HW1

Alan Wu
STAT 5440: Applied Bayesian Modeling

February 10, 2026

1 Question 1

Show that the Beta-Binomial model is conjugate where: the prior is $p \sim Beta(\alpha, \beta)$, and the likelihood is $X \sim Binomial(n, p)$

We solve for the beta-binomial conjugacy we must show the following:

- Given prior $p \sim Beta(\alpha, \beta)$ and $X \sim Binomial(n, p)$
- Want to show that $p(p|x) \propto p(x|p)p(p)$, where p(p—x) is the posterior distribution and takes on $p(p|x) \sim Beta(\alpha + x, \beta + n - x)$

Some facts to recall:

- Since $p \sim Beta(\alpha, \beta)$ then $p(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{Beta(\alpha,\beta)}$
- $X \sim Binomial(n, p)$, meaning $p(x|p) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$

We combine both these for:

$$
\begin{aligned}
p(p|x) &\propto p(x|p) \cdot p(p) \\
\propto \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \cdot \frac{p^{\alpha-1}(1-p)^{\beta-1}}{Beta(\alpha, \beta)} \\
&\propto p^{x+\alpha-1} \cdot (1-p)^{\beta+n-x-1} \\
&\sim Beta(\alpha + x, \beta + n - x)
\end{aligned}
$$

2 Question 2

In the Beta-Binomial model, interpret how the posterior changes if you observe 3 successes out of 4 trials when the prior success and failure counts are the following:

- (0,0)
- (1,1)
- (1/2, 1/2)
- $\alpha > \beta$
- $\alpha < \beta$

Overlay the posterior distributions in one plot for each of these situations. How does this change if the observed success rate is 75% out of 100 trials? Generate the same plot as before but with the new number of observations.

The results of the 4 situations appear as the following:

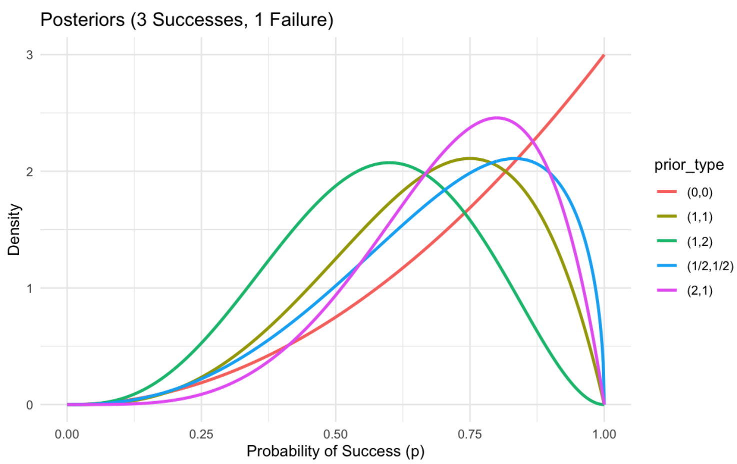

Figure 1: Overlaid posterior for 3 successes 1 failure

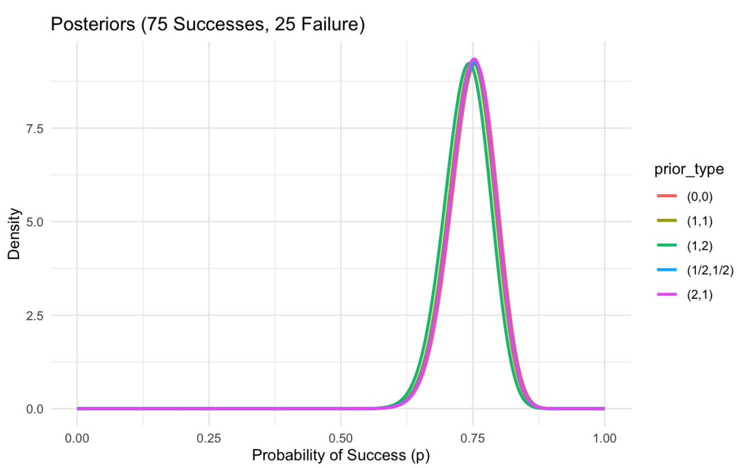

Figure 2: Overlaid posterior for 75 successes 25 failure

The overall trend that we realize is that when we introduce greater data, the posterior distribution tends to form to the parameters of the actual data. For the situation with 3 successes and 1 failure, at most we have a parameter of 2, either alpha or beta. Therefore, the prior has greater influence on the shape of the posterior distribution. However, when we increase the number of samples, despite the proportion of the number of successes being the same (75%), the posterior distribution shifts in favor of the data.

