

Midterm Practice

Concepts

- Likelihood (Lecture 1)
 - Identify proportionality constants that can be excluded
- Cromwell's Rule (Lecture 2)
- Sufficient Statistics (Lecture 1)
- Data Generating Process (Lecture 1)
- Bias, Variance, Mean Squared Error (Lecture 1)
- Mixture Model (Lecture 1)
- Conjugate Prior (Lecture 2, BR Sections 5.1,5.2)
 - Pseudo-counts interpretations of conjugate priors
 - Beta-Binomial (BR Chapter 3) and Gamma-Poisson model (BR Section 5.2) interpretations
- Improper Priors (Lecture 4)
- Posterior Predictive Distribution (Lecture 4, BR Section 8.3)
 - Integral definition involving likelihood and posterior (or prior)
- LOTUS: Law of the unconscious statistician (Lecture 5)
- Monte Carlo error (Lecture 5)
 - How does the variance of our Monte Carlo approximation decrease with sample size?
- Posterior Predictive Model Checking (Lecture 6)
 - Monte Carlo implementation
- Monte Carlo Sampling (Lecture 5)
 - Inversion Sampling
 - Rejection Sampling

Practice Problem 1

A machine produces products. Suppose the time needed for producing products, in minutes, are random and independent. Now we want to utilize our knowledge about Bayesian statistics to analyze this dataset. Suppose the time needed for one product follows an exponential distribution with rate λ , i.e. $Y_i \sim \text{i.i.d } \text{Expo}(\lambda)$.

1. What is the population? What is the sample? What is the estimand we are interested in?
2. Write out the data generating process for observation y_1, y_2, \dots, y_n .
3. Write out the log-likelihood function λ based on observation y_1, y_2, \dots, y_n .
4. Compute the MLE based on the given data.
5. Try to identify a conjugate prior distribution for λ (among the distributions discussed in class). The _____ prior distribution is the conjugate prior for the Exponential sampling model.
6. Compute the resulting posterior distribution based on the prior you identified in the above question.
7. Find the posterior mean of the posterior distribution. What is the mode of the posterior mode of the distribution? The posterior model is called the maximum a posteriori (MAP) estimate.
8. Assume y_1, \dots, y_n are $(2, 30, 40, 10, 24, 14, 6, 8, 20, 35)$. You believe *a priori* that the mean production time should be about 20 minutes. Choose prior parameters for your prior that reflects this belief (the prior variance can be anything you think is appropriate). Compute numeric value for the posterior mean and modes.

9. There are two kinds of intervals we learned for summarizing the posterior distribution: central interval and HPD (Highest posterior density), please describe how you would construct the 95% central interval and the 95% HPD in words.
10. Which interval/region for λ is shorter? Interpret the central interval you created from the above problem. How is this different from the frequentist confidence interval?
11. Let \tilde{y} be the time required to produce the next product. Compute the posterior predictive distribution. i.e. $p(\tilde{y}|y_1, \dots, y_n)$.

Practice Problem 2

1. Poisson-Gamma Practice.
 - (a) Assume $y_i \sim$ i.i.d $\text{Pois}(\lambda, \nu_i)$ where ν_i is the exposure for observation i . Write out the likelihood for λ given observations y_1, \dots, y_n . Simplify by removing proportionality constants.
 - (b) Assume the prior $p(\lambda) \sim \text{Gamma}(a, b)$. What is the posterior distribution for λ ?
 - (c) How do we interpret a and b in terms of pseudo-counts?
 - (d) Find the posterior mean. Write it in terms of a weighted average of the MLE and the prior mean.
2. Beta-Bernoulli practice
 - (a) Assume $y_i \sim$ i.i.d $\text{Bernoulli}(\theta)$. Write out the likelihood for θ given observations y_1, \dots, y_n . Simplify by removing proportionality constants.
 - (b) Assume the prior $p(\theta) \sim \text{Beta}(a, b)$. What is the posterior distribution for θ ?
 - (c) How do we interpret a and b in terms of pseudo-counts?
 - (d) Find the posterior mean. Write it in terms of a weighted average of the MLE and the prior mean.

Multiple Choice Quiz Questions to Review

Make sure you know *why* each of the right answers is true.

1. You observe $Y_i \sim \text{Binom}(10, \theta)$ for $i = 1, 2, \dots, n$, where $n = 100$ observations. Circle all sufficient statistics for θ
 - (a) y_1, y_2, \dots, y_n
 - (b) $\sum_{i=1}^n y_i$
 - (c) $\frac{\sum_{i=1}^n y_i}{n}$
 - (d) median(y_1, y_2, \dots, y_n)
 - (e) y_1
2. Circle all *true* statements:
 - (a) An estimator is a random variable and an estimate is a constant
 - (b) An estimator is a constant and an estimate is a random variable
 - (c) In classical frequentist inference, parameters are considered unknown constants and in Bayesian inference parameters are considered unknown random variables.
 - (d) In classical frequentist inference, parameters are considered unknown random variables and in Bayesian inference parameters are unknown considered constants.
 - (e) The bias of an estimator is defined as $E[\hat{\theta} - \theta]$
 - (f) The variance of an estimator is defined as $E[(\hat{\theta} - \theta)^2]$
3. In Bayesian statistics, *conjugacy* refers to the setting in which
 - (a) The sampling distribution and the prior distribution are in the same model family
 - (b) The sampling distribution and the posterior distribution are in the same model family
 - (c) The prior distribution and the posterior distribution are in the same model family
 - (d) The predictive distribution is in the same family as the sampling distribution

4. Fill in both blanks
- The _____ distribution is the conjugate prior distribution for the Binomial model.
 - The _____ distribution is the conjugate prior distribution for the Poisson model.
5. For a random variable with a $\text{Beta}(\alpha, \beta)$ distribution (circle one):
- As $\alpha + \beta$ increases, the variance of the random variable decreases
 - As $\alpha + \beta$ increases, the variance of the random variable increases
6. For a 95% Bayesian credible interval:
- The interval is random and will cover the parameter 95% of the time
 - The quantile interval is always shorter than the HPD interval
 - θ is random and $\Pr(l(y) < \theta < u(y) | y) = 0.95$
 - Only the likelihood is needed to derive the interval.
7. Select all answers that are true. The posterior predictive distribution
- depends on unknown parameters
 - is often used as a tool for model checking
 - incorporates sampling variability and variability due to uncertainty about the parameter
 - is obtained by integrating out future observations, \tilde{y} , from a joint model.
8. For both the Gamma-Poisson (Gamma prior, Poisson sampling model) and the Beta-Binomial model (Beta prior, Binomial sampling model), the posterior mean of the parameter
- can be expressed as a weighted average between the prior mean and the maximum likelihood estimate
 - can be expressed as a weighted average between the predictive mean and the maximum likelihood estimate
 - can be expressed as a weighted average between the prior mean and the predictive mean
 - does not depend on prior parameters