VectorBiTE Training 2018
Methods Workshop

Introduction to Bayesian Statistics

www.vectorbite.org

Recall: Bayes Theorem

Bayes Theorem allows us to related the conditional probabilities of two events A and B:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

The theory of Bayesian statistics is based on Bayes theorem.

What is Bayesian Statistics?

Bayesian statistics is an approach for finding information in data and making inference with uncertainty based on the probability distribution of parameters given data.

This probability distribution is called posterior and denoted $f(\theta|Y)$, where

$$\overbrace{f(\theta|Y)}^{\text{Posterior}} \propto \overbrace{f(Y|\theta)}^{\text{Likelihood}} \times \overbrace{f(\theta)}^{\text{Prior}}$$

Classical vs Bayesian

The fundamental differences between classical and Bayesian methods is what is fixed and what is random in an analysis

Paradigm	Fixed	Random
Classical	param (θ)	data (Y)
Bayesian	data (Y)	param (θ)

Why/Why Not Bayesian Statistics?

Pros

1. We are naturally Bayesian, in terms of being subjective.
2. It provides interpretable answers, such as the 95% CI.
3. It provides a wide range of possible models, such as hierarchical models and missing data problems.

Cons

1. There is no correct way to choose a prior.
2. It can produce posterior distributions that are heavily influenced by the priors.
3. It often requires high computational cost.

Steps to Making Inference

1. Research question
2. Data collection
3. Model $Y_i \approx f(X_i)$
4. Estimate the parameter in the model with uncertainty
5. Make inference

The difference between Classical and Bayesian lies in step 4: (C) uses maximum likelihood estimate (MLE), and (B) derives a posterior distribution.

Example

1. **RQ:** whether a cancer drug is effective or not?
2. **Data:**

Patient	Effectiveness	Numerical Data
1	N	0
2	N	0
3	Y	1
4	Y	1
5	N	0

3. **Model:**

$$Y_i|\rho \overset{iid}{\sim} f(Y|\rho) = \text{Bern}(\rho)$$

4. **Posterior:** First we need to decide on an appropriate prior for ρ. To look at the domain of ρ we recall that

$$\rho = \Pr(Y_i = 1), \Rightarrow \rho \in [0, 1]$$

Hence, either a Beta or a Uniform distribution can be a prior for ρ. Say we chose $\text{Beta}(a, b)$ with a and b specified as $a = b = 1$ or $a = 1, b = 3$. Thus,

$$f(\rho) = \text{Beta}(a, b).$$

4. **Posterior: continued**

Next, we need to derive the likelihood. We have $Y_i|\rho \overset{iid}{\sim} f(Y|\rho) = \text{Bern}(\rho)$, where $Y_i|\rho$ are assumed to be conditionally independent and identically distributed, thus the likelihood is given by

$$\begin{aligned} f(\tilde{Y}|\rho) &= f(Y_1, Y_2, \cdots Y_5|\rho) \\ &\overset{iid}{=} f(Y_1|\rho) f(Y_2|\rho) \cdots f(Y_5|\rho) \\ &= \prod_{i=1}^{n=5} f(Y_i|\rho). \end{aligned}$$

4. **Posterior: continued**

Now we can use Bayes formula to calculate the posterior distribution as follows

$$\begin{aligned} f(\rho|\tilde{Y}) &\propto \prod_{i=1}^{n=5} f(Y_i|\rho) f(\rho) \\ &\propto \prod_{i=1}^{n=5} \text{Bern}(Y_i|\rho)\text{Beta}(a,b) \\ &\propto \prod_{i=1}^{n=5} \rho^{Y_i}(1-\rho)^{n-Y_i} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \rho^{a-1}(1-\rho)^{b-1} \end{aligned}$$

Does this look like a known distribution? What do we do?

