STAT 535: Chapter 5: More Conjugate Priors

David B. Hitchcock
E-Mail: hitchcock@stat.sc.edu

Spring 2024

Why are Conjugate Priors Nice?

- Recall that a **conjugate prior** is a prior which (along with the data model) produces a posterior distribution that has the same functional form as the prior (but with new, updated parameter values).
- In the Beta-binomial setup, the beta prior was conjugate because the posterior was also a beta distribution.
- Conjugate priors are nice because
  1. we can typically derive the posterior without needing any difficult computation;
  2. it is typically easy to understand the respective contributions of the prior information and the data information to the posterior.
- We will now examine a couple of other Bayesian models with conjugate priors.

The Poisson Distribution

- Recall that the Poisson distribution is a common model for **count data**: Data whose possible values are the nonnegative integers $0, 1, 2, \ldots$.
- The Poisson distribution is indexed by a parameter $\lambda > 0$, and (given λ) the pdf of a Poisson random variable $Y|\lambda$ is:

$$f(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}$$

- If our data consists of a random sample on n such counts, then the likelihood function is the joint density function $f(y_1|\lambda)f(y_2|\lambda)\cdots f(y_n|\lambda)$, since $Y_1, Y_2, \ldots, Y_n$ are independent.

Choice of Prior

- When our data model is Poisson, what is a good choice for the prior for the parameter λ?
- Since $\lambda > 0$, we should use as a prior some distribution whose support is $(0, \infty)$.
- The Gamma distribution is a good choice for the prior, since its support is $(0, \infty)$.
- Note that the parameterization of the Gamma distribution that we will use in this class is different from the one in the STAT 511 course.
- We will consider a Gamma pdf with a **shape** parameter s and a **rate** parameter r:

$$f(\lambda) = \frac{r^s}{\Gamma(s)} \lambda^{s-1} e^{-r\lambda}, \quad \lambda > 0.$$

- Note that the rate parameter is the **reciprocal** of the **scale** parameter used in the other parameterization.

The Gamma/Poisson Bayesian Model

- If our data $Y_1, \ldots, Y_n$ are iid Poisson(λ), then a gamma(s, r) prior on λ is a **conjugate** prior.
  Likelihood:

$$L(\lambda|\boldsymbol{y}) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} = \frac{e^{-n\lambda}\lambda^{\sum y_i}}{\prod_{i=1}^{n}(y_i!)}$$

  Prior:

$$f(\lambda) = \frac{r^s}{\Gamma(s)}\lambda^{s-1}e^{-r\lambda}, \quad \lambda > 0.$$

  $\Rightarrow$ Posterior:

$$f(\lambda|\boldsymbol{y}) \propto \lambda^{\sum y_i + s - 1}e^{-(n+r)\lambda}, \quad \lambda > 0.$$

  $\Rightarrow f(\lambda|\boldsymbol{y})$ is gamma$\left(\sum y_i + s, n + r\right)$. **(Conjugate!)**

Properties of the Gamma (Mean)

- Under this shape/rate parameterization, the mean of the Gamma(s, r) prior distribution is

$$E(\lambda) = \frac{s}{r}$$

- Based on our prior beliefs, we would choose appropriate values of the hyperparameters s and r.
- Similarly, the mean of the Gamma$\left(\sum y_i + s, n + r\right)$ posterior distribution is

$$E(\lambda|\boldsymbol{y}) = \frac{\sum y_i + s}{n + r}$$

- This posterior mean could be used as a Bayesian estimator of λ.

Properties of the Gamma (Variance)

- If we have a good guess of the prior mean of λ, how can we specifically choose which s and r to use in our prior?
- Under this shape/rate parameterization, the variance of the Gamma(s, r) prior distribution is

$$Var(\lambda) = \frac{s}{r^2}$$

- The prior variance (and standard deviation) can guide our choices of s and r.
- Plotting the potential prior using the `plot_gamma` function in the `bayesrules` package can also be helpful in choosing the prior.

The Posterior Mean in the Gamma/Poisson Bayesian Model

- The posterior mean is:

$$\hat{\lambda}_B = \frac{\sum y_i + s}{n + r}$$
$$= \frac{\sum y_i}{n + r} + \frac{s}{n + r}$$
$$= \left[\frac{n}{n + r}\right]\left(\frac{\sum y_i}{n}\right) + \left[\frac{r}{n + r}\right]\left(\frac{s}{r}\right)$$

- Again, the data get weighted more heavily as $n \to \infty$.

