Conjugate Models

Patrick Lam

Outline

Outline

Outline

Conjugacy

Conjugacy

Suppose we have a Bayesian model with a likelihood $p(y|\theta)$ and a prior $p(\theta)$.

Conjugacy

Suppose we have a Bayesian model with a likelihood $p(y|\theta)$ and a prior $p(\theta)$.

If we multiply our likelihood and prior, we get our posterior $p(\theta|y)$ up to a constant of proportionality.

Conjugacy

Suppose we have a Bayesian model with a likelihood $p(y|\theta)$ and a prior $p(\theta)$.

If we multiply our likelihood and prior, we get our posterior $p(\theta|y)$ up to a constant of proportionality.

If our posterior is a distribution that is of the same family as our prior, then we have *conjugacy*.

Conjugacy

Suppose we have a Bayesian model with a likelihood $p(y|\theta)$ and a prior $p(\theta)$.

If we multiply our likelihood and prior, we get our posterior $p(\theta|y)$ up to a constant of proportionality.

If our posterior is a distribution that is of the same family as our prior, then we have *conjugacy*. We say that the prior is conjugate to the likelihood.

Conjugacy

Suppose we have a Bayesian model with a likelihood $p(y|\theta)$ and a prior $p(\theta)$.

If we multiply our likelihood and prior, we get our posterior $p(\theta|y)$ up to a constant of proportionality.

If our posterior is a distribution that is of the same family as our prior, then we have *conjugacy*. We say that the prior is conjugate to the likelihood.

Conjugate models are great because we know the exact distribution of the posterior so we can easily simulate or derive quantities of interest analytically.

Conjugacy

Suppose we have a Bayesian model with a likelihood $p(y|\theta)$ and a prior $p(\theta)$.

If we multiply our likelihood and prior, we get our posterior $p(\theta|y)$ up to a constant of proportionality.

If our posterior is a distribution that is of the same family as our prior, then we have *conjugacy*. We say that the prior is conjugate to the likelihood.

Conjugate models are great because we know the exact distribution of the posterior so we can easily simulate or derive quantities of interest analytically.

In practice, we rarely have conjugacy.

Brief List of Conjugate Models

Brief List of Conjugate Models

Likelihood	Prior	Posterior
Binomial	Beta	Beta
Negative Binomial	Beta	Beta
Poisson	Gamma	Gamma
Geometric	Beta	Beta
Exponential	Gamma	Gamma
Normal (mean unknown)	Normal	Normal
Normal (variance unknown)	Inverse Gamma	Inverse Gamma
Normal (mean and variance unknown)	Normal/Gamma	Normal/Gamma
Multinomial	Dirichlet	Dirichlet

Outline

A Binomial Example

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

$$Y \sim \mathrm{Binomial}(n, \pi)$$

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

$$Y \sim \mathrm{Binomial}(n, \pi)$$

Quantity of interest: π (voter turnout)

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

$$Y \sim \text{Binomial}(n, \pi)$$

Quantity of interest: π (voter turnout)

Assumptions:

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

$$Y \sim \text{Binomial}(n, \pi)$$

Quantity of interest: π (voter turnout)

Assumptions:

- Each voter's decision to vote follows the Bernoulli distribution.

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

$$Y \sim \text{Binomial}(n, \pi)$$

Quantity of interest: π (voter turnout)

Assumptions:

- Each voter's decision to vote follows the Bernoulli distribution.
- Each voter has the same probability of voting.

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

$$Y \sim \mathrm{Binomial}(n, \pi)$$

Quantity of interest: π (voter turnout)

Assumptions:

- Each voter's decision to vote follows the Bernoulli distribution.
- Each voter has the same probability of voting. (unrealistic)

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

$$Y \sim \text{Binomial}(n, \pi)$$

Quantity of interest: π (voter turnout)

Assumptions:

- Each voter's decision to vote follows the Bernoulli distribution.
- Each voter has the same probability of voting. (unrealistic)
- Each voter's decision to vote is independent.

A Binomial Example

Suppose we have vector of data on voter turnout for a random sample of n voters in the 2004 US Presidential election.

We can model the voter turnout with a binomial model.

$$Y \sim \text{Binomial}(n, \pi)$$

Quantity of interest: π (voter turnout)

Assumptions:

- Each voter's decision to vote follows the Bernoulli distribution.
- Each voter has the same probability of voting. (unrealistic)
- Each voter's decision to vote is independent. (unrealistic)

The Conjugate Beta Prior

The Conjugate Beta Prior

We can use the beta distribution as a prior for π, since the beta distribution is conjugate to the binomial distribution.

