Bayesian Basics

We are interested in obtaining the posterior distribution for making inferences.

$$posterior = \frac{likelihood * prior}{\int likelihood * prior}$$

It is often easier to work without the denominator, and recognize the posterior density.

$$posterior \propto likelihood * prior$$

Want to know characteristics about the posterior (such as posterior mean), and sample from the posterior. Think of the likelihood as a way to connect the data and the parameters. The posterior represents what you should believe about the parameters after seeing data.

Review of Marginal and Conditional Distributions

$$p(y|x) = \frac{p(x,y)}{p(y)} \text{ is conditional of y given x}$$
$$p(x) = \int p(x,y)\delta y = \int p(y|x)p(x)\delta y \text{ is marginal of x}$$
$$\text{Under independence, } p(x,y) = p(y|x)p(x)$$

Bayesian Framework: Posterior Distribution

$$p(\theta|y) = \frac{p(\theta,y)}{p(y)} = \frac{p(\theta,y)}{\int p(\theta,y)d\theta} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} \propto p(y|\theta)p(\theta)$$

Basic Hierarchical Model Setup

Normal-Gamma Gaussian Linear Model:

$$(y_i|\theta,\sigma^2) \sim N\left(\theta,\sigma^2\right), i = 1,...,n$$
$$(\theta|\sigma^2) \sim N\left(\mu, t\tau^2\sigma^2\right)$$
$$\sigma^2 \sim Inv - Gamma(\frac{d}{2},\frac{\eta}{2})$$

Reframed to work with precision $\omega = 1/\sigma^2$, called the Normal-Gamma Prior for (θ,ω):

$$(y_i|\theta,\omega) \sim N\left(\theta,(\omega)^{-1}\right), i = 1,...,n$$
$$(\theta|\omega) \sim N\left(\mu,(k\omega)^{-1}\right)$$
$$\omega \sim Gamma(\frac{d}{2},\frac{\eta}{2})$$

Hyperparameter interpretation:

- μ = Prior guess for θ.
- τ = Prior signal-to-noise ratio for σ^2.
- d = Prior sample size for σ^2.
- η = Prior sum of squares for σ^2; $\frac{\eta}{d} \approx$ prior guess for σ^2.
- $k = \frac{1}{\tau^2}$ is prior sample size for θ.

Prior

Joint prior:

$$p(\theta, \omega) = p(\theta|\omega)p(\omega)$$

Marginal priors:

$$\begin{aligned} p(\omega) &= p(\omega) \\ p(\theta|\omega) &= \int p(\theta, \omega) d\omega = \int p(\theta|\omega)p(\omega) d\omega \end{aligned}$$

Likelihood (Sampling Model)

$$p(y|\theta, \omega) = \prod_{i=1}^{n} p(y_i|\theta, \omega)$$

Posterior

Joint posterior (up to a constant of proportionality):

$$\begin{aligned} p(\theta, \omega|y) &\propto p(y|\theta, \omega)p(\theta, \omega) \\ &= p(y|\theta, \omega)p(\theta|\omega)p(\omega) \end{aligned}$$

Marginal posteriors.

$$\begin{aligned} p(\theta|y) &= \int p(\theta, \omega|y) d\omega \\ p(\omega|y) &= \int p(\theta, \omega|y) d\theta \end{aligned}$$

Conditional posteriors. (Full conditionals.)

$$\begin{aligned} p(\theta|y, \omega) &\propto p(y|\theta, \omega)p(\theta) \\ p(\omega|y, \theta) &\propto p(y|\theta, \omega)p(\omega) \end{aligned}$$

- Obtain the joint posterior, and read off forms of conditionals by considering all other variables constant.
- Need all full conditionals to sample from joint posterior using Gibbs Sampler or similar. (Is easier than sampling from joint posterior.)

More Complex Hierarchical Model Setup

Normal-Gamma Gaussian Multivariate Linear Model (Bayesian Regression):

$$\begin{aligned}
(y|\beta, \omega, \Lambda) &\sim N\left(X\beta, (\omega\Lambda)^{-1}\right) \\
\lambda_i &\overset{iid}{\sim} Gamma\left(\frac{h}{2}, \frac{h}{2}\right) \\
(\beta|\omega) &\sim N\left(m, (\omega K)^{-1}\right) \\
\omega &\sim Gamma\left(\frac{d}{2}, \frac{\eta}{2}\right)
\end{aligned}$$