Example: Fraud Risk Phone Calls

- The textbook gives an example using data on fraud risk phone calls per day, which can be modeled with a Poisson distribution.
- The parameter of interest is λ, the mean number of fraud risk calls per day.
- Prior belief: The mean number of such calls per day is around 5.
- So let's choose s and r so that $s/r = 5$.
- Also, we believe that λ is very likely between 2 and 7.
- Let's try to plot a few possible priors that have $s/r = 5$ (see R examples).

Example: Fraud Risk Phone Calls

- The choice of $s = 10$ and $r = 2$ seems to reflect our prior beliefs.
- We collect $n = 4$ counts as our data, and the data were: 6, 2, 2, 1 ($\sum_i y_i = 11$ and $\bar{y} = 2.75$).
- So our posterior is $\text{Gamma}\left(\sum y_i + s, n + r\right) = \text{Gamma}(11 + 10, 4 + 2) = \text{Gamma}(21, 6)$
- A Bayesian estimate of λ is thus the posterior mean $21/6 = 3.5$.
- See R plots to see how the data has updated our prior beliefs.

Bayesian Inference: Posterior Intervals

- Simple values like the posterior mean $E[\boldsymbol{\theta}|\boldsymbol{y}]$ and posterior variance $var[\boldsymbol{\theta}|\boldsymbol{y}]$ can be useful in learning about $\boldsymbol{\theta}$.
- Quantiles of $p(\boldsymbol{\theta}|\boldsymbol{y})$ (especially the posterior median) can also be a useful summary of $\boldsymbol{\theta}$.
- The ideal summary of $\boldsymbol{\theta}$ is an interval (or region) with a certain probability of containing $\boldsymbol{\theta}$.
- Note that a classical (frequentist) **confidence interval** does not exactly have this interpretation.

Bayesian Credible Intervals

- A **credible interval** (or in general, a **credible set**) is the Bayesian analogue of a confidence interval.
- A $100(1-\alpha)\%$ credible set $\mathcal{C}$ is a subset of Θ such that

$$\int_{\mathcal{C}} p(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} = 1 - \alpha.$$

- If the parameter space Θ is discrete, a sum replaces the integral.

Quantile-Based Intervals

- If θ_L^* is the $\alpha/2$ posterior quantile for θ, and θ_U^* is the $1-\alpha/2$ posterior quantile for θ, then (θ_L^*, θ_U^*) is a $100(1-\alpha)\%$ credible interval for θ.

Note: $P[\theta < \theta_L^* | \boldsymbol{y}] = \alpha/2$ and $P[\theta > \theta_U^* | \boldsymbol{y}] = \alpha/2$.

$$
\begin{aligned}
&\Rightarrow P\{\theta \in (\theta_L^*, \theta_U^*) | \boldsymbol{y}\} \\
&= 1 - P\{\theta \notin (\theta_L^*, \theta_U^*) | \boldsymbol{y}\} \\
&= 1 - \left(P[\theta < \theta_L^* | \boldsymbol{y}] + P[\theta > \theta_U^* | \boldsymbol{y}] \right) \\
&= 1 - \alpha.
\end{aligned}
$$

Quantile-Based Intervals

Picture:

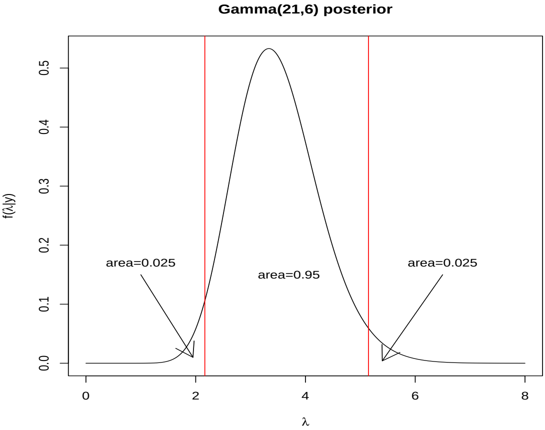


Figure: Between 2.17 and 5.15 is posterior probability 0.95.