The Conjugate Beta Prior

We can use the beta distribution as a prior for π, since the beta distribution is conjugate to the binomial distribution.

$$p(\pi|y) \propto p(y|\pi)p(\pi)$$

The Conjugate Beta Prior

We can use the beta distribution as a prior for π, since the beta distribution is conjugate to the binomial distribution.

$$
\begin{aligned}
p(\pi|y) &\propto p(y|\pi)p(\pi) \\
&= \mathrm{Binomial}(n,\pi) \times \mathrm{Beta}(\alpha,\beta)
\end{aligned}
$$

The Conjugate Beta Prior

We can use the beta distribution as a prior for π, since the beta distribution is conjugate to the binomial distribution.

$$
\begin{aligned}
p(\pi|y) &\propto p(y|\pi)p(\pi) \\
&= \mathrm{Binomial}(n,\pi) \times \mathrm{Beta}(\alpha,\beta) \\
&= \binom{n}{y}\pi^{y}(1-\pi)^{(n-y)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{(\alpha-1)}(1-\pi)^{(\beta-1)}
\end{aligned}
$$

The Conjugate Beta Prior

We can use the beta distribution as a prior for π, since the beta distribution is conjugate to the binomial distribution.

$$
\begin{aligned}
p(\pi|y) &\propto p(y|\pi)p(\pi) \\
&= \mathrm{Binomial}(n,\pi) \times \mathrm{Beta}(\alpha,\beta) \\
&= \binom{n}{y}\pi^{y}(1-\pi)^{(n-y)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{(\alpha-1)}(1-\pi)^{(\beta-1)} \\
&\propto \pi^{y}(1-\pi)^{(n-y)}\pi^{(\alpha-1)}(1-\pi)^{(\beta-1)}
\end{aligned}
$$

The Conjugate Beta Prior

We can use the beta distribution as a prior for π, since the beta distribution is conjugate to the binomial distribution.

$$
\begin{aligned}
p(\pi|y) &\propto p(y|\pi)p(\pi) \\
&= \mathrm{Binomial}(n,\pi) \times \mathrm{Beta}(\alpha,\beta) \\
&= \binom{n}{y} \pi^y (1-\pi)^{(n-y)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{(\alpha-1)} (1-\pi)^{(\beta-1)} \\
&\propto \pi^y (1-\pi)^{(n-y)} \pi^{(\alpha-1)} (1-\pi)^{(\beta-1)}
\end{aligned}
$$

$$
p(\pi|y) \propto \pi^{y+\alpha-1}(1-\pi)^{n-y+\beta-1}
$$

The Conjugate Beta Prior

We can use the beta distribution as a prior for π, since the beta distribution is conjugate to the binomial distribution.

$$\begin{aligned}
p(\pi|y) &\propto p(y|\pi)p(\pi) \\
&= \text{Binomial}(n,\pi) \times \text{Beta}(\alpha,\beta) \\
&= \binom{n}{y}\pi^{y}(1-\pi)^{(n-y)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{(\alpha-1)}(1-\pi)^{(\beta-1)} \\
&\propto \pi^{y}(1-\pi)^{(n-y)}\pi^{(\alpha-1)}(1-\pi)^{(\beta-1)}
\end{aligned}$$

$$p(\pi|y) \propto \pi^{y+\alpha-1}(1-\pi)^{n-y+\beta-1}$$

The posterior distribution is simply a $\text{Beta}(y+\alpha, n-y+\beta)$ distribution.

The Conjugate Beta Prior

We can use the beta distribution as a prior for π, since the beta distribution is conjugate to the binomial distribution.

$$
\begin{aligned}
p(\pi|y) &\propto p(y|\pi)p(\pi) \\
&= \mathrm{Binomial}(n,\pi) \times \mathrm{Beta}(\alpha,\beta) \\
&= \binom{n}{y}\pi^{y}(1-\pi)^{(n-y)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{(\alpha-1)}(1-\pi)^{(\beta-1)} \\
&\propto \pi^{y}(1-\pi)^{(n-y)}\pi^{(\alpha-1)}(1-\pi)^{(\beta-1)}
\end{aligned}
$$

$$
p(\pi|y) \propto \pi^{y+\alpha-1}(1-\pi)^{n-y+\beta-1}
$$

The posterior distribution is simply a Beta$(y+\alpha, n-y+\beta)$ distribution. Effectively, our prior is just adding $\alpha-1$ successes and $\beta-1$ failures to the dataset.

The Uninformative (Flat) Uniform Prior

The Uninformative (Flat) Uniform Prior

Suppose we have no strong prior beliefs about the parameters.

The Uninformative (Flat) Uniform Prior

Suppose we have no strong prior beliefs about the parameters. We can choose a prior that gives equal weight to all possible values of the parameters, essentially an uninformative or "flat" prior.

The Uninformative (Flat) Uniform Prior

Suppose we have no strong prior beliefs about the parameters. We can choose a prior that gives equal weight to all possible values of the parameters, essentially an uninformative or "flat" prior.

$$p(\pi) = \text{constant}$$

for all values of π.

The Uninformative (Flat) Uniform Prior

Suppose we have no strong prior beliefs about the parameters. We can choose a prior that gives equal weight to all possible values of the parameters, essentially an uninformative or "flat" prior.

$$p(\pi) = \text{constant}$$

for all values of π.

