More on Model Comparison and Assessment

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

02/24/2021

Fractional Bayes Factors

- An alternative to intrinsic *BF* is the *fractional Bayes factor*:

$$BF_b = \frac{p(\mathbf{y}, b \mid M_1)}{p(\mathbf{y}, b \mid M_2)}$$

  where

$$p(\mathbf{y}, b \mid M_i) = \frac{\int p(\mathbf{y} \mid \theta_i, M_i) p(\theta_i \mid M_i)\, d\theta_i}{\int p(\mathbf{y} \mid \theta_i, M_i)^b p(\theta_i \mid M_i)\, d\theta_i}$$

  for $b \in (0, 1)$.

- Often choose $b = 1/n$ if $BF_{1/n}$ is well-defined
- Fractional BF satisfies likelihood principle, intrinsic BF does not.

Fractional Bayes Factors

- Note that

$$p(\mathbf{y}, b \mid M_i) = \int p(\mathbf{y} \mid \theta_i, M_i)^{1-b} p(\theta_i \mid \mathbf{y}, b, M_i)\, d\theta_i$$

  where

$$p(\theta_i \mid \mathbf{y}, b, M_i) \propto p(\mathbf{y} \mid \theta_i, M_i)^b p(\theta_i \mid M_i)$$

- For $\{y_j\}_{j=1}^n$ iid given θ_i,

$$p(\mathbf{y} \mid \theta_i, M_i)^b = \left[\prod_{j=1}^{n} p(Y_j \mid \theta_i, M_i)\right]^b$$

  - So $b = 1/n$ gives the geometric mean for the likelihood of one observation.

Example: traffic accidents (continued)

- Note that

$$p(\lambda \mid \mathbf{y}, 1/n, M_2) = \text{Gamma}(\bar{y} + 1, 1),$$

which gives

$$p(\mathbf{y}, 1/n \mid M_2) = \frac{\Gamma(\sum y_i + 1)}{(\prod y_i!)^{\frac{n-1}{n}} \Gamma(\bar{y} + 1)(n)^{\sum y_i + 1}}$$

Example: traffic accidents (continued)

- Similarly,

$$p(\mathbf{y}, 1/n \mid M_1) = \frac{3^{\bar{y}+3}\Gamma(\sum y_i + 3)}{(\prod y_i!)^{\frac{n-1}{n}}\Gamma(\bar{y}+3)(2+n)^{\sum y_i + 3}}$$

- For 5 weeks data, the fractional BF for M_1 over M_2 is

$$BF_{1/5} = 1.28$$

http://www.ericfrazerlock.com/Model_Comparison_Rcode2.r

Cross-validated likelihood

- The *conditional predictive distribution* for y_i is

$$p(y_i \mid \mathbf{y}_{(i)}) = \int p(y_i|\theta, \mathbf{y}_{(i)}) p(\theta \mid \mathbf{y}_{(i)})\, d\theta$$

  - $\mathbf{y}_{(i)} = (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$
  - Low $p(y_i \mid \mathbf{y}_{(i)})$ indicates the model is a poor fit for y_i

- The product

$$\prod_{i=1}^{n} p(y_i \mid \mathbf{y}_{(i)})$$

  is sometimes called the *pseudo marginal likelihood*.

  - Higher values indicate better overall model fit

Bayesian residuals

- Use conditional predictive distribution to compute Bayesian *residuals*:

$$r_i' = y_i - E(Y_i \mid \mathbf{y}_{(i)}),$$

- Standardized residual:

$$d_i' = \frac{y_i - E(Y_i \mid \mathbf{y}_{(i)})}{\sqrt{Var(Y_i \mid \mathbf{y}_{(i)})}},$$

- Can be used to detect systematic departures from model assumptions

Example: IQ (continued)

- Recall: IQs have Normal$(100, 225)$ population distribution.
- An IQ test has assumed error variance 64.
- A sample of 100 randomly selected participants are each tested 5 times
  - Let y_{ij} be the score for the jth trial of the ith participant
  - Then μ_i be the IQ of subject i
  - $y_{ij} \sim \text{Normal}(\mu_i, 64)$

Example: IQ (continued)

- Let $\mathbf{y}_{(ij)}$ represent all scores but score j for subject i
- Let $\bar{y}_{i(j)}$ represent the mean for all scores but j for subject i

$$p(\mu_i \mid \mathbf{y}_{(ij)}) = \text{Normal}(6.64 + 0.934\bar{y}_{i(j)}, 14.94)$$