Changing the Width of the Credible Interval

- The **credible interval** $(2.17, 5.15)$ in the picture on the previous slide is based on a Gamma$(21, 6)$ posterior distribution.
- The posterior probability that the true daily mean number of fraud risk calls is between 2.17 and 5.15 is 0.95.
- What could we do if we desired a narrower (more precise) credible interval?
- We could use, say, a 90% credible interval, with area 0.05 in each tail.
- See R code for example of deriving a 90% credible interval with this posterior distribution.
- The 90% credible interval is $(2.35, 4.84)$ here. We will soon see a different approach to getting a 90% credible interval that is even narrower.

Example 2: Quantile-Based Interval

- Consider 10 flips of a coin having $P\{\text{Heads}\} = \theta$.
- Suppose we observe 2 "heads".
- We model the count of heads as binomial:

$$p(y|\theta) = \binom{10}{y} \theta^y (1-\theta)^{10-y}, \;\; y = 0, 1, \ldots, 10.$$

- Let's use a uniform prior for θ:

$$p(\theta) = 1, \;\; 0 \leq \theta \leq 1.$$

Example 2: Quantile-Based Interval

- Then the posterior is:

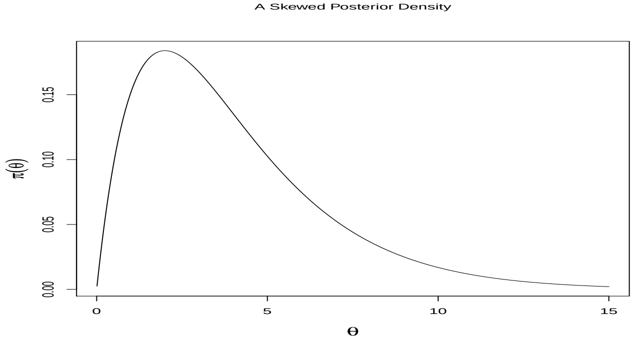
$$\begin{aligned} p(\theta|y) &\propto p(\theta)L(\theta|y) \\ &= (1)\binom{10}{y}\theta^y(1-\theta)^{10-y} \\ &\propto \theta^y(1-\theta)^{10-y}, \;\; 0 \le \theta \le 1. \end{aligned}$$

- This is a **beta** distribution for θ with parameters $y+1$ and $10-y+1$.
- Since $y = 2$ here, $p(\theta|y=2)$ is beta(3,9).
- The 0.025 and 0.975 quantiles of a beta(3,9) are (.0602, .5178), which is a 95% credible interval for θ.

HPD Intervals / Regions

- The equal-tail credible interval approach is ideal when the posterior distribution is symmetric.
- But what if $p(\theta|y)$ is skewed?

Picture:



HPD Intervals / Regions

- Note that values of θ around 1 have **much** higher posterior probability than values around 7.5.
- Yet 7.5 is in the equal-tails interval and 1 is not!
- A better approach here is to create our interval of θ-values having the **Highest Posterior Density**.

HPD Intervals / Regions

Defn: A $100(1-\alpha)\%$ HPD region for θ is a subset $\mathcal{C} \in \Theta$ defined by

$$\mathcal{C} = \{\theta : p(\theta|\boldsymbol{y}) \geq k\}$$

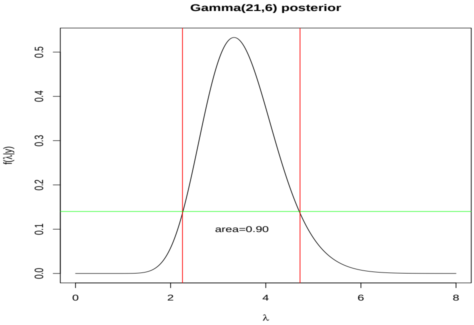
where k is the **largest** number such that

$$\int_{\theta : p(\theta|\boldsymbol{y}) \geq k} p(\theta|\boldsymbol{y})\, d\theta = 1 - \alpha.$$

- The value k can be thought of as a horizontal line placed over the posterior density whose intersection(s) with the posterior define regions with probability $1 - \alpha$.