For the binomial model, one example of a flat prior is the Beta(1,1) prior:

The Uninformative (Flat) Uniform Prior

Suppose we have no strong prior beliefs about the parameters. We can choose a prior that gives equal weight to all possible values of the parameters, essentially an uninformative or "flat" prior.

$$p(\pi) = \text{constant}$$

for all values of π.

For the binomial model, one example of a flat prior is the Beta(1,1) prior:

$$p(\pi) \quad = \quad \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \pi^{(1-1)}(1-\pi)^{(1-1)}$$

The Uninformative (Flat) Uniform Prior

Suppose we have no strong prior beliefs about the parameters. We can choose a prior that gives equal weight to all possible values of the parameters, essentially an uninformative or "flat" prior.

$$p(\pi) = \text{constant}$$

for all values of π.

For the binomial model, one example of a flat prior is the Beta(1,1) prior:

$$\begin{aligned} p(\pi) &= \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \pi^{(1-1)} (1-\pi)^{(1-1)} \\ &= 1 \end{aligned}$$

The Uninformative (Flat) Uniform Prior

Suppose we have no strong prior beliefs about the parameters. We can choose a prior that gives equal weight to all possible values of the parameters, essentially an uninformative or "flat" prior.

$$p(\pi) = \text{constant}$$

for all values of π.

For the binomial model, one example of a flat prior is the Beta(1,1) prior:

$$\begin{aligned} p(\pi) &= \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \pi^{(1-1)} (1-\pi)^{(1-1)} \\ &= 1 \end{aligned}$$

which is the Uniform distribution over the $[0, 1]$ interval.

Since we know that a Binomial likelihood and a Beta(1,1) prior produces a Beta$(y+1, n-y+1)$ posterior, we can simulate the posterior in R.

Since we know that a Binomial likelihood and a Beta(1,1) prior produces a Beta($y+1, n-y+1$) posterior, we can simulate the posterior in R.

Suppose our turnout data had 500 voters, of which 285 voted.

Since we know that a Binomial likelihood and a Beta(1,1) prior produces a Beta$(y+1, n-y+1)$ posterior, we can simulate the posterior in R.

Suppose our turnout data had 500 voters, of which 285 voted.

```
> table(turnout)

turnout
  0   1
215 285
```

Since we know that a Binomial likelihood and a Beta(1,1) prior produces a Beta($y+1, n-y+1$) posterior, we can simulate the posterior in R.

Suppose our turnout data had 500 voters, of which 285 voted.

```
> table(turnout)

turnout
  0   1
215 285
```

Setting our prior parameters at $\alpha = 1$ and $\beta = 1$,

Since we know that a Binomial likelihood and a Beta(1,1) prior produces a Beta($y+1, n-y+1$) posterior, we can simulate the posterior in R.

Suppose our turnout data had 500 voters, of which 285 voted.

```
> table(turnout)

turnout
  0   1
215 285
```

Setting our prior parameters at $\alpha = 1$ and $\beta = 1$,

```
> a <- 1
> b <- 1
```

Since we know that a Binomial likelihood and a Beta(1,1) prior produces a Beta($y + 1, n - y + 1$) posterior, we can simulate the posterior in R.

Suppose our turnout data had 500 voters, of which 285 voted.

```
> table(turnout)

turnout
  0   1
215 285
```

Setting our prior parameters at $\alpha = 1$ and $\beta = 1$,

```
> a <- 1
> b <- 1
```

we get the posterior

Since we know that a Binomial likelihood and a Beta(1,1) prior produces a Beta($y+1, n-y+1$) posterior, we can simulate the posterior in R.

Suppose our turnout data had 500 voters, of which 285 voted.

```
> table(turnout)

turnout
  0   1
215 285
```

Setting our prior parameters at $\alpha = 1$ and $\beta = 1$,

```
> a <- 1
> b <- 1
```

we get the posterior

```
> posterior.unif.prior <- rbeta(10000, shape1 = 285 + a, shape2 = 500 -
+     285 + b)
```

Outline

Outline

Normal Model with Unknown Mean, Known Variance

Normal Model with Unknown Mean, Known Variance

Suppose we wish to estimate a model where the likelihood of the data is normal with an unknown mean μ and a known variance σ^2.

Normal Model with Unknown Mean, Known Variance

Suppose we wish to estimate a model where the likelihood of the data is normal with an unknown mean μ and a known variance σ^2.

Our parameter of interest is μ.

Normal Model with Unknown Mean, Known Variance

Suppose we wish to estimate a model where the likelihood of the data is normal with an unknown mean μ and a known variance σ^2.

Our parameter of interest is μ.

We can use a conjugate Normal prior on μ, with mean μ_0 and variance τ_0^2.

Normal Model with Unknown Mean, Known Variance

Suppose we wish to estimate a model where the likelihood of the data is normal with an unknown mean μ and a known variance σ^2.

Our parameter of interest is μ.

