Multiple Parameter Models

Most of the time, it is more likely that we have more than one unknown parameter. For example, we may know that data is normally distributed, but we do not know their mean or variance. In such scenarios, we need to generalize the methods from single-parameter models to hold for the case of multiple unknown parameters.

Multinomial Moldel

In Chapter 2, we considered the binomial model. For the binomial model, we had binary data (success/failure), where we ultimately wanted to draw inference on the probability of success. We now extend this discrete distribution to the case of a variable with more than two categories. In this case, we use the multinomial distribution for modeling more than two outcomes (e.g. yes, maybe, no)

If $y \sim \text{Multi}(n, \theta_1, ..., \theta_k)$ then the likelihood function is as follows:

Recall:

Binomial: $p(y|\theta) = \frac{n!}{y!(n-y!)}\theta^y(1-\theta)^{n-y}$

Multinomial: $p(\vec{\boldsymbol{y}}|\vec{\boldsymbol{\theta}}) = \frac{n!}{y_1! \times y_2! \times \cdots \times \ y_k!}{\theta_1}^{y_1} \times {\theta_2}^{y_2} \times \cdots \times {\theta_k}^{y_k}$

$$p(\vec{\boldsymbol{y}}|\vec{\boldsymbol{\theta}}) \propto \prod_{j=1}^{k} {\theta_j}^{y_j}$$

where:
θ_j = probability of observing category j
y_j = number of observations in category j

In this case, y_k is the count of the number of observations in the k-th category. Additionally, note that $\sum_{i=1}^{k}\theta_i = 1$.

For convenience, we typically write $\theta = (\theta_1,\ ...,\ \theta_k)$. Ultimately, our goal is to draw inference regarding θ, or the probability of falling into each of the k categories.

When working with the binomial model, we used a beta prior. Now that we have extended the binomial distribution to the multinomial distribution, we also need to extend the beta prior. By extending the beta prior, we are able to get a Dirichlet distribution

Recall:

Beta: $p(\theta|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ \theta^{a-1}(1-\theta)^{b-1}$

Dirichlet: $p(\theta|\alpha_1, ..., \alpha_k) = \frac{\Gamma(\sum_{i=1}^{k}\alpha_i)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\ \prod_{i=1}^{k}\theta_i^{\alpha_i-1}$

where: $\alpha_j - 1$ = prior number of events in category j

$$\mathbb{E}(\theta_j) = \frac{\alpha_j}{\sum_{i=1}^{k}\alpha_i} \quad \& \quad \mathbb{V}(\theta_j) = \frac{\alpha_j(\sum_{i=1}^{k}\alpha_i - \alpha_j)}{(\sum_{i=1}^{k}\alpha_i)^2(\sum_{i=1}^{k}\alpha_i + 1)}$$

Using the multinomial likelihood and a general Dirichlet prior, what is the posterior distribution for θ?

$$p(\theta|\vec{\boldsymbol{y}}, \vec{\boldsymbol{\alpha}}) \propto p(\vec{\boldsymbol{y}}|\vec{\boldsymbol{\theta}})\ p(\vec{\boldsymbol{\theta}}|\vec{\boldsymbol{\alpha}}) \propto \prod_{j=1}^{k}{\theta_j}^{y_j}\ \prod_{i=1}^{k}\theta_i^{\alpha_i-1}$$

$$\left.\propto \prod_{j=1}^{k}{\theta_j}^{y_j+\alpha_j-1}\right\}\text{Kernel of Dirichlet}(\vec{\boldsymbol{y}} + \vec{\boldsymbol{\alpha}})$$

Just as we did with the beta distribution, if we set the hyperparameters for the Dirichlet distribution to all equal 1, then this is a uniform prior, which would be uninformative. Another uninformative option is to set all the hyperparameters equal to 0. This would be an improper prior, but it will lead to a proper posterior so long as each category has at least one observation from the data.

Example:

During the 2020 presidential election, many polls were taken to at- tempt to predict the outcome of the election.

