Bayesian Inference for Surveys

Roderick Little and Trivellore Raghunathan

Model Checking, Comparison and Averaging

UNIVERSITY OF MICHIGAN

Model Checking

- Model checking is an important step in a Bayesian analysis
- Two approaches
  - Generate new data from the marginal distribution of observables under the model and compare with the observed data. Prior-predictive check
  - Generate new data from the posterior predictive distribution and compare with the observed data. Posterior-predictive check

$y =$ Observed data

$\theta =$ Parameters in the model

$f(y \mid \theta)$: Conditional distribution of observables given the parameter θ

$\pi(\theta)$=Prior density

Prior Predictive Check

- Prior predictive distribution (marginal distribution of observables)

$$f(y_{\text{new}}) = \int f(y_{\text{new}} \mid \theta)\pi(\theta)d\theta$$

- Generate new data from the prior-predictive distribution and compare with the observed data y_{inc}

$$\theta^* = \text{draw from } \pi(\theta)$$
$$y_{\text{new}} = \text{ draw from } f(y \mid \theta^*)$$

Not possible to implement if the prior distribution is improper

Posterior Predictive Check

- Posterior predictive distribution

$$f(y_{\text{new}} \mid y_{\text{inc}}) = \int f(y_{\text{new}} \mid \theta, y_{\text{inc}}) \pi(\theta \mid y_{\text{inc}}) d\theta$$
$$= \int f(y_{\text{new}} \mid \theta) \pi(\theta \mid y_{\text{inc}}) d\theta$$

- Generate new data from the posterior-predictive distribution and compare with the observed data y_{inc}

$$\theta^* = \text{draw from } \pi(\theta \mid y_{\text{inc}})$$
$$y_{\text{new}} = \text{ draw from } f(y \mid \theta^*)$$

If the prior distribution is diffuse generate new data from "likelihood" and compare with the observed data

Modification of Posterior Predictive check

- Use a subsample to construct posterior distribution

This is useful when a survey is conducted by drawing replicates

Using half the replicates for model fitting and the other half for model checking

$$y_{inc} = (y_{inc}^{(1)}, y_{inc}^{(2)})$$

$$\theta \sim \pi(\theta \mid y_{inc}^{(1)})$$

$$y_{new}^{(2)} \sim f(y \mid \theta)$$

$$Compare: y_{new}^{(2)}, y_{inc}^{(2)}$$

Example 1: Prior and Posterior Predictive Distributions

Objective: To check the independence in a sequence of Bernoulli trials

Model:

$y_i \sim \text{iid Bin}(1, \theta)$

$\theta \sim \text{Unif}(0,1)$

Observed sequence:

$y_{\text{inc}} = \{1,1,$

$0,0,0,0,0,$

$1,1,1,1,1,$

$0,0,0,0,0,0,0,0\}$

$T(y_{\text{inc}}) = 3$

Prior Predictive

$\theta^{\text{rep}} \sim \text{Unif}(0,1)$

$y_i^{\text{rep}} \sim \text{Bin}(1, \theta^{\text{rep}})$

$i = 1, 2, \ldots, 20$

$T(y^{\text{rep}})$

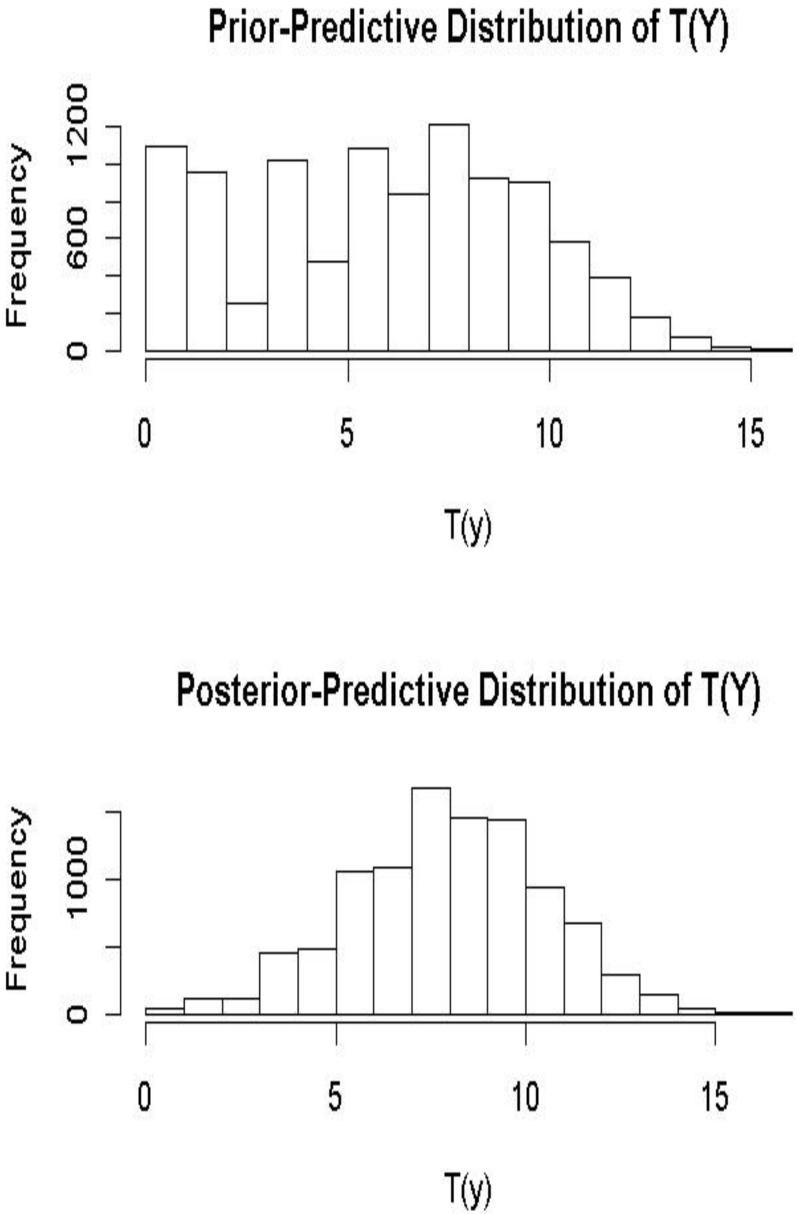
Posterior predictive

$\theta^{\text{rep}} \sim \text{Beta}(8,14)$

$y_j^{\text{rep}} \sim \text{Bin}(1, \theta^{\text{rep}})$

$j = 1, 2, \ldots, 20$

$T(y^{\text{rep}})$



Posterior Predictive frequency	Prior Predictive Prior frequency	T(y)
31	994	0
7	99	1
120	955	2
111	253	3
459	1022	4
477	472	5
1064	1090	6
1084	838	7
1671	1210	8
1457	920	9
1435	900	10
934	579	11
669	394	12
296	173	13
137	74	14
33	21	15
12	6	16
3	0	17

Comparing new and observed data

- Develop "discrepancy measure" $T(y)$ or $T(y, q)$. Note that the discrepancy measure can depend upon the parameter q.
- Compare $T(y_{\text{inc}})$ with $T(y_{\text{new}})$ or

$$T(y_{\text{inc}},\theta^*) \text{ with } T(y_{\text{new}},\theta^*)$$

- Discrepancy measures depends upon the problem
- General goodness of fit measure:

$$T(y,\theta) = \sum_{i=1}^{n} \frac{(y_i - E(y_i \mid \theta))^2}{\text{var}(y_i \mid \theta)}$$

Model Checking: Quality Control Example

- In the quality control example with beta-binomial model, generating replicates is fairly straightforward given the draws of (a, b, q) (independent binomial samples)
- The discrepancy measure used:

$$T = (\mathrm{Min}(y_i - k_i\theta_i), \mathrm{Max}(y_i - k_i\theta_i))$$

- Computed fraction of times, out of 2500 replicates, the observed minimum was less than the replicate minimum and/or the replicate maximum was greater than the observed maximum (Bayesian p-value)
- 2247 replicates did not capture the observed spread (p-value=0.1012)

