The Gamma/Poisson Bayesian Model

- If our data $X_1, \ldots, X_n$ are iid Poisson(λ), then a gamma(α, β) prior on λ is a **conjugate** prior.
  Likelihood:

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^{n}(x_i!)}$$

  Prior:

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}, \quad \lambda > 0.$$

  $\Rightarrow$ Posterior:

$$\pi(\lambda|\mathbf{x}) \propto \lambda^{\sum x_i + \alpha - 1}e^{-(n+\beta)\lambda}, \quad \lambda > 0.$$

  $\Rightarrow \pi(\lambda|\mathbf{x})$ is gamma$\left(\sum x_i + \alpha, n + \beta\right)$. **(Conjugate!)**

The Gamma/Poisson Bayesian Model

- The posterior mean is:

$$\begin{aligned}
\hat{\lambda}_B &= \frac{\sum x_i + \alpha}{n + \beta} \\
&= \frac{\sum x_i}{n + \beta} + \frac{\alpha}{n + \beta} \\
&= \left[\frac{n}{n + \beta}\right]\left(\frac{\sum x_i}{n}\right) + \left[\frac{\beta}{n + \beta}\right]\left(\frac{\alpha}{\beta}\right)
\end{aligned}$$

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

Bayesian Learning

- We can use the Bayesian approach to update our information about the parameter(s) of interest sequentially as new data become available.
- Suppose we formulate a prior for our parameter θ and observe a random sample $\mathbf{x}_1$.
- Then the posterior is

$$\pi(\theta|\mathbf{x}_1) \propto p(\theta)L(\theta|\mathbf{x}_1)$$

- Then we observe a new (independent) sample $\mathbf{x}_2$.

Bayesian Learning

- We can use our previous posterior as the **new prior** and derive a **new** posterior:

$$
\begin{aligned}
\pi(\theta|\mathbf{x}_1, \mathbf{x}_2) &\propto \pi(\theta|\mathbf{x}_1)L(\theta|\mathbf{x}_2) \\
&\propto p(\theta)L(\theta|\mathbf{x}_1)L(\theta|\mathbf{x}_2) \\
&= p(\theta)L(\theta|\mathbf{x}_1, \mathbf{x}_2) \\
&\text{(since } \mathbf{x}_1, \mathbf{x}_2 \text{ independent)}
\end{aligned}
$$

- Note this is the same posterior we would have obtained had $\mathbf{x}_1$ and $\mathbf{x}_2$ arrived at the same time!
- This "sequential updating" process can continue indefinitely in the Bayesian setup.

CHAPTER 3 SLIDES START HERE

Why Normal Models?

- 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 $\bar{X}$ 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 $\bar{X}$ 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 $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$, with μ unknown and σ^2 known.
- We will make inference about μ.
- The likelihood is

$$L(\mu|\mathbf{x}) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

- 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:

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

A Conjugate analysis with Normal Data (variance known)

So the posterior is:

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

(where k is some constant)

A Conjugate analysis with Normal Data (variance known)

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

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

A Conjugate analysis with Normal Data (variance known)

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

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

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 $E[\mu|\mathbf{x}]$ is simply

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

  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 $\bar{x}$.
- Clearly as $n \to \infty$, $E[\mu|\mathbf{x}] \approx \bar{x}$, and $var[\mu|\mathbf{x}] \approx \frac{\sigma^2}{n}$ 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.