Let:

θ_1 = proportion of voters favoring Trump

θ_2 = proportion of voters favoring Biden

θ_3 = proportion favoring other candidates

Additionally, suppose that an exit poll found that out of $n = 2104$ voters:

$y_1 = 945$ supported Trump

$y_2 = 1021$ supported Biden

$y_3 = 138$ supported other candidates

If we use a flat prior, what is the posterior distribution for $(\theta_1,\ \theta_2,\ \theta_3)$?

$$\begin{aligned} p(\vec{\boldsymbol{\theta}}|\vec{\boldsymbol{y}},\vec{\boldsymbol{\alpha}}) &= \text{Dirichlet}(y_1+\alpha_1,\ y_2+\alpha_2,\ y_3+\alpha_3) \\ &= \text{Dirichlet}(945+1,\ 1021+1,\ 138+1) \\ &= \text{Dirichlet}(946,\ 1022,\ 139) \end{aligned}$$

What is the posterior distribution for θ_1?

$$\begin{aligned} p(\theta_1|\vec{\boldsymbol{y}},\vec{\boldsymbol{\alpha}}) &= \text{Beta}(y_1+\alpha_1,\ (y_2+\alpha_2)+(y_3+\alpha_3)) \\ &= \text{Beta}(946,\ 1022+139) \\ &= \text{Beta}(946,\ 1161) \end{aligned}$$

Suppose we wanted to know the posterior probability that Biden had more support than Trump. What do we want to estimate, and how could we estimate it with a 95% posterior interval?

$$p(\theta_2 > \theta_1) \quad \Rightarrow \quad P(\theta_2 - \theta_1 > 0)$$

$$\text{95\% Credible Interval for } \theta_2 - \theta_1$$

$$[-0.004677508,\ 0.07740148]$$

Normal Model

When dealing with a normal model, we typically have data $y_1, ..., y_n \sim \mathbb{N}(\mu, \sigma^2)$ where both μ and σ^2 are unknown. Since our goal is to always make inference on the unknown parameters, we are generally interested in looking to determine the joint posterior distribution $p(\mu, \sigma^2|y)$. Practically, we will do this iteratively in order to draw inference on μ and σ^2 individually in order to simplify inference. First, we need to determine our posterior distribution, $p(\mu, \sigma^2|y$. Then, we factor this into two pieces: $p(\mu|\sigma^2, y)p(\sigma^2|y)$. From here, we can determine the marginal posterior of μ as $p(\mu|y)$ and the posterior predictive distribution, $p(\tilde{y}|y)$.

Note: $p(\mu, \sigma^2|y) = \dfrac{p(y|\mu, \sigma^2)p(\mu, \sigma^2)}{p(y)}$

Joint Prior: $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$

Posterior: $p(\mu, \sigma^2|y) \propto p(y|\mu, \sigma^2)p(\mu, \sigma^2) = \underline{p(y|\mu, \sigma^2)}\, p(\mu|\sigma^2)p(\sigma^2)$

$$p(y|\mu, \sigma^2) \propto (\sigma^2)^{\frac{-n}{2}} \exp\left\{\frac{-\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right\}$$

The prior distribution can equivalently be written as $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$

Uninformative Prior In conducting our posterior inference, we must first decided on our prior. Let's begin by considering the case of an uninformative prior. As previously discussed for the single parameter models, using $p(\mu) = \mathcal{N}(0, 1)$ and $p(\sigma^2) = \frac{1}{\sigma^2}$. Combining these two pieces gives a joint prior of $p(\mu, \sigma^2) = 1\sigma^2$

$$\mu|\sigma^2, y \sim \mathcal{N}(0, \infty) \implies \text{Constant}$$
$$\sigma^2 \sim \text{Inv-Gamma}(0, 0) = \frac{1}{\sigma^2}$$
$$p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2) \propto c \cdot \frac{1}{\sigma^2} = \frac{1}{\sigma^2} \implies \text{Joint Prior}$$

Let's use this to calculate the joint posterior density:

$$\begin{aligned}
p(\mu, \sigma^2|y) &\propto p(y|\mu, \sigma^2)p(\mu, \sigma^2) \\
&\propto \left[\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}\right] \frac{1}{\sigma^2} \\
&\propto (\cancel{2\pi}\sigma)^{-\frac{n}{2}} \cdot \exp\left\{\frac{-\sum (y_i - \mu)^2}{2\sigma^2}\right\} \cdot (\sigma^2)^{-1} \\
&\propto (\sigma^2)^{-(\frac{n}{2}+1)} \cdot \exp\left\{\frac{-\sum (y_i - \mu)^2}{2\sigma^2}\right\} \\
&\propto (\sigma^2)^{-(\frac{n}{2}+1)} \cdot \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n \left((y_i - \bar{y}) + (\bar{y} - \mu)\right)^2\right\} \\
&\propto (\sigma^2)^{-(\frac{n}{2}+1)} \cdot \exp\left\{\frac{-1}{2\sigma^2} \sum_{i=1}^n \left((y_i - \bar{y})^2 + 2(y_i - \bar{y})(\bar{y} - \mu) + (\bar{y} - \mu)^2\right)\right\} \\
&\propto (\sigma^2)^{-(\frac{n}{2}+1)} \cdot \exp\left\{\frac{-1}{2\sigma^2} \left(\sum_{i=1}^n (y_i - \bar{y})^2 + \cancel{2(\bar{y} - \mu)\sum_{i=1}^n (y_i - \bar{y})} + n(\bar{y} - \mu)^2\right)\right\}
\end{aligned}$$

$$\text{Joint Posterior:} \quad \propto (\sigma^2)^{-(\frac{n}{2}+1)} \cdot \exp\left\{\frac{-1}{2\sigma^2} \left((n-1)S^2 + n(\bar{y} - \mu)^2\right)\right\} \quad \text{Where } S^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y})^2$$

We now want to factor this into two pieces: $p(\mu|\sigma^2, y) \times p(\sigma^2|y)$

$$p(\mu, \sigma^2|y) = p(\mu|\sigma^2, y)p(\sigma^2|y)$$

We begin by first finding the distribution of $p(\sigma^2|y)$ **(Marginal Posterior)**

$$\begin{aligned}
p(\sigma^2|y) &= \int p(\mu, \sigma^2|y) d\mu \\
&\propto \int (\sigma^2)^{-(\frac{n}{2}+1)} \cdot \exp\left\{\frac{-1}{2\sigma^2}\left((n-1)S^2 + n(\bar{y}-\mu)^2\right)\right\} d\mu \\
&\propto (\sigma^2)^{-(\frac{n}{2}+1)} \int \exp\left\{\frac{-(n-1)S^2}{2\sigma^2}\right\} \exp\left\{\frac{-n(\bar{y}-\mu)^2}{2\sigma^2}\right\} d\mu \\
&\propto (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left\{\frac{-(n-1)S^2}{2\sigma^2}\right\} \int \exp\left\{\frac{-n(\bar{y}-\mu)^2}{2\sigma^2}\right\} d\mu \\
&\propto (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left\{\frac{-(n-1)S^2}{2\sigma^2}\right\} \sqrt{2\pi\sigma^2/n} \int \underbrace{\frac{1}{\sqrt{2\pi\sigma^2/n}} \exp\left\{\frac{-(\mu-\bar{y})^2}{2\sigma^2/n}\right\}}_{\text{Kernel of } \mathcal{N}(\bar{y}, \frac{\sigma^2}{n})} d\mu \\
&\propto (\sigma^2)^{-(\frac{n}{2}+1)} \exp\left\{\frac{-(n-1)S^2}{2\sigma^2}\right\} \cdot (\sigma^2)^{\frac{1}{2}} \\
&\propto (\sigma^2)^{-(\frac{n-1}{2}+1)} \exp\left\{\frac{-(n-1)S^2}{2\sigma^2}\right\} \Longrightarrow \text{Inv-Gamma}\left(\frac{n-1}{2}, \frac{(n-1)S^2}{2}\right) \\
&\qquad = \text{Inv-}\chi^2(n-1,\ S^2)
\end{aligned}$$

Then, we obtain the distribution of $p(\mu|\sigma^2, y)$. **(Conditional Posterior)**

$$\begin{aligned}
p(\mu|\sigma^2, y) &= \frac{p(\mu, \sigma^2|y)}{p(\sigma^2|y)} \propto p(\mu, \sigma^2|y) \\
&\propto \cancel{(\sigma^2)^{-(\frac{n}{2}+1)}} \cdot \exp\left\{\frac{-1}{2\sigma^2}\left((n-1)S^2 + n(\bar{y}-\mu)^2\right)\right\} \\
&\propto \cancel{\exp\left\{\frac{-1}{2\sigma^2}\left((n-1)S^2\right)\right\}} \exp\left\{\frac{-1}{2\sigma^2}\left(n(\bar{y}-\mu)^2\right)\right\} \\
&\propto \exp\left\{\frac{-n(\bar{y}-\mu)^2}{2\sigma^2}\right\} = \exp\left\{\frac{-(\mu-\bar{y})^2}{2\sigma^2/n}\right\} \Longrightarrow \mathcal{N}(\bar{y},\ \sigma^2/n)
\end{aligned}$$

How do we actually draw inference on μ and σ^2 in practice?