We can use a conjugate Normal prior on μ, with mean μ_0 and variance τ_0^2.

$$p(\mu|\mathbf{y}, \sigma^2) \quad \propto \quad p(\mathbf{y}|\mu, \sigma^2)p(\mu)$$

Normal Model with Unknown Mean, Known Variance

Suppose we wish to estimate a model where the likelihood of the data is normal with an unknown mean μ and a known variance σ^2.

Our parameter of interest is μ.

We can use a conjugate Normal prior on μ, with mean μ_0 and variance τ_0^2.

$$
\begin{aligned}
p(\mu|\mathbf{y}, \sigma^2) &\propto p(\mathbf{y}|\mu, \sigma^2)p(\mu) \\
\mathrm{Normal}(\mu_1, \tau_1^2) &= \mathrm{Normal}(\mu, \sigma^2) \times \mathrm{Normal}(\mu_0, \tau_0^2)
\end{aligned}
$$

Let θ represent our parameter of interest, in this case μ.

Let θ represent our parameter of interest, in this case μ.

$$p(\theta|\mathbf{y}) \quad \propto \quad \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right) \times \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right)$$

Let θ represent our parameter of interest, in this case μ.

$$\begin{aligned}
p(\theta|\mathbf{y}) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right) \times \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&\propto \exp\left(-\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu_0)^2}{2\tau_0^2}\right)
\end{aligned}$$

Let θ represent our parameter of interest, in this case μ.

$$\begin{aligned}
\mathbf{p(\theta|y)} &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right) \times \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&\propto \exp\left(-\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&= \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right)\right]
\end{aligned}$$

Let θ represent our parameter of interest, in this case μ.

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right) \times \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&\propto \exp\left(-\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&= \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2 \sum_{i=1}^{n}(y_i-\theta)^2 + \sigma^2(\theta-\mu_0)^2\right)\right]
\end{aligned}
$$

Let θ represent our parameter of interest, in this case μ.

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\theta)^2}{2\sigma^2}\right) \times \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left(-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&\propto \exp\left(-\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu_0)^2}{2\tau_0^2}\right) \\
&= \exp\left[-\frac{1}{2}\left(\sum_{i=1}^{n} \frac{(y_i-\theta)^2}{\sigma^2} + \frac{(\theta-\mu_0)^2}{\tau_0^2}\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2 \sum_{i=1}^{n} (y_i-\theta)^2 + \sigma^2(\theta-\mu_0)^2\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2 \sum_{i=1}^{n} (y_i^2-2\theta y_i+\theta^2) + \sigma^2(\theta^2-2\theta\mu_0+\mu_0^2)\right)\right]
\end{aligned}
$$

We can multiply the $2\theta y_i$ term in the summation by $\frac{n}{n}$ in order to get the equations in terms of the sufficient statistic $\bar{y}$.

We can multiply the $2\theta y_i$ term in the summation by $\frac{n}{n}$ in order to get the equations in terms of the sufficient statistic $\bar{y}$.

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}(y_i^2 - 2\theta\frac{n}{n}y_i + \theta^2) + \sigma^2(\theta^2 - 2\theta\mu_0 + \mu_0^2)\right)\right]$$

We can multiply the $2\theta y_i$ term in the summation by $\frac{n}{n}$ in order to get the equations in terms of the sufficient statistic $\bar{y}$.

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}(y_i^2 - 2\theta\frac{n}{n}y_i + \theta^2) + \sigma^2(\theta^2 - 2\theta\mu_0 + \mu_0^2)\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}y_i^2 - \tau_0^2 2\theta n\bar{y} + \tau_0^2 n\theta^2 + \theta^2\sigma^2 - 2\theta\mu_0\sigma^2 + \mu_0^2\sigma^2\right)\right]
\end{aligned}
$$

We can multiply the $2\theta y_i$ term in the summation by $\frac{n}{n}$ in order to get the equations in terms of the sufficient statistic $\bar{y}$.

$$
\begin{aligned}
p(\theta|\mathbf{y}) \quad &\propto \quad \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}(y_i^2 - 2\theta\frac{n}{n}y_i + \theta^2) + \sigma^2(\theta^2 - 2\theta\mu_0 + \mu_0^2)\right)\right] \\
&= \quad \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}y_i^2 - \tau_0^2 2\theta n\bar{y} + \tau_0^2 n\theta^2 + \theta^2\sigma^2 - 2\theta\mu_0\sigma^2 + \mu_0^2\sigma^2\right)\right]
\end{aligned}
$$

We can then factor the terms into several parts. Since $\mu_0^2\sigma^2$ and $\tau_0^2\sum_{i=1}^{n} y_i^2$ do not contain θ, we can represent them as some constant k, which we will drop into the normalizing constant.

We can multiply the $2\theta y_i$ term in the summation by $\frac{n}{n}$ in order to get the equations in terms of the sufficient statistic $\bar{y}$.

