Bayesian Formulation

March 12, 2025

Outline

1. Conjugate families

2. Bayesian formulations

Prior distributions for exponential families

- We have seen that beta prior distributions for the probability of success in Bernoulli trials leads to the conjugate families of priors, families to which the posterior after sampling also belongs.
- Suppose $X_1, ..., X_n$ is a sample from the k-parameter exponential family, and, as we always do in the Bayesian context, write $p(x|\theta)$ for $p(x;\theta)$. Then

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} h(x_i) \exp\left(\sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}) \sum_{i=1}^{n} T_j(x_i) - nB(\boldsymbol{\theta})\right)$$

  where $\boldsymbol{\theta}$ is k-dimensional.
- A conjugate exponential family is obtained from it by letting $t_j = \sum_{i=1}^{n} T_j(x_i)$, $j = 1, \ldots, k$ be "parameters" and treating $\boldsymbol{\theta}$ as the 'variables of interest'.

Conjugate exponential family

- We have for $\mathbf{t} = (t_1, \ldots, t_{k+1}) \in \Omega$

$$\omega(\mathbf{t}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(\sum_{j=1}^{k} t_j \eta_j(\boldsymbol{\theta}) - t_{k+1} B(\boldsymbol{\theta}) \right) d\theta_1 \ldots d\theta_k$$

where Ω is the set of $\mathbf{t}$ for which the above is finite.

Theorem

The $(k+1)$-parameter exponential family given by

$$\pi_{\mathbf{t}}(\boldsymbol{\theta}) = \exp\left(\sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}) t_j - t_{k+1} B(\boldsymbol{\theta}) - \log \omega(\mathbf{t}) \right)$$

where $\mathbf{t} \in \Omega$ is a conjugate prior to the exponential family $p(\mathbf{x}|\boldsymbol{\theta})$.

The proof

$$
\begin{aligned}
\pi(\boldsymbol{\theta}|\mathbf{x}) &\propto p(\mathbf{x}|\theta)\pi_{\mathbf{t}}(\boldsymbol{\theta}) \\
&\propto \exp\left(\sum_{j=1}^{k}\eta_j(\boldsymbol{\theta})\left(\sum_{i=1}^{n}T_j(x_i)+t_j\right)-(t_{k+1}+n)B(\boldsymbol{\theta})\right) \\
&\propto \pi_{\mathbf{s}}(\boldsymbol{\theta}),
\end{aligned}
$$

where

$$
\mathbf{s} = (s_1,\ldots,s_{k+1}) = \left(\sum_{i=1}^{n}T_1(x_i)+t_1,\ldots,\sum_{i=1}^{n}T_k(x_i)+t_k,\right)
$$

and $\propto$ indicates that the two sides are proportional functions of θ.

Because two probability densities that are proportional must be equal, $\pi(\theta|\mathbf{x})$ is the member of the exponential family given by the above expression and our assertion follows.

- Note that we have an updating formula in the sense that as data $x_1, \ldots, x_n$ become available, the parameter $\mathbf{t}$ of the prior distribution is updated to $\mathbf{s} = (\mathbf{t}+\mathbf{a})$, where $\mathbf{a} = \left(\sum_{i=1}^{n} T_1(x_i), \ldots, \sum_{i=1}^{n} T_k(x_i)\right)$.
- It is easy to check that the beta distributions are obtained as conjugate to the binomial in this way.

Example

Suppose $X_1, ..., X_n$ is a $N(\theta, \sigma_0^2)$ sample, where σ_0^2 is known and θ is unknown. To choose a prior distribution for θ, we consider the conjugate family of the model. For $n = 1$

$$p(x|\theta) \propto \exp\left(\frac{\theta x}{\sigma^2} - \frac{\theta^2}{2\sigma_0^2}\right).$$

This is a one-parameter exponential family with

$$T_1(x) = x,\ \eta_1(\theta) = \theta/(\sigma_0^2),\ B(\theta) = \theta^2/(2\sigma_0^2).$$

The conjugate two-parameter exponential family has density

$$\pi_t(\theta) = \exp(\theta t_1/\sigma_0^2 - \theta^2 t_2/(2\sigma_0^2) - \log \omega(t_1, t_2).$$

Upon completing the square we obtain

$$\pi_t(\theta) \propto \exp\left(-\frac{t_2}{2\sigma_0^2}(\theta - t_1/t_2)^2\right).$$

Thus, $\pi_t(\theta)$ is defined only for $t_2 > 0$ and all t_1 and is the $N(t_1/t_2, \sigma_0^2/t_2)$ density. Our conjugate family, therefore, consists of all $N(\eta_0, \tau_0^2)$ distributions where η_0 varies freely and τ_0^2 is positive.

Building hierarchical models

Homework *In the hierarchical Bayesian modeling, randomize parameters of the prior distribution and put some hierarchical prior. Since the exponential family of distribution has a conjugate exponential prior, one can utilize this property to build convenient hierarchical priors. Formalize this approach.*

Homework *Use example from the previous slides and build a one-level hierarchical structure of priors. For this model, take (simulate?) some realistic data, and illustrate the difference in the posterior when one uses a non-hierarchical prior vs. hierarchical prior.*

Homework *Provide a similar discussion for an exponential family that is different from normal distributions.*

Outline

Credibility intervals and regions

- In the Bayesian framework we define bounds and intervals, called level $1-\alpha$ credible bounds and intervals, that determine subsets of the parameter space that are assigned probability at least $1-\alpha$ by the posterior distribution of the parameter θ given the data x.
- In the case of a normal prior $\pi(\theta)$ and normal model $p(x|\theta)$, the level $(1-\alpha)$ credible interval is similar to the frequentist interval except it is pulled in the direction μ_0 of the prior mean and it is little narrower.
- However, the interpretations are different: In the frequentist confidence interval, the probability of coverage is computed with the data X random and θ fixed, whereas in the Bayesian credible interval, the probability of coverage is computed with $X = x$ fixed and θ random with probability distribution $\Pi(\theta|X = x)$.
- The Bayesian credibility interval is, in fact, the prediction interval for the unobserved θ, i.e. if one would repeat sampling from the full model for θ and x, then in $1-\alpha 100\%$ cases the Bayesian credibility interval would include θ (that would be changing from sample to sample).

Credibility intervals and regions

- Let $\Pi(\cdot|x)$ denote the posterior probability distribution of θ given $X = x$, then $\underline{\theta}$ and $\overline{\theta}$ are level $1-\alpha$ lower and upper credible bounds for θ if they respectively satisfy
$$\Pi(\underline{\theta} \leq \theta|x) \geq 1-\alpha, \ \Pi(\theta \leq \overline{\theta}|x) \geq 1-\alpha$$
- Turning to Bayesian credible intervals and regions, it is natural to consider the collection of θ that is "most likely" under the distribution $\Pi(\theta|x)$.
$$C_k = \{\theta : \pi(\theta|x) \geq k\}$$
is called a level $1-\alpha$ credible region for θ if $\Pi(C_k|x) \geq 1-\alpha$.

Homework *Show that there can be different from C_k credibility regions. Consider a specific example in which you can provide different sensible credibility intervals from C_k. Which of these credibility regions has the smaller volume? HINT: Consider the case of asymmetric posterior distributions and take symmetric credibility regions vs C_k that have the same credibility level.*

Difficult problem

Homework *Do you think that there is a result about the optimality of C_k. Either give an counter example or give a proof:)*