We only know y and n. However...

We need to know σ^2 to draw a conclusion about μ.

Procedure:

1. Draw a sample of σ^2 from its posterior $\Longrightarrow$ $\sigma^2|y \sim \text{Inv-}\chi^2(n-1,\ S^2)$
2. Use the sampled value of σ^2 from step 1 to draw a μ from its posterior $\Rightarrow$ $\mu|\sigma^2, y \sim \mathcal{N}\left(\bar{y},\ \sigma^2/n\right)$
3. Repeat steps 1 and 2 many many many times
4. Summarize the posteriors by summarizing the samples of μ and σ^2 (mean, histograms, quantiles, etc...)
5. For the posterior predictive distribution, sample from the likelihood using the sampled values of μ and σ^2

While it is useful to know the conditional posterior distribution of $p(\mu|\sigma^2, y)$, we generally are more interested in determining the marginal posterior distribution of $p(\mu|y)$. This marginal posterior distribution allows us to draw inference directly on the mean of the distribution without having to worry about σ^2. Often times, even when there are multiple unknown parameters, we primarily are concerned with drawing conclusions about one particular parameter, or at least for one parameter at a time. Because of this, we want to marginalize out the effect of any nuisance parameter so we have theposterior distribution of the parameter of interest. In the case of thenormal example, we are at the point where we want to be able todraw inference about μ by itself, so we σ^2 becomes a nuisance parameter that we need to marginalize out. We do this by integrating over all possible values of σ^2.

$$p(\mu|y) = \int p(\mu, \sigma^2|y)d\sigma^2 \propto \int (\sigma^2)^{-(\frac{n}{2}+1)} \cdot \exp\left\{\frac{-1}{2\sigma^2}\left((n-1)S^2 + n(\bar{y}-\mu)^2\right)\right\} d\sigma^2$$

$$= \text{Kernel of Inv-Gamma}\left(\frac{n}{2},\ \frac{(n-1)s^2 + n(\bar{y}-\mu)^2}{2}\right)$$

↳ Since this is the kernel of a probability distribution, we know that when multiplied by a normalizing constant it must integrate to one. As a result of this, we can rewrite this integral as the reciprocal of the normalizing constant of the Inv-Gamma distribution.

$$= \frac{\Gamma(\frac{n}{2})}{\left(\frac{(n-1)s^2+n(\bar{y}-\mu)^2}{2}\right)^{n/2}} \propto \underbrace{\Gamma\left(\frac{n}{2}\right)\left((n-1)s^2\right)^{-\frac{n}{2}}\left(1+\frac{1}{n-1}\left(\frac{(\mu-\bar{y})^2}{s^2/n}\right)\right)^{-\frac{n}{2}}}$$

↳ This is the pdf of a t_{n-1} distribution!

$$\mu|y \sim t_{n-1}\left(\bar{y},\ \frac{s^2}{n}\right) \qquad\qquad \frac{\mu-\bar{y}}{\sqrt{s^2/n}} \sim t_{n-1}$$

If we want to make posterior predictions for future values, we can similarly determine the posterior predictive distribution by integrating out all parameters:

$$p(\tilde{y}|y) = \int\int p(\tilde{y}|\mu, \sigma^2, y)p(\mu, \sigma^2|y)d\mu\ d\sigma^2$$

By similar methods as to what was used above, we can first integrate out μ to get $p(\tilde{y}|\sigma^2, y) \sim \mathcal{N}\left(\tilde{y},\ \frac{(n+1)\sigma^2}{n}\right)$. If we then integrate out σ^2, we are left with $p(\tilde{y}|y) \sim t_{n-1}\left(\tilde{y},\ \frac{(n+1)s^2}{n}\right)$

Overall, we have thus found the following:

Conditional Approach:	$p(\sigma^2|y) = \text{Inv-Gamma}\left(\frac{n-1}{2},\ \frac{(n-1)s^2}{2}\right)$ $p(\mu|\sigma^2, y) = \mathcal{N}\left(\bar{y},\ \frac{\sigma^2}{n}\right)$