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}(y_i^2 - 2\theta\frac{n}{n}y_i + \theta^2) + \sigma^2(\theta^2 - 2\theta\mu_0 + \mu_0^2)\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n} y_i^2 - \tau_0^2 2\theta n\bar{y} + \tau_0^2 n\theta^2 + \theta^2\sigma^2 - 2\theta\mu_0\sigma^2 + \mu_0^2\sigma^2\right)\right]
\end{aligned}
$$

We can then factor the terms into several parts. Since $\mu_0^2\sigma^2$ and $\tau_0^2\sum_{i=1}^{n} y_i^2$ do not contain θ, we can represent them as some constant k, which we will drop into the normalizing constant.

$$
p(\theta|\mathbf{y}) \propto \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\theta^2\left(\sigma^2 + \tau_0^2 n\right) - 2\theta\left(\mu_0\sigma^2 + \tau_0^2 n\bar{y}\right) + k\right)\right]
$$

We can multiply the $2\theta y_i$ term in the summation by $\frac{n}{n}$ in order to get the equations in terms of the sufficient statistic $\bar{y}$.

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}(y_i^2 - 2\theta\frac{n}{n}y_i + \theta^2) + \sigma^2(\theta^2 - 2\theta\mu_0 + \mu_0^2)\right)\right] \\ &= \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}y_i^2 - \tau_0^2 2\theta n\bar{y} + \tau_0^2 n\theta^2 + \theta^2\sigma^2 - 2\theta\mu_0\sigma^2 + \mu_0^2\sigma^2\right)\right] \end{aligned}$$

We can then factor the terms into several parts. Since $\mu_0^2\sigma^2$ and $\tau_0^2\sum_{i=1}^{n} y_i^2$ do not contain θ, we can represent them as some constant k, which we will drop into the normalizing constant.

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\theta^2\left(\sigma^2 + \tau_0^2 n\right) - 2\theta\left(\mu_0\sigma^2 + \tau_0^2 n\bar{y}\right) + k\right)\right] \\ &= \exp\left[-\frac{1}{2}\left(\theta^2\left(\frac{\sigma^2 + \tau_0^2 n}{\sigma^2\tau_0^2}\right) - 2\theta\left(\frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\sigma^2\tau_0^2}\right) + k\right)\right] \end{aligned}$$

We can multiply the $2\theta y_i$ term in the summation by $\frac{n}{n}$ in order to get the equations in terms of the sufficient statistic $\bar{y}$.

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}(y_i^2 - 2\theta\frac{n}{n}y_i + \theta^2) + \sigma^2(\theta^2 - 2\theta\mu_0 + \mu_0^2)\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\tau_0^2\sum_{i=1}^{n}y_i^2 - \tau_0^2 2\theta n\bar{y} + \tau_0^2 n\theta^2 + \theta^2\sigma^2 - 2\theta\mu_0\sigma^2 + \mu_0^2\sigma^2\right)\right]
\end{aligned}
$$

We can then factor the terms into several parts. Since $\mu_0^2\sigma^2$ and $\tau_0^2\sum_{i=1}^{n} y_i^2$ do not contain θ, we can represent them as some constant k, which we will drop into the normalizing constant.

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2\sigma^2\tau_0^2}\left(\theta^2\left(\sigma^2 + \tau_0^2 n\right) - 2\theta\left(\mu_0\sigma^2 + \tau_0^2 n\bar{y}\right) + k\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\theta^2\left(\frac{\sigma^2 + \tau_0^2 n}{\sigma^2\tau_0^2}\right) - 2\theta\left(\frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\sigma^2\tau_0^2}\right) + k\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\theta^2\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right) - 2\theta\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right) + k\right)\right]
\end{aligned}
$$

Let's multiply by $\frac{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$ in order to simplify the θ^2 term.

Let's multiply by $\frac{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$ in order to simplify the θ^2 term.

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta^2\left(\frac{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)-2\theta\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)+k\right)\right]$$

Let's multiply by $\frac{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$ in order to simplify the θ^2 term.

$$\begin{aligned}
p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta^2\left(\frac{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)-2\theta\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)+k\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta^2-2\theta\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)+k\right)\right]
\end{aligned}$$

Let's multiply by $\frac{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$ in order to simplify the θ^2 term.

$$\begin{aligned}
p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta^2\left(\frac{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)-2\theta\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)+k\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta^2-2\theta\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)+k\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta-\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)\right)^2\right]
\end{aligned}$$

Let's multiply by $\frac{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$ in order to simplify the θ^2 term.

$$
\begin{aligned}
p(\theta|\mathbf{y}) &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta^2\left(\frac{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)-2\theta\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)+k\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta^2-2\theta\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)+k\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta-\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)\right)^2\right]
\end{aligned}
$$

Finally, we have something that looks like the density function of a Normal distribution!