4. **Posterior: continued**

We play the game name that distribution

a. Simplify the best you can
b. $*$, $\div$ by constants
c. Remember what is the RV in the expression
d. Look for Kernels of known distributions

Definition: The Kernel of a probability density function (or mass function) are all components which include the random variable.

eg: Let $Y_i \sim \mathbf{N}(\mu, \sigma^2)$, with random variable μ

$$f(Y_i) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{constant} \underbrace{e^{-\frac{1}{2\sigma^2}(Y_i-\mu)^2}}_{Kernel}$$

4. Posterior: continued

Back to our example:

$$
\begin{aligned}
f(\rho|\tilde{Y}) &\propto \prod_{i=1}^{n=5} \rho^{Y_i}(1-\rho)^{n-Y_i}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\rho^{a-1}(1-\rho)^{b-1} \\
&\propto \rho^{\sum Y_i}(1-\rho)^{n-\sum Y_i}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\rho^{a-1}(1-\rho)^{b-1} \\
&\propto \rho^{\sum Y_i+a-1}(1-\rho)^{n-\sum Y_i+b-1} \\
&\propto \mathbf{Beta}\left(\underbrace{\sum Y_i + a}_{a^*}, \underbrace{n-\sum Y_i + b}_{b^*}\right),
\end{aligned}
$$

where $\mathbf{Beta}(a^*, b^*) = \dfrac{\Gamma(a^*+b^*)}{\Gamma(a^*)\Gamma(b^*)}\rho^{a^*-1}(1-\rho)^{b^*-1}$. Now what do we do with this posterior?

5. **Make inference:** We summarize the posterior to make inference using the following statistics
   a. Expectation $\mathrm{E}(\rho|\tilde{Y})$
   b. Maximum A Posteriori (MAP) (*Mode*)($f(\rho|\tilde{Y})$
   c. Variance $\mathrm{Var}(\rho|\tilde{Y})$
   d. Calculate $\Pr(\rho < .5|\tilde{Y})$
   e. Q Credible Interval $\Pr(\rho \in [L, U]|\tilde{Y}) = Q$, with $L = \frac{1-Q}{2}$, and $U = \frac{1-Q}{2} + Q$
   f. Q Highest Posterior Density Interval (HPDI) $\Pr(\rho \in [L, U]|\tilde{Y}) = Q$, where L and U are chosen so that $U - L$ is the shortest

5. **Inference Statements**

   For a, b, c: Given the data, we predict that the probability that the cancer drug is effective is $\mathrm{E}(\rho|\tilde{Y})$ or $\mathrm{MAP}(\rho|\tilde{Y})$ with variance $\mathrm{Var}(\rho|\tilde{Y})$.

   For d: Given the data, we infer the prob of the cancer drug being effective is low because $\Pr(\rho < .5|\tilde{Y}) = .96$ or high because $\Pr(\rho < .5|\tilde{Y}) = .2$

   For e, f: Given the data, the probability that $\rho \in [L, U]$ is Q.

   Note: **QCI** Describes the behavior of the tail of the posterior distribution and **HPDI** Describes where the mean of the posterior distribution is.

Bayesian Models:

Beta-Bernoulli	Gamma-Poisson
$Y_i \mid \rho \sim \textbf{Bern}(\rho)$	$Y_i \mid \lambda \sim \textbf{Pois}(\rho)$
$\rho \sim \textbf{Beta}(a, b)$	$\lambda \sim \textbf{Gamma}(a, b)$
$\rho \mid \tilde{Y} \sim \textbf{Beta}(a^*, b^*)$	$\lambda \mid \tilde{Y} \sim \textbf{Gamma}(a^*, b^*)$

Question: Are all posteriors in Bayesian analysis the same family as the priors?

Answer: No

Definition of Conjugacy:

Consider a parameter θ, if the derived posterior distribution $f(\theta|\tilde{Y})$ has the same form as the prior $f(\theta)$, then the prior is called conjugate prior distribution

Examples of Conjugate Priors