Prior

Joint prior:

$$p(\beta, \omega, \Lambda) = p(\beta|\omega)p(\omega)p(\Lambda)$$

Likelihood

$$p(y|\beta, \omega, \Lambda) \propto N\left(X\beta, (\omega\Lambda)^{-1}\right)$$

Posterior

Joint posterior:

$$p(\beta, \omega, \Lambda|y) \propto p(y|\beta, \omega, \Lambda)p(\beta|\omega)p(\omega)p(\Lambda)$$

Full posterior conditionals.

$$\begin{aligned}
p(\beta|\omega, \Lambda, y) &\propto p(y|\beta, \omega, \Lambda)p(\beta|\omega)p(\omega)p(\Lambda) \\
&\propto p(y|\beta, \omega, \Lambda)p(\beta|\omega)
\end{aligned}$$

$$\begin{aligned}
p(\omega|\beta, \Lambda, y) &\propto p(y|\beta, \omega, \Lambda)p(\beta|\omega)p(\omega)p(\Lambda) \\
&\propto p(y|\beta, \omega, \Lambda)p(\beta|\omega)p(\omega)
\end{aligned}$$

$$\begin{aligned}
p(\lambda_i|y, \beta, \omega) &\propto p(y|\lambda_i, \beta, \omega)p(\lambda_i|\beta, \omega)p(\beta|\omega)p(\omega) \\
&\propto p(y|\lambda_i, \beta, \omega)p(\lambda_i)
\end{aligned}$$

- For full conditionals, read these off of the joint posterior, by considering all variables on the right side of the $|$ to be constant.

Messy Joint Posteriors

If cannot recognize the form of the joint posterior, can still try to characterize the joint posterior:

$$\begin{aligned} p(\theta, \omega, y) &= p(\theta|\omega, y)p(\omega|y), \text{ i.e.} \\ joint &= conditional * marginal \end{aligned}$$

So we can stare at a complex joint posterior $p(\theta, y)$ to read the conditional, considering only θ as a random variable, and ω, y as constants. Get rid of terms which are constant wrt θ, and note the form.

Then obtain the marginal, which we can do **without integration**.

$$p(\theta, \omega|y) = p(\theta|\omega, y)p(\omega|y) \qquad \rightarrow \qquad p(\omega|y) = \frac{p(\theta, \omega|y)}{p(\theta|\omega, y)}$$

$$joint = conditional * marginal \qquad \rightarrow \qquad marginal = \frac{joint}{conditional}$$

Normal Prior-Posterior Form (Avoid Completing the Square)

A normal-normal model has an easy form to avoid completing the square.

$$\begin{aligned} p(y|\theta) &\sim N(\theta, \Sigma) \\ p(\theta) &\sim N(m, V) \end{aligned}$$

Posterior takes form

$$\begin{aligned} p(\theta|y) &\sim p(y|\theta)p(\theta) \\ &\propto \exp\left\{-\frac{1}{2}\left[(\theta - m)^T V^{-1}(\theta - m) + (y - \theta)^T \Sigma^{-1}(y - \theta)\right]\right\} \end{aligned}$$

We could complete the square, or we can recognize this form: $-\frac{1}{2}(\theta - \mu)^T C^{-1}(\theta - \mu)$

This is $N(\mu, C)$ with

$$\begin{aligned} C^{-1} &= V^{-1} + \Sigma^{-1} \\ \mu &= C\left(V^{-1}m + n\Sigma^{-1}\bar{y}\right) \end{aligned}$$

Summarizing the Posterior: How To Do Inference

Want to summarize the posterior to make inferences about the parameters. There are a few ways to summarize. All of these methods are based on the idea of integrating some function of the posterior.

Point Estimation of Parameters

Posterior mean: $\bar{\theta} = \int \theta p(\theta|y) d\theta$

Uncertainty of Point Estimation of Parameters

Posterior variance of θ: $Var(\theta|y) = \int (\theta - \bar{\theta})^2 p(\theta|y) d\theta$

Prediction of New Data Value

Have $y_i \overset{iid}{\sim} p(y_i|\theta)$, for $i = 1...n$. Want to infer y_{n+1}.

$$p(y_{n+1}|\underbrace{y_1, ..., y_n}_{\text{"y"}}) = p(y_{n+1}|y)$$

Can always write

$$p(y) = \int \underbrace{p(y|x)p(x)}_{\text{p(y,x)}} dx$$

Therefore, the posterior predictive distribution is

$$p(y_{n+1}|y) = \int p(y_{n+1}|\theta, y) p(\theta|y) d\theta$$