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta-\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)\right)^2\right]$$

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta-\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)\right)^2\right]$$

Posterior Mean: $\mu_1 = \dfrac{\left(\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta-\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)\right)^2\right]$$

Posterior Mean: $\mu_1 = \dfrac{\left(\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$

Posterior Variance: $\tau_1^2 = \left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)^{-1}$

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta-\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)\right)^2\right]$$

Posterior Mean: $\mu_1 = \dfrac{\left(\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$

Posterior Variance: $\tau_1^2 = \left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)^{-1}$

Posterior Precision: $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2}+\frac{n}{\sigma^2}$

$$p(\theta|\mathbf{y}) \quad \propto \quad \exp\left[-\frac{1}{2}\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)\left(\theta-\left(\frac{\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}}\right)\right)^2\right]$$

Posterior Mean: $\mu_1 = \dfrac{\left(\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$

Posterior Variance: $\tau_1^2 = \left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)^{-1}$

Posterior Precision: $\frac{1}{\tau_1^2} = \frac{1}{\tau_0^2}+\frac{n}{\sigma^2}$

Posterior Precision is just the sum of the prior precision and the data precision.

We can also look more closely at how the prior mean μ_0 and the posterior mean μ_1 relate to each other.

We can also look more closely at how the prior mean μ_0 and the posterior mean μ_1 relate to each other.

$$\mu_1 = \frac{\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)}$$

We can also look more closely at how the prior mean μ_0 and the posterior mean μ_1 relate to each other.

$$\begin{aligned}
\mu_1 &= \frac{\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)} \\
&= \frac{\frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\tau_0^2\sigma^2}}{\frac{\sigma^2 + n\tau_0^2}{\tau_0^2\sigma^2}}
\end{aligned}$$

We can also look more closely at how the prior mean μ_0 and the posterior mean μ_1 relate to each other.

$$
\begin{aligned}
\mu_1 &= \frac{\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)} \\
&= \frac{\frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\tau_0^2\sigma^2}}{\frac{\sigma^2 + n\tau_0^2}{\tau_0^2\sigma^2}} \\
&= \frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\sigma^2 + n\tau_0^2}
\end{aligned}
$$

We can also look more closely at how the prior mean μ_0 and the posterior mean μ_1 relate to each other.

$$
\begin{aligned}
\mu_1 &= \frac{\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)} \\
&= \frac{\frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\tau_0^2\sigma^2}}{\frac{\sigma^2 + n\tau_0^2}{\tau_0^2\sigma^2}} \\
&= \frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\sigma^2 + n\tau_0^2} \\
&= \frac{\mu_0\sigma^2}{\sigma^2 + n\tau_0^2} + \frac{\tau_0^2 n\bar{y}}{\sigma^2 + n\tau_0^2}
\end{aligned}
$$

We can also look more closely at how the prior mean μ_0 and the posterior mean μ_1 relate to each other.

$$\begin{aligned}
\mu_1 &= \frac{\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)} \\
&= \frac{\frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\tau_0^2\sigma^2}}{\frac{\sigma^2 + n\tau_0^2}{\tau_0^2\sigma^2}} \\
&= \frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\sigma^2 + n\tau_0^2} \\
&= \frac{\mu_0\sigma^2}{\sigma^2 + n\tau_0^2} + \frac{\tau_0^2 n\bar{y}}{\sigma^2 + n\tau_0^2}
\end{aligned}$$

- As n increases, data mean dominates prior mean.

We can also look more closely at how the prior mean μ_0 and the posterior mean μ_1 relate to each other.

$$
\begin{aligned}
\mu_1 &= \frac{\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)} \\
&= \frac{\frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\tau_0^2\sigma^2}}{\frac{\sigma^2 + n\tau_0^2}{\tau_0^2\sigma^2}} \\
&= \frac{\mu_0\sigma^2 + \tau_0^2 n\bar{y}}{\sigma^2 + n\tau_0^2} \\
&= \frac{\mu_0\sigma^2}{\sigma^2 + n\tau_0^2} + \frac{\tau_0^2 n\bar{y}}{\sigma^2 + n\tau_0^2}
\end{aligned}
$$

- As n increases, data mean dominates prior mean.
- As τ_0^2 decreases (less prior variance, greater prior precision), our prior mean becomes more important.

A Simple Example

A Simple Example

Suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

A Simple Example

Suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

```
> known.sigma.sq <- 16
> unknown.mean <- 68
> n <- 100
> heights <- rnorm(n, mean = unknown.mean, sd = sqrt(known.sigma.sq))
```

A Simple Example

Suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

```
> known.sigma.sq <- 16
> unknown.mean <- 68
> n <- 100
> heights <- rnorm(n, mean = unknown.mean, sd = sqrt(known.sigma.sq))
```

We believe that the heights are normally distributed with some unknown mean μ and a known variance $\sigma^2 = 16$.

A Simple Example

Suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

```
> known.sigma.sq <- 16
> unknown.mean <- 68
> n <- 100
> heights <- rnorm(n, mean = unknown.mean, sd = sqrt(known.sigma.sq))
```

We believe that the heights are normally distributed with some unknown mean μ and a known variance $\sigma^2 = 16$.

Suppose before we see the data, we have a prior belief about the distribution of μ. Let our prior mean $\mu_0 = 72$ and our prior variance $\tau_0^2 = 36$.