HPD Intervals / Regions

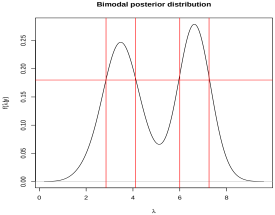
Picture: (90% HPD Interval)



$\Rightarrow P\{\theta_L^* < \theta < \theta_U^*\} = 0.90.$
The values between $\theta_L^* = 2.25$ and $\theta_U^* = 4.72$ here have the **highest posterior density**.

HPD Intervals / Regions

- The HPD region will be an **interval** when the posterior is **unimodal**.
- If the posterior is multimodal, the HPD region might be a **discontiguous set**.
  Picture:



- The set $\{\theta : \theta \in (2.85, 4.1) \cup (6.0, 7.25)\}$ is the HPD region for θ here.

Example 1 Revisited: HPD Interval

- See course web page for finding an HPD interval in `R` for λ in the fraud risk call example.
- A 90% quantile-based credible interval for λ is $(2.345, 4.844)$.
- Also note the `hpd` function in `TeachingDemos` package in `R`.
- See code for Example 2 (coin-flipping data) in R.

The Normal-Normal Model

- Why is it so common to model data using a normal distribution?
- Approximately normally distributed quantities appear often in nature.
- CLT tells us any variable that is basically a sum of independent components should be approximately normal.
- Note $\bar{y}$ and S^2 are independent when sampling from a normal population — so if beliefs about the mean are independent of beliefs about the variance, a normal model may be appropriate.

Why Normal Models?

- The normal model is analytically convenient (exponential family, sufficient statistics $\bar{y}$ and S^2)
- Inference about the population mean based on a normal model will be correct as $n \to \infty$ even if the data are truly non-normal.
- When we assume a normal likelihood, we can get a wide class of posterior distributions by using different priors.

A Conjugate analysis with Normal Data (variance known)

- Simple situation: Assume data $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$, with μ unknown and σ^2 known.
- We will make inference about μ.
- The likelihood is

$$L(\mu|\boldsymbol{y}) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2}(Y_i-\mu)^2}$$

- The parameter of interest μ can take values from $-\infty$ to ∞.
- A conjugate prior for μ is $\mu \sim N(\delta, \tau^2)$:

$$p(\mu) = (2\pi\tau^2)^{-1/2} e^{-\frac{1}{2\tau^2}(\mu-\delta)^2}$$

A Conjugate analysis with Normal Data (variance known)

So the posterior is:

$$
\begin{aligned}
p(\mu|\boldsymbol{y}) &\propto L(\mu|\boldsymbol{y})p(\mu) \\
&\propto \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(Y_i-\mu)^2} e^{-\frac{1}{2\tau^2}(\mu-\delta)^2} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)^2 + \frac{1}{\tau^2}(\mu-\delta)^2\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i^2-2Y_i\mu+\mu^2) + \frac{1}{\tau^2}(\mu^2-2\mu\delta+\delta^2)\right]\right\}
\end{aligned}
$$

A Conjugate analysis with Normal Data (variance known)

So the posterior is:

$$
\begin{aligned}
p(\mu|\boldsymbol{y}) &\propto \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2\tau^2}\left(\tau^2\sum Y_i^2 - 2\tau^2\mu n\bar{y} + n\mu^2\tau^2\right.\right. \\
&\qquad \left.\left. + \sigma^2\mu^2 - 2\sigma^2\mu\delta + \sigma^2\delta^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2}\frac{1}{\sigma^2\tau^2}\left[\mu^2(\sigma^2 + n\tau^2) - 2\mu(\delta\sigma^2 + \tau^2 n\bar{y})\right.\right. \\
&\qquad \left.\left. + \left(\delta^2\sigma^2 + \tau^2\sum Y_i^2\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\mu^2\left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right) - 2\mu\left(\frac{\delta}{\tau^2} + \frac{n\bar{y}}{\sigma^2}\right) + k\right]\right\}
\end{aligned}
$$

(where k is some constant)

$$\text{Hence } p(\mu|\boldsymbol{y}) \propto \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)\left(\mu^2 - 2\mu\left(\frac{\frac{\delta}{\tau^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right)+k\right)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right)\left(\mu - \frac{\frac{\delta}{\tau^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right)^2\right]\right\}$$

A Conjugate analysis with Normal Data (variance known)

- Hence the posterior for μ is simply a normal distribution with mean

$$\frac{\frac{\delta}{\tau^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}$$

and variance

$$\left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right)^{-1} = \frac{\tau^2\sigma^2}{\sigma^2 + n\tau^2}$$

- The **precision** is the reciprocal of the **variance**.
- Here, $\frac{1}{\tau^2}$ is the **prior precision** ...
- $\frac{n}{\sigma^2}$ is the **data precision** ...
- ... and $\frac{1}{\tau^2} + \frac{n}{\sigma^2}$ is the **posterior precision**.