Remarks

- Model checking is an important step. The posterior predictive check may give indication about lack of fit
- Prior knowledge is also very important in making judicious choice of the models.
- It may be difficult to pin down one model that may be satisfactory in every aspect
- It is better to consider a continuum of models and perform sensitivity analysis by inspecting inferences under these models
- Generally need some prior information to handle extrapolation outside the range of observed data

Model Comparisons 1

- Model checking requires thought about measuring deviations between observed data and what one would expect under the model.
- Many such choices have to be investigated to ensure that the model is an adequate abstraction for glean useful information about the population based on the observed values and express the associated uncertainty.
- An alternative mechanism: To compute the posterior probability for the model being "correct".
  - Bayes factor

Model Comparisons 2

- K models under consideration

$$M_j, j = 1, 2, \ldots, K$$

- Prior probabilities for model j being correct

$$\{p_j, j = 1, 2, \ldots, K\}; \sum_{j=1}^{K} p_j = 1$$

- Model $\quad M_j : [f_j(y \mid \theta_j), \pi_j(\theta_j)]$
- Posterior probability for model j being correct

$$\Pr(M_j \mid y) = \frac{p_j L_j(M_j \mid y)}{\sum_{j=1}^{K} p_j L_j(M_j \mid y)},$$

$$L(M_j \mid y) = \int f_j(y \mid \theta_j) \pi_j(\theta_j) d\theta_j$$

Model Comparisons 3

Posterior odds = Prior odds x Bayes Factor:

$$\frac{\Pr(M_j \mid y)}{\Pr(M_l \mid y)} = \frac{p_j}{p_l} \times \frac{L(M_j \mid y)}{L(M_l \mid y)}$$

Bayes factor

$\log(\text{Bayes Factor}) \approx$

$\log(L(\hat{\theta}_j \mid y, M_j)) - \log(L(\hat{\theta}_l \mid y, M_l)) + (k_j - k_l)\log(n)/2$

$L(\hat{\theta} \mid y, M) =$ likelihood evaluated at

ML estimate under model M

k = number of parameters

Model Comparisons 4

- Bayes Information Criterion

$$\mathrm{BIC}(M) = \log(L(\hat{\theta} \mid y, M)) - k\log(n)/2$$

$$\text{Bayes factor}(M_j \ vs \ M_l) \approx \mathrm{BIC}(M_j) - \mathrm{BIC}(M_l)$$

- The approximation has been derived for nested models
- The Bayes factor, however, can be applied for non-nested models
- The marginal distribution has to be proper. This is not guaranteed when a non-informative prior is used for the parameter

Model Averaging

- Original model: $\{f(y|\theta), \pi(\theta)\}$
- Consider an expanded class of model

$$\{f(y|\theta,\phi), \pi(\theta|\phi), p(\phi)\}$$

where the original model is a member of this expanded class

- Inference from

$$\pi(\theta|y) = \int \pi(\theta,\phi|y)d\phi$$
$$\propto \int f(y|\theta,\phi)\pi(\theta|\phi)p(\phi)d\phi$$

Example

- Original model

$$y \sim N(\mu, \sigma^2), \pi(\mu, \sigma^2) \propto \sigma^{-2}$$

- Expanded model

$$y \sim t_\nu(\mu, \sigma^2), \pi(\mu, \sigma^2, \nu) \propto \sigma^{-2}\nu^{-2}$$

$$t_\infty(\mu, \sigma^2) \equiv N(\mu, \sigma^2)$$

$$t_1(\mu, \sigma^2) \equiv \text{Cauchy}$$

- Prior distribution gives more weight towards normal range

Example

- Original model

$$y \sim N(\mu,\sigma^2),\ \pi(\mu,\sigma^2) \propto \sigma^{-2}$$

- Expanded model

$$\alpha N(\mu,\sigma^2) + (1-\alpha) g(\mu+\delta, \rho^2\sigma^2)$$

where g is a member of some location-scale family of distribution