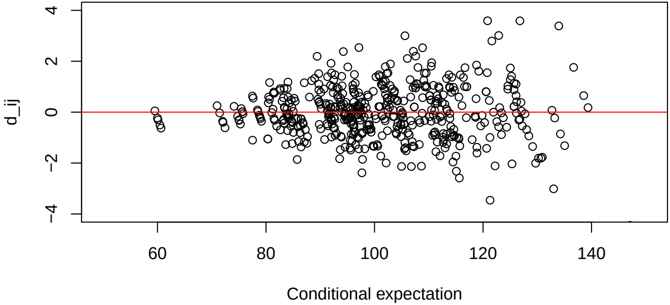
and

$$p(y_{ij} \mid \mathbf{y}_{(ij)}) = \text{Normal}(6.64 + 0.934\bar{y}_{i(j)}, 78.94)$$

- The standardized residuals are

$$d'_{ij} = \frac{y_{ij} - 6.64 - 0.934\bar{y}_{i(j)}}{8.88}.$$

- Plot of d'_{ij} vs. $E(Y_{ij} \mid \mathbf{y}_{(ij)})$:
  - Is the model appropriate?



- http://www.ericfrazerlock.com/More_on_model_comparison_and_assessment_Rcode1.r

Bayesian p-values

- Recall: the frequentist p-value based on observed statistic $T(\mathbf{y})$:
$$P(T(\mathbf{Y}) \geq T(\mathbf{y}) \mid H_0)$$
  - $\mathbf{Y}$ and $\mathbf{y}$ are iid given θ
- The *posterior predictive p-value* under model
$$M_0 : \mathbf{y} \sim p(\mathbf{y} \mid \theta), \ \theta \sim p(\theta)$$
for statistic $T(\mathbf{y}, \theta)$ is
$$\begin{aligned} p_T &= P(T(\mathbf{Y}, \theta) \geq T(\mathbf{y}, \theta) \mid M_0, \mathbf{y}) \\ &= \int P(T(\mathbf{Y}, \theta) \geq T(\mathbf{y}, \theta) \mid \theta) p(\theta \mid \mathbf{y}, M_0) \, d\theta \end{aligned}$$
  - Sometimes called a "Bayesian p-value"
  - More generally, replace '$\geq$' with 'more extreme than'.

Bayesian p-values

- T can be a function of data ($\mathbf{y}$) and parameters (θ)
  - E.g., the goodness-of-fit statistic

$$T(\mathbf{y}, \theta) = \sum_{i=1}^{n} \frac{[y_i - E(Y_i \mid \theta)]^2}{\mathrm{Var}(Y_i \mid \theta)}$$

- Low p_T can indicate problems with M_0
  - The observed data are not plausible under the predictive model
- Little decision-theoretic justification
- Not recommended as sole basis for comparing models

Example: parental remorse

- Interested in θ: proportion of parents who regret having children [1]
- Prior for θ: $\text{Beta}(1, 9)$.
- In a small survey of $n = 10$ longtime parents, $y = 9$ say they regret the choice
- Posterior for θ: $\text{Beta}(10, 10)$
- What is the posterior predictive p-value?

[1] Motivated by a newspaper survey:
https://userpages.umbc.edu/~nmiller/POLI300/stat353annlanders.pdf

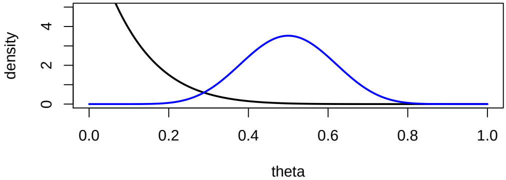
Beta-Binomial marginal

- If $y \sim \text{Binomial}(n, \theta)$ and $p(\theta) = \text{Beta}(a, b)$,

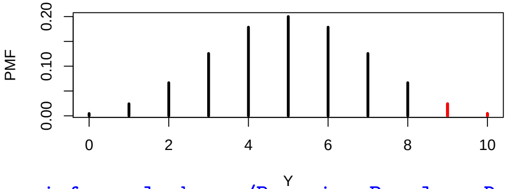
$$P(y = k) = \binom{n}{k} \frac{B(a + k, b + n - k)}{B(a, b)},$$

where B is the Beta function.

- **Prior** and posterior densities for θ:



- Posterior predictive pmf $P(Y \mid y = 9)$:



http://www.ericfrazerlock.com/Bayesian_P-values_Rcode1.r

Example: parental remorse

- The posterior predictive p-value is

$$P(Y \geq 9 \mid y = 9) = 0.029.$$

- Suggests that our prior was overly strong / should have been less optimistic about parental remorse.

Bayesian p-values

- Posterior predictive p-values have been criticized for "using the data twice"
  - For reference distribution $p(\theta \mid \mathbf{y})$
  - For observed statistic $T(\mathbf{y}, \theta)$.
- The practical implication of this has been debated:
  - See https://xianblog.wordpress.com/2014/02/04/posterior-predictive-p-values/ and related discussion in comments.
- Alternatively, for training data $\mathbf{y}_1$ and test data $\mathbf{y}_2$, compute

$$p'_T = P(T(\mathbf{Y_2}, \theta) \geq T(\mathbf{y_2}, \theta) \mid M_0, \mathbf{y_1})$$