Marginal Approach:	$p(\mu|y) = t_{n-1}(\bar{y},\ \frac{s^2}{n})$

Recall that the t distribution takes the following form:

$$P(\theta|\mu, \sigma^2, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\nu\pi\sigma^2\right)^{-\frac{1}{2}}\left(1+\frac{1}{\nu}\left(\frac{\theta-\mu}{\sigma}\right)^2\right)^{-\left(\frac{\nu+1}{2}\right)}$$

Example How long do guests usually stay in Las Vegas? Based on the sample of 100 guests, you find that the average length of stay is 4.1 nights, with a standard deviation of 3.2 nights. Suppose you have no previous information about the length of stays and use an uninformative prior. What is the probability that the average length of stay is longer than 5 nights?

Using A Conjugate Prior Let's again consider a normal model with unknown mean and variance. In this case we have that:

$$y_1, y_2, ..., y_n \sim \mathcal{N}(\mu, \sigma^2)$$

Instead of placing an uninformative prior of $p(\mu, \sigma^2)$, let's use a conjugate prior. To do this, we need to break up our prior into its two pieces.

$$p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$$

We will assign priors as follows:

$$\mu|\sigma^2 \sim \mathcal{N}\left(\mu_0,\ \frac{\sigma^2}{\kappa_0}\right) \qquad \sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2) = \text{Inv-Gamma}\left(\frac{\nu_0}{2},\ \frac{\nu_0\sigma_0^2}{2}\right)$$

Where...
μ_0 = prior mean
κ_0 = prior sample size
ν_0 = prior degrees of freedom
σ_0^2 = prior scale of σ^2

In the joint prior μ, and σ^2 are thus dependent on each other. If σ^2 is large, then the prior for μ will also have large variability. In other words, the variance of the prior is consistent with the sampling variability of y.

Putting these pieces together gives the joint prior:

$$\begin{aligned}
p(\mu, \sigma^2) &= p(\mu|\sigma^2)p(\sigma^2) \\
&\propto (\sigma^2)^{-\frac{1}{2}}\exp\left\{\frac{-1}{2\sigma^2}\left(\kappa_0(\mu-\mu_0)^2\right)\right\} \cdot (\sigma^2)^{-\frac{\nu_0}{2}+1}\exp\left\{\frac{-\nu_0\sigma_0^2}{2\sigma^2}\right\} \\
&\propto (\sigma^2)^{-(\frac{\nu_0}{2}+1)}\exp\left\{\frac{-1}{2\sigma^2}\left(\kappa_0(\mu-\mu_0)^2 + \nu_0\sigma_0^2\right)\right\} \\
&\implies \text{Kernel of Normal-Inverse } \chi^2\left(\mu_0,\ \frac{\sigma_0^2}{\kappa_0},\ \nu_0,\ \sigma_0^2\right)
\end{aligned}$$

This distribution is the creatively-named Normal-Inverse χ^2 distribution with parameters μ_0, $\frac{\sigma_0^2}{\kappa_0}$, ν_0, and σ_0^2. Now that we have our likelihood and our prior, we can determine our joint posterior:

$$\begin{aligned}
p(\mu, \sigma^2|y) &\propto p(y|\mu, \sigma^2)p(\mu, \sigma^2) \\
&\propto \underbrace{\left[\prod_{i=1}^{n}(\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}\right]}_{\text{Likelihood}} \underbrace{\left[(\sigma^2)^{-(\frac{\nu_0}{2}+1)}\exp\left\{\frac{-1}{2\sigma^2}\left(\kappa_0(\mu-\mu_0)^2+\nu_0\sigma_0^2\right)\right\}\right]}_{\text{Prior}} \\
&\propto (\sigma^2)^{-\frac{n}{2}}\exp\left\{\frac{-1}{2\sigma^2}\left(\sum_{i=1}^{n}(y_i-\mu)^2 + \kappa_0(\mu-\mu_0)^2\right)\right\} \cdot (\sigma^2)^{-(\frac{\nu_0}{2}+1)}\exp\left\{\frac{-\nu_0\sigma_0^2}{2\sigma^2}\right\} \\
&\propto (\sigma^2)^{-(\frac{\nu_0+n+1}{2}+1)}\exp\left\{\frac{-1}{2\sigma^2}\left((\kappa_0+n)\left(\mu-\frac{\kappa_0\mu_0+n\bar{y}}{\kappa_0+n}\right)^2+\nu_0\sigma_0^2\right)\right\} \\
&\implies \text{Kernel of Normal-Inverse } \chi^2\left(\mu_0,\ \frac{\sigma_0^2}{\kappa_0},\ \nu_0,\ \sigma_0^2\right)
\end{aligned}$$