A Conjugate analysis with Normal Data (variance known)

- Note the posterior mean $E[\mu|\mathbf{y}]$ is simply

$$\frac{1/\tau^2}{1/\tau^2 + n/\sigma^2}\delta + \frac{n/\sigma^2}{1/\tau^2 + n/\sigma^2}\bar{y},$$

  a combination of the **prior mean** and the **sample mean**.
- If the prior is highly precise, the weight is large on δ.
- If the data are highly precise (e.g., when n is large), the weight is large on $\bar{y}$.
- Clearly as $n \to \infty$, $E[\mu|\mathbf{y}] \approx \bar{y}$, and $var[\mu|\mathbf{y}] \approx \frac{\sigma^2}{n}$ if we choose a large prior variance τ^2.
- This implies that for τ^2 large and n large, Bayesian and frequentist inference about μ will be nearly identical.

A Conjugate analysis with Normal Data (mean known)

- Now suppose $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$ with μ known and σ^2 unknown.
- We will make inference about σ^2.
- Our likelihood

$$L(\sigma^2|\mathbf{y}) \propto (\sigma^2)^{-\frac{n}{2}} e^{-\frac{n}{2\sigma^2}[\frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu)^2]}$$

- Let W denote the sufficient statistic $\frac{1}{n}\sum(Y_i - \mu)^2$.
- The conjugate prior for σ^2 is the **inverse gamma** distribution.
- If a r.v. $Y \sim$ gamma, then $1/Y \sim$ inverse gamma (IG).
- The prior for σ^2 is

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)} e^{-(\beta/\sigma^2)} \quad \text{for } \sigma^2 > 0$$

where $\alpha > 0, \beta > 0$.

A Conjugate analysis with Normal Data (mean known)

- Note the prior mean and variance are

$$E(\sigma^2) = \frac{\beta}{\alpha - 1} \text{ provided that } \alpha > 1$$

$$var(\sigma^2) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \text{ provided that } \alpha > 2$$

- So the posterior for σ^2 is:

$$\begin{aligned} p(\sigma^2|\boldsymbol{y}) &\propto L(\sigma^2|\boldsymbol{y})p(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{n}{2}} e^{-\frac{n}{2\sigma^2}w} (\sigma^2)^{-(\alpha+1)} e^{-(\beta/\sigma^2)} \\ &= (\sigma^2)^{-(\alpha+\frac{n}{2}+1)} e^{-\frac{\beta+\frac{n}{2}w}{\sigma^2}} \end{aligned}$$

- Hence the posterior is clearly an IG$(\alpha + \frac{n}{2}, \beta + \frac{n}{2}w)$ distribution, where $w = \frac{1}{n}\sum(Y_i - \mu)^2$. **Conjugate!**

A Conjugate analysis with Normal Data (mean known)

- How to choose the prior parameters α and β?
- Note

$$\alpha = \frac{[E(\sigma^2)]^2}{var(\sigma^2)} + 2 \text{ and } \beta = E(\sigma^2)\left\{\frac{[E(\sigma^2)]^2}{var(\sigma^2)} + 1\right\}$$

so we could make guesses about $E(\sigma^2)$ and $var(\sigma^2)$ and use these to determine α and β.

A Model for Normal Data (mean and variance both unknown)

- When $Y_1, \ldots, Y_n$ are iid $N(\mu, \sigma^2)$ with both μ, σ^2 **unknown**, the conjugate prior for the mean explicitly depends on the variance:

$$p(\sigma^2) \propto (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2}$$

$$p(\mu|\sigma^2) \propto (\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2/s_0}(\mu-\delta)^2}$$

- The prior parameter s_0 measures the analyst's confidence in the prior specification.
- When s_0 is large, we strongly believe in our prior.