3 Question 3

Show that the posterior distribution for the Gamma-Poisson is conjugate.

- The prior is $\lambda \sim Gamma(\alpha, \beta)$
- The likelihood is $X_i \sim Poisson(\lambda)$ for $i \in 1, ..., n$

We can derive the Gamma conjugate prior with poisson likelihood by

- Given prior $\lambda \sim Gamma(\alpha, \beta)$ and $X_i \sim Poisson(\lambda)$ for $i \in 1, \ldots, n$
- Want to show that $p(\lambda|X) \propto p(X|\lambda)p(\lambda)$, where p(p—x) is the posterior distribution and takes on $p(p|x) \sim Gamma(\alpha + \sum_{i=1}^{n} X_i, \beta + n)$

Some facts to recall:

- Since $\lambda \sim Gamma(\alpha, \beta)$ then $p(\lambda) = \lambda^{\alpha-1} \cdot e^{-\beta\lambda} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)}$
- $X_i \sim Poisson(\lambda)$, meaning each individual likelihood $p(x_i|\lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$.

From here, before we prove the full proportional derivation of the posterior, we must compute the joint likelihood for the entire dataset, where $p(X|\lambda) = \prod_{i=1}^{n} p(x_i|\lambda)$. The joint likelihood can be derived through:

$$p(X|\lambda) = \prod_{i=1}^{n} p(x_i|\lambda)$$

$$= \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$= \prod_{i=1}^{n} \frac{1}{x_i!} \cdot e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i}$$

We now use $p(X|\lambda)$ the joint likelihood to derive the posterior distribution:

$$
\begin{aligned}
p(\lambda|X) &\propto p(X|\lambda) \cdot p(\lambda) \\
&\propto \prod_{i=1}^{n} \frac{1}{x_i!} \cdot e^{-n\lambda} \cdot \lambda^{\sum_{i=1}^{n} x_i} \cdot \lambda^{\alpha-1} \cdot e^{-\beta\lambda} \\
&\propto \lambda^{\alpha+\sum_{i=1}^{n} x_i - 1} \cdot e^{-(\beta+n)\lambda} \\
&\sim Gamma(\alpha + \sum_{i=1}^{n} x_i, \beta + n)
\end{aligned}
$$

4 Question 7

Assume a normal likelihood $y \sim N(\mu, \sigma^2)$, where:

- σ^2 is known
- μ has a conjugate prior: $\mu \sim N(\mu_0, \sigma_0^2)$

Prove the posterior distribution for is a normal distribution. How does the choice of both prior parameters correspond to having a weighted average of prior belief and observed information? HINT: Review how to "complete the square"

Our parameter of interest is μ and σ^2 is known. We want to prove that $p(\mu|y) \propto p(y|\mu) \cdot p(\mu) \sim Normal$, where the result is some normal distribution with shifted parameters based on the prior distribution.