From here, we can again decompose into two pieces: $p(\mu|\sigma^2, y)$ and $p(\sigma^2|y)$:

$$p(\mu, \sigma^2|y) = p(\mu|\sigma^2, y)p(\sigma^2|y)$$

Conditional Approach:

$$p(\sigma^2|y) = \text{Inv - Gamma}\left(\frac{\nu_n}{2}, \ \frac{\nu_n\sigma_n^2}{2}\right) = \text{Inv-}\chi^2(\nu_n, \ \sigma_n^2)$$

$$p(\mu|\sigma^2, y) = \mathcal{N}\left(\mu_n, \ \frac{\sigma^2}{\kappa_n}\right)$$

Marginal Approach: $p(\mu|y) = t_{n-1}\left(\mu_n, \ \frac{\sigma_n^2}{\kappa_n}\right)$

Where:

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

$$\kappa_n = \kappa_0 + n$$

$$\nu_n = \nu_0 + n$$

$$\sigma_n^2 = \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2}{\nu_0 + n}$$

Regression - A Special Case of the Normal Model

Suppose we have a simple linear regression model for a random sample of n observations where we are trying to predict the response $\boldsymbol{y}$ based on a single predictor $\boldsymbol{x}$ (which is take to be fixed and known). In this case we have the general form for the regression model:

$$y_i = \alpha + \beta x_i + \epsilon_i \quad \text{for } i = 1, 2, ..., n$$

In this notation, (x_i, y_i) is the data for observation i, α is the intercept, β is the population slope, and ϵ_i is the error term for observation i. We usually assume that the errors are normally distributed with mean 0 and variance σ^2. You may assume exchangeability. We therefore have the following likelihood:

$$y_i x_i, \alpha, \beta.\sigma^2 \sim \mathcal{N}(\alpha + \beta x_i, \ \sigma^2) \quad \text{for } i = 1, 2, ..., n$$

Our goal is ultimately to make inference on α, β, and σ^2. To start, we therefore set some prior distribution on these parameters, which we can do as follows:

$$p(\alpha, \beta, \sigma^2|\boldsymbol{x}) \propto \frac{1}{\sigma^2}$$

We can calculate the joint posterior distribution, $p(\alpha, \beta, \sigma^2|\boldsymbol{x}, \boldsymbol{y})$ as follows:

$$\begin{aligned}
p(\alpha, \beta, \sigma^2|x, y) &\propto p(y|\alpha, \beta, \sigma^2, x)p(\alpha, \beta, \sigma^2|x) \\
&\propto \left[\prod_{i=1}^{n}(\sigma^2)^{-\frac{1}{2}}\exp\left\{\frac{-1}{2\sigma^2}(y_i - (\alpha + \beta x_i))^2\right\}\right]\frac{1}{\sigma^2} \\
&= (\sigma^2)^{-(\frac{n}{2}-1)}\exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2\right\} \\
\text{Joint Posterior: } &= (\sigma^2)^{-(\frac{n}{2}-1)}\exp\left\{\frac{-1}{2\sigma^2}\left(\beta^2\sum_{i=1}^{n}x_i^2 - 2\beta\sum_{i=1}^{n}x_i y_i + n\alpha^2 - 2\alpha n\bar{y} + 2\alpha\beta n\bar{x} + \sum_{i=1}^{n}y_i^2\right)\right\}
\end{aligned}$$