A Simple Example

Suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

```
> known.sigma.sq <- 16
> unknown.mean <- 68
> n <- 100
> heights <- rnorm(n, mean = unknown.mean, sd = sqrt(known.sigma.sq))
```

We believe that the heights are normally distributed with some unknown mean μ and a known variance $\sigma^2 = 16$.

Suppose before we see the data, we have a prior belief about the distribution of μ. Let our prior mean $\mu_0 = 72$ and our prior variance $\tau_0^2 = 36$.

```
> mu0 <- 72
> tau.sq0 <- 36
```

Our posterior is a Normal distribution with Mean $\frac{\left(\frac{\mu_0}{\tau_0^2}+\frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)}$ and Variance $\left(\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}\right)^{-1}$

Our posterior is a Normal distribution with Mean $\dfrac{\left(\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}\right)}{\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)}$ and Variance $\left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}$

```
> post.mean <- (mu0/tau.sq0 + (n * mean(heights)/known.sigma.sq))/(1/tau.sq0 +
+     n/known.sigma.sq)
> post.mean

[1] 68.03969

> post.var <- 1/(1/tau.sq0 + n/known.sigma.sq)
> post.var

[1] 0.1592920
```

Outline

Normal Model with Known Mean, Unknown Variance

Normal Model with Known Mean, Unknown Variance

Now suppose we wish to estimate a model where the likelihood of the data is normal with a known mean μ and an unknown variance σ^2.

Normal Model with Known Mean, Unknown Variance

Now suppose we wish to estimate a model where the likelihood of the data is normal with a known mean μ and an unknown variance σ^2.

Now our parameter of interest is σ^2.

Normal Model with Known Mean, Unknown Variance

Now suppose we wish to estimate a model where the likelihood of the data is normal with a known mean μ and an unknown variance σ^2.

Now our parameter of interest is σ^2.

We can use a conjugate inverse gamma prior on σ^2, with shape parameter α_0 and scale parameter β_0.

Normal Model with Known Mean, Unknown Variance

Now suppose we wish to estimate a model where the likelihood of the data is normal with a known mean μ and an unknown variance σ^2.

Now our parameter of interest is σ^2.

We can use a conjugate inverse gamma prior on σ^2, with shape parameter α_0 and scale parameter β_0.

$$p(\sigma^2|\mathbf{y}, \mu) \quad \propto \quad p(\mathbf{y}|\mu, \sigma^2)p(\sigma^2)$$

Normal Model with Known Mean, Unknown Variance

Now suppose we wish to estimate a model where the likelihood of the data is normal with a known mean μ and an unknown variance σ^2.

Now our parameter of interest is σ^2.

We can use a conjugate inverse gamma prior on σ^2, with shape parameter α_0 and scale parameter β_0.

$$\begin{aligned} p(\sigma^2|\mathbf{y},\mu) &\propto p(\mathbf{y}|\mu,\sigma^2)p(\sigma^2) \\ \text{Invgamma}(\alpha_1,\beta_1) &= \text{Normal}(\mu,\sigma^2) \times \text{Invgamma}(\alpha_0,\beta_0) \end{aligned}$$

Let θ represent our parameter of interest, in this case σ^2.

Let θ represent our parameter of interest, in this case σ^2.

$$p(\theta|\mathbf{y}, \mu) \quad \propto \quad \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i - \mu)^2}{2\theta}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right)$$

Let θ represent our parameter of interest, in this case σ^2.

$$\begin{aligned}
p(\theta|\mathbf{y},\mu) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&\propto \prod_{i=1}^{n} \theta^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right)
\end{aligned}$$

Let θ represent our parameter of interest, in this case σ^2.

$$
\begin{aligned}
p(\theta|\mathbf{y},\mu) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&\propto \prod_{i=1}^{n} \theta^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right)
\end{aligned}
$$

Let θ represent our parameter of interest, in this case σ^2.

$$\begin{aligned}
p(\theta|\mathbf{y},\mu) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&\propto \prod_{i=1}^{n} \theta^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0}{\theta} + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right)\right]
\end{aligned}$$

Let θ represent our parameter of interest, in this case σ^2.

$$
\begin{aligned}
p(\theta|\mathbf{y}, \mu) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&\propto \prod_{i=1}^{n} \theta^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0}{\theta} + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right)\right] \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{2\beta_0 + 2\left(\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right)}{2\theta}\right)\right]
\end{aligned}
$$

Let θ represent our parameter of interest, in this case σ^2.

$$\begin{aligned}
p(\theta|\mathbf{y},\mu) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)}\theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&\propto \prod_{i=1}^{n} \theta^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0}{\theta} + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right)\right] \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{2\beta_0 + 2\left(\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right)}{2\theta}\right)\right] \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0 + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}}{\theta}\right)\right]
\end{aligned}$$

Let θ represent our parameter of interest, in this case σ^2.