- We are given $\mu \sim N(\mu_0, \sigma_0^2)$, meaning $p(\mu) = \frac{1}{\sigma_0\sqrt{2\pi}} exp[-\frac{1}{2}(\frac{\mu-\mu_0}{\sigma_0})^2]$
- We are also given $y \sim N(\mu, \sigma^2)$, so $p(y|\mu) = \frac{1}{\sigma\sqrt{2\pi}} exp[-\frac{1}{2}(\frac{y-\mu}{\sigma})^2]$

From here, we want to derive the posterior distribution by solving for $p(\mu|y) \propto p(y|\mu)p(\mu)$. For the first portion, we multiply the two probability functions:

$$
\begin{aligned}
p(\mu|y) &\propto p(y|\mu)p(\mu) \\
&\propto \frac{1}{\sigma\sqrt{2\pi}} exp[-\frac{1}{2}(\frac{y-\mu}{\sigma})^2] \cdot \frac{1}{\sigma_0\sqrt{2\pi}} exp[-\frac{1}{2}(\frac{\mu-\mu_0}{\sigma_0})^2] \\
&\propto \exp[-\frac{1}{2}(\frac{(y-\mu)^2}{\sigma_0^2} + \frac{(\mu-\mu_0)^2}{\sigma^2})] \\
&\propto \exp[-\frac{1}{2}(\frac{y^2-2\mu y+\mu^2}{\sigma^2} + \frac{\mu^2-2\mu\mu_0+\mu_0^2}{\sigma_0^2})] \\
&\propto \exp[-\frac{1}{2}((\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2})\mu^2 - (\frac{2\mu_0}{\sigma_0^2} + \frac{2y}{\sigma^2})\mu + \frac{\mu_0^2}{\sigma_0^2} + \frac{y^2}{\sigma^2})] \\
&\propto \exp[-\frac{1}{2}((\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2})\mu^2 - (\frac{2\mu_0}{\sigma_0^2} + \frac{2y}{\sigma^2})\mu)] \cdot \exp[-\frac{1}{2}(\frac{\mu_0^2}{\sigma_0^2} + \frac{y^2}{\sigma^2})] \\
&\propto \exp[-\frac{1}{2}((\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2})\mu^2 - (\frac{2\mu_0}{\sigma_0^2} + \frac{2y}{\sigma^2})\mu)]
\end{aligned}
$$

At this point, let us define some variables, where $a = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}$ and $b = \frac{2\mu_0}{\sigma_0^2} + \frac{2y}{\sigma^2}$ to make the derivation easier.

We can complete the square by adding and subtracting $(\frac{b}{2a})^2$ to our original expressions. Therefore, our expression becomes

$$\begin{aligned}
p(\mu|y) &\propto \exp[-\frac{1}{2}(a\mu^2 - b\mu + (\frac{b}{2a})^2 - (\frac{b}{2a})^2)] \\
&\propto \exp[-\frac{1}{2}(a\mu^2 - b\mu + (\frac{b}{2a})^2)] \cdot \exp[\frac{1}{2}(\frac{b}{2a})^2] \\
&\propto \exp[-\frac{a}{2}(\mu - \frac{b}{2a})^2]
\end{aligned}$$

We solve for the $\frac{b}{2a}$ and obtain:

$$\frac{b}{2a} = \frac{\frac{2\mu_0}{\sigma_0^2} + \frac{2y}{\sigma^2}}{2 \cdot (\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2})}$$

$$= \frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}} \cdot \mu_0 + \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}} \cdot y$$

From here, we return to the original expression:

$$p(\mu|y) \propto \exp[-\frac{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}}{2}(\mu - (\frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}} \cdot \mu_0 + \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}} \cdot y))^2]$$

This result represents a normal distribution with new parameters, where:

- $\mu = \frac{\frac{1}{\sigma_0^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}} \cdot \mu_0 + \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}} \cdot y$
- $\sigma^2 = \frac{1}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}}$

The posterior distribution reflects a synthesis of prior knowledge and new evidence, where the posterior mean functions as a **precision-weighted average**. In this framework, **precision**—defined as the reciprocal of the variance ($1/\sigma^2$)—quantifies the "certainty" or "information" contained within each source. The posterior mean μ_n is pulled toward the prior mean μ_0 or the observed data y based on their relative precisions; a highly certain (low-variance) prior will "anchor" the estimate, while highly precise data will "wash out" the prior belief. Simultaneously, the posterior precision is the sum of the prior and data precisions, $(\frac{1}{\sigma_n^2} = \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2})$, ensuring that the resulting posterior variance σ_n^2 is strictly smaller than both original variances.