme!
  - If we let the normal variance $\to \infty$ then our prior on μ is $\propto 1$ (recall the Jeffreys' prior on μ for known σ^2).
  - If we let the gamma variance get very large (e.g., $a, b \to 0$), then the prior on σ^2 is $\propto \dfrac{1}{\sigma^2}$.
- $\pi(\mu, \sigma^2) \propto \dfrac{1}{\sigma^2}$ is improper (does not integrate to 1) but does lead to a proper posterior distribution that yields inferences similar to frequentist ones.
- For that choice, we have

$$
\begin{aligned}
\mu | Y, \tau &\sim \mathcal{N}\left(\bar{y}, \frac{1}{n\tau}\right) \\
\tau | Y &\sim \text{Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)
\end{aligned}
$$

Analysis with noninformative priors

- So our joint posterior is

$$
\begin{aligned}
\mu_A|Y_A,\tau_A &\sim \mathcal{N}\left(\bar{y}_A, \frac{1}{n_A\tau_A}\right) = \mathcal{N}\left(15.2, \frac{1}{6\tau_A}\right)\\
\tau_A|Y_A &\sim \text{Gamma}\left(\frac{n_A-1}{2}, \frac{(n_A-1)s_A^2}{2}\right) = \text{Gamma}\left(\frac{6-1}{2}, \frac{(6-1)(22.17)}{2}\right)\\
\mu_N|Y_N,\tau_N &\sim \mathcal{N}\left(\bar{y}_N, \frac{1}{n_N\tau_N}\right) = \mathcal{N}\left(6.2, \frac{1}{6\tau_N}\right)\\
\tau_N|Y_N &\sim \text{Gamma}\left(\frac{n_N-1}{2}, \frac{(n_N-1)s_A^2}{2}\right) = \text{Gamma}\left(\frac{6-1}{2}, \frac{(6-1)(13.37)}{2}\right)
\end{aligned}
$$

Monte Carlo sampling

It is easy to sample from these posteriors:

```
aA <- (6-1)/2
aN <- (6-1)/2
bA <- (6-1)*22.17/2
bN <- (6-1)*13.37/2
muA <- 15.2
muN <- 6.2
tauA_postsample_impr <- rgamma(10000,aA,bA)
thetaA_postsample_impr <- rnorm(10000,muA,sqrt(1/(6*tauA_postsample_impr)))
tauN_postsample_impr <- rgamma(10000,aN,bN)
thetaN_postsample_impr <- rnorm(10000,muN,sqrt(1/(6*tauN_postsample_impr)))
sigma2A_postsample_impr <- 1/tauA_postsample_impr
sigma2N_postsample_impr <- 1/tauN_postsample_impr
```

Monte Carlo sampling

- Is the average improvement for the accelerated group larger than that for the no growth group?
  - What is $\Pr[\mu_A > \mu_N | Y_A, Y_N)$?

```
mean(thetaA_postsample_impr > thetaN_postsample_impr)
```

```
## [1] 0.9941
```

- Is the variance of improvement scores for the accelerated group larger than that for the no growth group?
  - What is $\Pr[\sigma_A^2 > \sigma_N^2 | Y_A, Y_N)$?

```
mean(sigma2A_postsample_impr > sigma2N_postsample_impr)
```

```
## [1] 0.7113
```

- How does the new choice of prior affect our conclusions?

Recall the job training data

- Data:
  - No training group (N): sample size $n_N = 429$.
  - Training group (T): sample size $n_A = 185$.
- Summary statistics for change in annual earnings:
  - $\bar{y}_N = 1364.93$; $s_N = 7460.05$
  - $\bar{y}_T = 4253.57$; $s_T = 8926.99$
- Using the same approach we used for the Pygmalion data, answer the questions of interest.