A Model for Normal Data (mean and variance both unknown)

The joint posterior for (μ, σ^2) is:

$$p(\mu, \sigma^2|\boldsymbol{y}) \propto L(\mu, \sigma^2|\boldsymbol{y})p(\sigma^2)p(\mu|\sigma^2)$$

$$\propto (\sigma^2)^{-\alpha-\frac{n}{2}-\frac{3}{2}} e^{-\frac{\beta}{\sigma^2}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)^2-\frac{1}{2\sigma^2/s_0}(\mu-\delta)^2}$$

$$= (\sigma^2)^{-\alpha-\frac{n}{2}-\frac{3}{2}} e^{-\frac{\beta}{\sigma^2}-\frac{1}{2\sigma^2}(\sum Y_i^2-2n\bar{y}\mu+n\mu^2)-\frac{1}{2\sigma^2/s_0}(\mu^2-2\mu\delta+\delta^2)}$$

$$= \left[(\sigma^2)^{-\alpha-\frac{n}{2}-\frac{1}{2}} e^{-\frac{\beta}{\sigma^2}-\frac{1}{2\sigma^2}(\sum Y_i^2-n\bar{y}^2)}\right]$$

$$\times \left[(\sigma^2)^{-1} e^{-\frac{1}{2\sigma^2}\{(n+s_0)\mu^2-2(n\bar{y}+\delta s_0)\mu+(n\bar{y}^2+s_0\delta^2)\}}\right]$$

Note the second part is simply a **normal kernel** for μ.

A Model for Normal Data (mean and variance both unknown)

- To get the posterior for σ^2, we integrate out μ:

$$p(\sigma^2|\boldsymbol{y}) = \int_{-\infty}^{\infty} p(\mu, \sigma^2|\boldsymbol{y})\,\mathrm{d}\mu$$
$$\propto (\sigma^2)^{-\alpha-\frac{n}{2}-\frac{1}{2}} e^{-\frac{1}{\sigma^2}[\beta+\frac{1}{2}(\sum Y_i^2 - n\bar{y}^2)]}$$

  since the second piece (which depends on μ) just integrates to a normalizing constant.

- Hence since $-\alpha - \frac{n}{2} - \frac{1}{2} = -(\alpha + \frac{n}{2} - \frac{1}{2}) - 1$, we see the posterior for σ^2 is inverse gamma:

$$\sigma^2|\boldsymbol{y} \sim IG\left(\alpha + \frac{n}{2} - \frac{1}{2}, \beta + \frac{1}{2}\sum(Y_i - \bar{y})^2\right)$$

A Model for Normal Data (mean and variance both unknown)

- Note that

$$p(\mu|\sigma^2, \boldsymbol{y}) = \frac{p(\mu, \sigma^2|\boldsymbol{y})}{p(\sigma^2|\boldsymbol{y})}$$

- After lots of cancellation,

$$\begin{aligned} p(\mu|\sigma^2, \boldsymbol{y}) \propto \sigma^{-2} \exp\{-\tfrac{1}{2\sigma^2}[(n+s_0)\mu^2 - 2(n\bar{y} + \delta s_0)\mu \\ + (n\bar{y}^2 + s_0\delta^2)]\} \\ = \sigma^{-2} \exp\left\{-\tfrac{1}{2\sigma^2/(n+s_0)}\left[\mu^2 - 2\tfrac{n\bar{y}+\delta s_0}{n+s_0}\mu + \tfrac{n\bar{y}^2+s_0\delta^2}{n+s_0}\right]\right\} \end{aligned}$$

- Clearly $p(\mu|\sigma^2, \boldsymbol{y})$ is **normal**:

$$\mu|\sigma^2, \boldsymbol{y} \sim N\left(\frac{n\bar{y} + \delta s_0}{n + s_0}, \frac{\sigma^2}{n + s_0}\right)$$

A Model for Normal Data (mean and variance both unknown)

- Note as $s_0 \to 0$, $\mu | \sigma^2, \boldsymbol{y} \dot{\sim} N\left(\bar{y}, \frac{\sigma^2}{n}\right)$.
- Note also the conditional posterior mean is

$$\left(\frac{n}{n + s_0}\right)\bar{y} + \left(\frac{s_0}{n + s_0}\right)\delta.$$

- The relative sizes of n and s_0 determine the weighting of the sample mean $\bar{y}$ and the prior mean δ.

