MVE550 2019 Lecture 7

Petter Mostad

Chalmers University

November 26, 2019

Overview

- The Multinomial Dirichlet conjugacy.
- Bayesian inference for Markov chains.
- Bayesian inference for HMMs.
- Bayesian inference for Branching processes.
- If time, the Normal Normal conjugacy.

The Multinomial Dirchlet conjugacy

- A vector $x = (x_1, \ldots, x_k)$ of non-negative integers has a Multinomial distribution with parameters n and p, where $n > 0$ is an integer and p is a probability vector of length k if $\sum_{i=1}^k x_i = n$ and the probability mass function is given by

$$\pi(x \mid n, p) = \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}.$$

- A vector $\theta = (\theta_1, \ldots, \theta_k)$ of non-negative real numbers satisfying $\sum_{i=1}^k \theta_i = 1$ has a Dirichlet distribution with parameter vector $\alpha = (\alpha_1, \ldots, \alpha_k)$, if it has probability density function

$$\pi(\theta \mid \alpha) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_k)}{\Gamma(\alpha_1)\Gamma(\alpha_2) \cdot \Gamma(\alpha_k)} \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \ldots \theta_k^{\alpha_k - 1}.$$

- We have conjugacy in this case.
- The predictive distribution is given by

$$\pi(x) = \frac{n!}{x_1! \ldots x_k!} \cdot \frac{\Gamma(\alpha_1 + x_1)}{\Gamma(\alpha_1)} \ldots \frac{\Gamma(\alpha_k + x_k)}{\Gamma(\alpha_k)} \cdot \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\Gamma(\sum_{i=1}^k \alpha_i + x_i)}.$$

Bayesian inference for discrete state space Markov chains

- The parametres are P, the transition matrix, and p, the probability vector for the initial value X_0.
- Idea: Specify a prior for the parameters, find the posterior given available data, and use the posteriors for predictions.
- One possibility: p fixed and

$$\pi(P) = \prod_{i=1}^{s} \text{Dirichlet}(P_i; \alpha_i)$$

  where s is the size of the state space, P_i is the i'th row of P, and α_i is a vector of length s of positive parameters: Most often, $\alpha = (1, 1, \ldots, 1)$.
- We get the posterior

$$\pi(P \mid \text{data}) = \prod_{i=1}^{s} \text{Dirichlet}(P_i; \alpha_i + c_i)$$

  where c_i is the vector of counts of observed transitions starting at state i.

Prediction

- Assume you have observed $x_0, x_1, \dots, x_k$ as the first $k+1$ steps of a Markov chain, and would like to predict the probability distribution for x_{k+1}. Then

$$\pi(x_{k+1} \mid x_0, \dots, x_k) = \int P_{x_k, x_{k+1}} \pi(P_{x_k} \mid x_0, \dots, x_k) \, dP_{x_k}.$$

- For each possible value of x_{k+1} this is the expectation of the posterior for $P_{x_k, x_{k+1}}$.
- Using the Dirichlet distributions above in the prior, we get

$$\pi(x_{k+1} \mid x_0, \dots, x_k) = \frac{\alpha_{x_k} + c_{x_k}}{\alpha_{x_k,1} + \cdots + \alpha_{x_k,s} + c_{x_k,1} + \cdots + c_{x_k,s}}.$$

- To predict longer sequences $x_{k+1}, x_{k+2}, \dots$, it is possible to derive formulas, or one can simulate them stepwise: Then, at each step, the previously simulated values are added to the data.

Bayesian inference for HMMs

- Many different inference questions can be raised, depending on the data that is available.
- We will assume
  - We have observed $X_0, \ldots, X_n$ and $Y_0, \ldots, Y_n$
  - We use a model where the parameters are p and P for the underlying X chain, and a matrix Q with $Q_{ij} = \Pr(Y_k = j \mid X_k = i)$ of *emittance probabilities*.
- Then, the inference for p and P, and for Q, can be done separately.
- The posterior for Q will of course depend on the choice of prior for Q.
- Examples.

Bayesian inference for Branching processes

- The parameter of a Branching process is the probability vector a for the offspring process.
- We assume the data is a set of counts $y_1, y_2, \ldots, y_n$ representing the outcomes of n realizations of the offspring process.
- As usual, we choose a prior for for the parameter a, obtain the posterior given the data, and use the posterior for predictions.
- Examples.

The Normal Normal conjugacy

- Assume $y \sim \text{Normal}\left(\theta, \frac{1}{\tau_y}\right)$ where θ is unknown and the *precision* τ_y is known and fixed. Then the normal family is a conjugate family for θ.
- In fact, if $\theta \sim \text{Normal}\left(\mu, \frac{1}{\tau_\mu}\right)$ then

$$\theta \mid y \sim \text{Normal}\left(\frac{\tau_y y + \tau_\mu \mu}{\tau_y + \tau_\mu}, \frac{1}{\tau_y + \tau_\mu}\right)$$

.

- The predictive distribution is also normal. In fact,

$$y \sim \text{Normal}\left(\mu, \frac{1}{\tau_y} + \frac{1}{\tau_\mu}\right).$$