We can decompose this into its pieces, calculating first $p(\sigma^2|\boldsymbol{x}, \boldsymbol{y})$

$$p(\sigma^2|x, y) = \int\int p(\alpha, \beta, \sigma^2|x, y)d\alpha \ d\beta$$

$$p(\sigma^2|x, y) = \text{Inv-Gamma}\left(\frac{n-2}{2}, \frac{\sum(y_i - \hat{y})^2}{2}\right)$$

where: $\hat{y} = \hat{\alpha} + \hat{\beta}x_i$

$$\hat{\beta} = \frac{rS_y}{S_x}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Then $p(\beta|\sigma^2, \boldsymbol{x}, \boldsymbol{y}) \propto \int p(\alpha, \beta, \sigma^2|x, y)d\alpha$

$$\propto \int (\sigma^2)^{-(\frac{n}{2}-1)} \exp\left\{\frac{-1}{2\sigma^2}\left(\beta^2\sum_{i=1}^n x_i^2 - 2\beta\sum_{i=1}^n x_iy_i + n\alpha^2 - 2\alpha n\bar{y} + 2\alpha\beta n\bar{x} + \sum_{i=1}^n y_i^2\right)\right\} d\alpha$$

$$\propto \cancel{(\sigma^2)^{-(\frac{n}{2}-1)}} \exp\left\{\frac{-1}{2\sigma^2}\left(\beta^2\sum_{i=1}^n x_i^2 - 2\beta\sum_{i=1}^n x_iy_i + \cancel{\sum_{i=1}^n y_i^2}\right)\right\} \int \exp\left\{\frac{-1}{2\sigma^2}\left(n\alpha^2 - 2\alpha n\bar{y} + 2\alpha\beta n\bar{x}\right)\right\} d\alpha$$

$$\propto c \times \int \exp\left\{\frac{-1}{2\sigma^2}\left(n\alpha^2 - 2\alpha n\bar{y} + 2\alpha\beta n\bar{x}\right) + (\bar{y} - \beta\bar{x})^2\right\} \exp\left\{\frac{-n}{2\sigma^2}(-(\bar{y} - \beta\bar{x})^2)\right\} d\alpha$$

where $c = \exp\left\{\frac{-1}{2\sigma^2}\left(\beta^2\sum_{i=1}^n x_i^2 - 2\beta\sum_{i=1}^n x_iy_i\right)\right\}$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}\left(\beta^2\sum_{i=1}^n x_i^2 - 2\beta\sum_{i=1}^n x_iy_i\right)\right\} \times \exp\left\{\frac{-n}{2\sigma^2}(-(\bar{y} - \beta\bar{x})^2)\right\}$$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}\left(\beta^2\sum_{i=1}^n x_i^2 - 2\beta\sum_{i=1}^n x_iy_i + n(\bar{y} - \beta\bar{x})^2\right)\right\}$$

$$\propto \exp\left\{\frac{-(\sum x_i^2 - n\bar{x}^2)}{2\sigma^2}\left(\beta - \frac{\sum x_iy_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}\right)^2\right\}$$

$$\implies \quad p(\beta|\sigma^2, x, y) = \mathcal{N}\left(\frac{\sum_{i=1}^n x_iy_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2},\ \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\right)$$

And finally $p(\alpha|\beta, \sigma^2, \boldsymbol{x}, \boldsymbol{y})$

$$\propto p(\alpha, \beta, \sigma^2|x, y)$$

$$\propto \cancel{(\sigma^2)^{-(\frac{n}{2}-1)}} \exp\left\{\frac{-1}{2\sigma^2}\left(\cancel{\beta^2\sum_{i=1}^n x_i^2 - 2\beta\sum_{i=1}^n x_iy_i} + n\alpha^2 - 2\alpha n\bar{y} + 2\alpha\beta n\bar{x} + \cancel{\sum_{i=1}^n y_i^2}\right)\right\}$$

$$\propto \exp\left\{\frac{-1}{2\sigma^2}\left(n\alpha^2 - 2\alpha n\bar{y} + 2\alpha\beta n\bar{x}\right)\right\}$$

$$\propto \exp\left\{\frac{-n}{2\sigma^2}\left(\alpha^2 - 2\alpha(\bar{y} - \beta\bar{x})\right)\right\} \quad \text{(Complete the Square)}$$

$$\propto \exp\left\{\frac{-n}{2\sigma^2}\left(\alpha^2 - 2\alpha(\bar{y} - \beta\bar{x}) + (\bar{y} - \beta\bar{x})^2\right)\right\}$$

$$= \exp\left\{\frac{-n}{2\sigma^2}(\alpha - (\bar{y} - \beta\bar{x}))^2\right\} \quad \implies \quad p(\alpha|\beta, \sigma^2, x, y) = \mathcal{N}\left(\bar{y} - \beta\bar{x},\ \frac{\sigma^2}{n}\right)$$