Homework 1: Bayesian Inference

Paul Sabin

Spring 2025

Homework 1: Bayesian Inference

Complete the following problems and upload written responses from either a tablet or a picture of your work on canvas. For responses requiring code submit your responses as PDF document (via RMarkdown or Quarto) with code included in the problem. Code and typed responses should look professional.

Question 1: Deriving the Beta-Binomial Posterior

Problem

Show that the Beta-Binomial model is conjugate where:

- The prior is $p \sim \text{Beta}(\alpha, \beta)$,
- The likelihood is $X_i \sim \text{Binomial}(n, p)$ for $i \in 1, \ldots, n$.

Question 2: Interpreting the Posterior

Problem

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.

Question 3: Deriving Gamma Conjugate Prior with Poisson Likelihood

Problem

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

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

Question 4: Posterior Distribution on Real Data

Problem

The file `manchester_city_2024.csv` shows the goals scored `team_goals` and allowed `opp_goals` for the Manchester City soccer club in the 2023-24 season. You want to build a Bayesian model for how many goals you expect the club to score in a future match. Specify a conjugate likelihood and prior that fits this data. Justify your choice of prior distribution.

Report the posterior distribution in closed form and plot it over a histogram of the observed goals scored in the season. Do they line up? Is this expected?

Question 5: Posterior Distribution on Real Data

Problem

In the previous problem you defined a posterior distribution on how many goals you expect Manchester City to score in a match. Now you are curious about how many goals you expect them to allow (`opp_goals`) in a match. Repeat the same process in (4) by defining your likelihood & prior then deriving your posterior distribution. Be sure to justify your choice of prior distribution.

Question 6: Posterior Predictive

Your friend learned about your models from the previous problems and wants to know the probabilites that Manchester City wins, ties, or loses their next match. Use 1,000 samples from the posterior predictive distribution from your model in Problems 4&5 to estimate these probabilites.

Question 7: Deriving a Conjugate Prior for Mean with Known Variance

Problem

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 μ is a normal distribution. How does the choice of both prior parameters correspond to having a weighted average of prior belief and observed information?

Question 8: Preview of the Gibbs Sampler

Problem

Consider a normal model with unknown mean and unknown variance. Using the following steps, describe the process of iteratively sampling from the posterior distribution:

1. Sample $y_1, y_2, \ldots, y_n \sim N(10, 3.3)$ where $n = 100$. Prior to sampling set your seed with `set.seed(5)`.
2. Assign independent conjugate priors:
    - $\mu \sim N(0, 50)$,
    - $\sigma^2 \sim \text{Inverse-Gamma}(2, 2)$.

3. Iteratively sample 10,000 times where you use your most recent value of μ and σ^2 sampled in the conditional posterior distributions of the other parameter:
   - $\mu \mid \sigma^2, y$,
   - $\sigma^2 \mid \mu, y$.

Report a 95% posterior interval for both μ and σ^2 based on your samples of each. Are your samples of μ and σ^2 correlated with each other?