$$
\begin{aligned}
p(\theta|\mathbf{y},\mu) &\propto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&\propto \prod_{i=1}^{n} \theta^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right) \times \theta^{-(\alpha_0+1)} \exp\left(-\frac{\beta_0}{\theta}\right) \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0}{\theta} + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\theta}\right)\right] \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{2\beta_0 + 2\left(\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}\right)}{2\theta}\right)\right] \\
&= \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0 + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}}{\theta}\right)\right]
\end{aligned}
$$

This looks like the density of an inverse gamma distribution!

$$p(\theta|\mathbf{y},\mu) \quad \propto \quad \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0+\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}}{\theta}\right)\right]$$

$$
\begin{aligned}
p(\theta|\mathbf{y},\mu) &\propto \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0+\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}}{\theta}\right)\right] \\
\alpha_1 &= \alpha_0+\frac{n}{2}
\end{aligned}
$$

$$p(\theta|\mathbf{y},\mu) \quad \propto \quad \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0+\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}}{\theta}\right)\right]$$

$$\begin{aligned}
\alpha_1 &= \alpha_0 + \frac{n}{2} \\
\beta_1 &= \beta_0 + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}
\end{aligned}$$

$$p(\theta|\mathbf{y},\mu) \quad \propto \quad \theta^{-(\alpha_0+\frac{n}{2}+1)} \exp\left[-\left(\frac{\beta_0+\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}}{\theta}\right)\right]$$

$$\alpha_1 \quad = \quad \alpha_0 + \frac{n}{2}$$

$$\beta_1 \quad = \quad \beta_0 + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2}$$

Our posterior is an $\mathrm{Invgamma}(\alpha_0 + \frac{n}{2}, \beta_0 + \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2})$ distribution.

A Simple Example

A Simple Example

Again suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

A Simple Example

Again suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

```
> known.mean <- 68
> unknown.sigma.sq <- 16
> n <- 100
> heights <- rnorm(n, mean = known.mean, sd = sqrt(unknown.sigma.sq))
```

A Simple Example

Again suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

```
> known.mean <- 68
> unknown.sigma.sq <- 16
> n <- 100
> heights <- rnorm(n, mean = known.mean, sd = sqrt(unknown.sigma.sq))
```

We believe that the heights are normally distributed with a known mean $\mu = 68$ and some unknown variance σ^2.

A Simple Example

Again suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

```
> known.mean <- 68
> unknown.sigma.sq <- 16
> n <- 100
> heights <- rnorm(n, mean = known.mean, sd = sqrt(unknown.sigma.sq))
```

We believe that the heights are normally distributed with a known mean $\mu = 68$ and some unknown variance σ^2.

Suppose before we see the data, we have a prior belief about the distribution of σ^2. Let our prior shape $\alpha_0 = 5$ and our prior scale $\beta_0 = 20$.

A Simple Example

Again suppose we have some (fake) data on the heights (in inches) of a random sample of 100 individuals in the U.S. population.

```
> known.mean <- 68
> unknown.sigma.sq <- 16
> n <- 100
> heights <- rnorm(n, mean = known.mean, sd = sqrt(unknown.sigma.sq))
```

We believe that the heights are normally distributed with a known mean $\mu = 68$ and some unknown variance σ^2.

Suppose before we see the data, we have a prior belief about the distribution of σ^2. Let our prior shape $\alpha_0 = 5$ and our prior scale $\beta_0 = 20$.

```
> alpha0 <- 5
> beta0 <- 20
```

Our posterior is a inverse gamma distribution with shape $\alpha_0 + \frac{n}{2}$ and scale $\beta_0 + \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2}$

Our posterior is a inverse gamma distribution with shape $\alpha_0 + \frac{n}{2}$ and scale $\beta_0 + \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2}$

```
> alpha1 <- alpha0 + n/2
> beta1 <- beta0 + sum((heights - known.mean)^2)/2
> library(MCMCpack)
> posterior <- rinvgamma(10000, alpha1, beta1)
> post.mean <- mean(posterior)
> post.mean

[1] 12.88139

> post.var <- var(posterior)
> post.var

[1] 3.136047
```

Our posterior is a inverse gamma distribution with shape $\alpha_0 + \frac{n}{2}$ and scale $\beta_0 + \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2}$

```
> alpha1 <- alpha0 + n/2
> beta1 <- beta0 + sum((heights - known.mean)^2)/2
> library(MCMCpack)
> posterior <- rinvgamma(10000, alpha1, beta1)
> post.mean <- mean(posterior)
> post.mean

[1] 12.88139

> post.var <- var(posterior)
> post.var

[1] 3.136047
```

Hmm ... what if we increased our sample size?

```
> n <- 1000
> heights <- rnorm(n, mean = known.mean, sd = sqrt(unknown.sigma.sq))
> alpha1 <- alpha0 + n/2
> beta1 <- beta0 + sum((heights - known.mean)^2)/2
> posterior <- rinvgamma(10000, alpha1, beta1)
> post.mean <- mean(posterior)
> post.mean

[1] 15.92281

> post.var <- var(posterior)
> post.var

[1] 0.5058952
```