A Model for Normal Data (mean and variance both unknown)

The marginal posterior for μ is:

$$p(\mu|\boldsymbol{y}) = \int_0^\infty p(\mu, \sigma^2|\boldsymbol{y})\, d\sigma^2$$
$$= \int_0^\infty (\sigma^2)^{-\alpha-\frac{n}{2}-\frac{3}{2}} \exp\left[-\frac{2\beta + (s_0+n)(\mu-\delta)^2}{2\sigma^2}\right] d\sigma^2$$

Letting $A = 2\beta + (s_0 + n)(\mu - \delta)^2$, $z = \frac{A}{2\sigma^2} \Rightarrow \sigma^2 = \frac{A}{2z}$ and $d\sigma^2 = -\frac{A}{2z^2}dz$,

A Model for Normal Data (mean and variance both unknown)

$$
\begin{aligned}
p(\mu|\boldsymbol{y}) &\propto \int_0^\infty \left(\frac{A}{2z}\right)^{-\alpha-\frac{n}{2}-\frac{3}{2}} \frac{A}{2z^2} e^{-z}\,\mathrm{d}z \\
&= \int_0^\infty \left(\frac{A}{2z}\right)^{-\alpha-\frac{n}{2}-\frac{1}{2}} \frac{1}{z} e^{-z}\,\mathrm{d}z \\
&\propto A^{-\alpha-\frac{n}{2}-\frac{1}{2}} \int_0^\infty z^{-\alpha-\frac{n}{2}-\frac{1}{2}-1} e^{-z}\,\mathrm{d}z
\end{aligned}
$$

This integrand is the kernel of a gamma density and thus the integral is a constant. So

A Model for Normal Data (mean and variance both unknown)

$$
\begin{aligned}
p(\mu|\boldsymbol{y}) &\propto A^{-\alpha-\frac{n}{2}-\frac{1}{2}} \\
&= \left[2\beta + (s_0 + n)(\mu - \delta)^2\right]^{-\frac{2\alpha+n+1}{2}} \\
&\propto \left[1 + \frac{(s_0 + n)(\mu - \delta)^2}{2\beta}\right]^{-\frac{2\alpha+n+1}{2}}
\end{aligned}
$$

which is a (scaled) noncentral t kernel having noncentrality parameter δ and degrees of freedom $n + 2\alpha$.

Example 1: Midge Data

- **Example 1**: $Y_1, \ldots, Y_9$ are a random sample of midge wing lengths (in mm). Assume the $Y_i's \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$.
- Example 1(a): If we know $\sigma^2 = 0.01$, make inference about μ. (See R example)
- A Bayesian point estimate for the population mean midge wing length is the posterior mean, 1.806 mm.
- A 95% credible interval for μ is $(1.741, 1.871)$, so with posterior probability 0.95, the population mean midge wing length is between 1.741 and 1.871 mm.

Example 1: Midge Data

- Example 1(b): Make inference about μ **and** σ^2, both **unknown** (see R example).
- This requires choosing the hyperparameters α and β of the inverse gamma prior on σ^2.
- 95% credible interval for σ^2: $(0.012, 0.028)$, with posterior median 0.0188.
- To approximate the posterior distribution for μ: We will randomly generate many values from the posterior distribution of σ^2.
- Then we will generate many values from the posterior of μ, given each respective generated value of σ^2.
- 95% credible interval for μ: $(1.727, 1.90)$, with posterior median 1.81 mm.

Example 2: Brain Data

- The textbook has an example of Bayesian inference about the mean hippocampal volume of the brain in a population of college football players who have a history of concussions.
- **Example 2**: $Y_1, \ldots, Y_{25}$ are a random sample of hippocampal volumes (in cm^3) of such football players. Assume the $Y_i's \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$.
- Example 2(a): If we know $\sigma = 0.5 \Rightarrow \sigma^2 = 0.25$, make inference about μ. We assume a $N(6.5, 0.4^2)$ prior on μ.
- The posterior mean is 5.78 cm^3. With posterior probability 0.95, the mean hippocampal volume of the brains for the population of concussed players is between 5.59 and $5.97 cm^3$.