Model checking

SeokHoon Park

September 29, 2020

Seoul National University

Table of Contents

Table of Contents

The places of model checking in applied Bayesian statistics

- Once we have accomplished constructing a probability model and computing the posterior distribution of all estimands,
  We should assess the fit of the model to the data and to out substantive knowledge.

- It is difficult to include in a probability distribution all of one's knowledge about a problem
  So it is wise to investigate what aspects of reality are not captured by the model

Sensitivity analysis and model improvment

- The sensitivty analysis:
  How much do posterior inferences changes when other reasonable probability models are used in place of the present model?
- In theory, both model checking and sensitivity analysis can be incorporated into the usual prior-to-posterior analysis.
  Under this perpective, model checking is done by setting up a comprehensive joint distribution.

Sensitivity analysis and model improvment

- In practice, however, setting up such a super-model to include all possibilities and all substantive knowledge is both conceptually impossible and computationally infeasible in all but simplest problems.

Table of Contents

External validation

- We can check a model by external validation using the model to make predictions about future data, and then collecting those data and comparing to their predictions.

- In this chapter and the next, we discuss methods which can approximate external validation using the data we already have.

Table of Contents

Posterior predictive checking

- Our basic technique for checking the fit of a model to data is to draw simulated values from the joint posterior predictive distrbution of replicated data and compare these samples to the observed data.

Notation for replications

- y =observed data
- θ =vector of parameters
- y^{rep} =the replicated data
- $\tilde{y}$ =the future observable value

Test quantities

- Test quantity is a scalar summary of parameters and data that is used as standard when comparing data to predictive sumulations := $T(y)$ *or* $T(y, \theta)$

- The choice of test quantity requires careful consideration of the type of inferences required for the problem being considered

Posterior predictive p-values

$$\begin{aligned}
p_B &= Pr(T(y^{rep}, \theta) \geq T(y, \theta) | y) \\
&= \int \int I_{(T(y^{rep}, \theta) \geq T(y, \theta))} p(y^{rep}, \theta | y) dy^{rep} d\theta \\
&= \int \int I_{(T(y^{rep}, \theta) \geq T(y, \theta))} \frac{p(y^{rep}, \theta, y)}{p(y)} dy^{rep} d\theta \\
&= \int \int I_{(T(y^{rep}, \theta) \geq T(y, \theta))} \frac{p(y^{rep}, \theta, y)}{p(y, \theta)} \frac{p(y, \theta)}{p(y)} dy^{rep} d\theta \\
&= \int \int I_{(T(y^{rep}, \theta) \geq T(y, \theta))} p(y^{rep} | \theta, y) p(\theta | y) dy^{rep} d\theta
\end{aligned}$$

Posterior predictive p-values

$$\begin{aligned} p(y^{rep}|\theta, y) &= \frac{p(y^{rep}, \theta, y)}{p(\theta, y)} \\ &= \frac{p(y^{rep}, y|\theta)}{p(y|\theta)} \\ &= \frac{p(y^{rep}|\theta)p(y|\theta)}{p(y|\theta)} \\ &= p(y^{rep}|\theta) \end{aligned}$$

$$So, P_B = \int \int I_{(T(y^{rep},\theta) \geq T(y,\theta))} p(y^{rep}|\theta)p(\theta|y)dy^{rep}d\theta$$

Posterior predictive checking in practice

1. Compute the posterior predictive distribution using simulation.
2. If we simulated N times from the posteiror density of θ , we draw one y^{rep} from the predictive distribution for each simulated θ
3. Comparison between the realized test quantities $T(y, \theta^s)$ and the predictive test quantities $T(y^{reps}, \theta^s) \quad , s = 1, \cdots N$

Estimated p-value

- The estimated p-value is just the proportion of these N simulations for $T(y^{rep}, \theta^s) \geq T(y, \theta^s) \quad s = 1, \cdots N$

Example: Checking the assumption of independence in binomial trials

- $y_1, \cdots, y_n \sim^{iid} Bernoulli(p)$
- $p(\theta|y) \propto \theta^{\sum y}(1-\theta)^{n-\sum y}$
- The observed data : 1,1,0,0,0,0,0,1,1,1,1,1,0,0,0,0,0,0,0,0
- Test quantity T= of switches between 0 and 1 in the sequence
- # of simuation =10000
- T(y)=3
- # of $T(y^{rep}, \theta^s) \geq T(y, \theta^s) = 9838$
- Estimated p-value= 0.9838

Interpreting posterior predictive p-values

- $p_B = Pr(T(y^{rep}, \theta) \geq T(y, \theta)|y)$
- p-value$\approx 0 \quad or \quad 1$: model cannot be expected to capture this aspect of the data
- p-value≈ 0.5 : model can be expected to capture this aspect of the data

Marginal posterior p-value

- $p_i = Pr(T(y_i^{rep}) \leq T(y_i)|y) \quad , y = [y_1, \cdots, y_n]$
- If y_i is scalar and continuous, a natural quantity is $T(y_i) = y_i$, $\mathrm{p}_i = Pr(y_i^{rep} \leq y_i|y)$
- For ordered discrete data, we can compute a "mid" p-value $p_i = Pr(y_i^{rep} < y_i|y) + \frac{1}{2}Pr(y_i^{rep} = yi|y)$
- we will see different behavior than from the joint checks

Marginal posterior p-value

- marginal posterior p-value 0 *or* 1
  : Data is over-dispersed compared to the model

- marginal posterior p-value ≈ 0.5
  : Data is under-dispersed compared to the model

Predictive checks with cross-validation

- $p_i = Pr(y_i^{rep} \leq y_i | y_{-i}), \quad y_{-i} =$ all other data except y_i
- We will address cross-validation in the 7 chapter

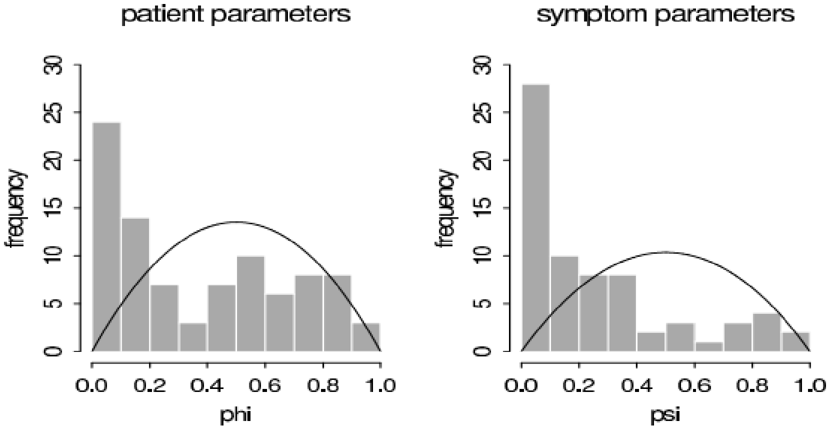
Table of Contents

Graphical posterior predictive checks

Example

- $\phi_1, \cdots, \phi_{90}, \psi_1, \cdots, \psi_{69} \sim^{iid} Beta(2, 2)$
- The full Bayesan model fitted and yields posterior simulations for all these parameters.

Graphical posterior predictive checks



Graphical posterior predictive checks

- The histogram and prior distribution does not fit
- Replace the offending $Beta(2, 2)$ prior distribution by mixture of two beta distribution.

$$p(\phi_j) = 0.5Beta(1, 6) + 0.5Beta(1, 1)$$
$$p(\psi_j) = 0.5Beta(1, 16) + 0.5Beta(1, 1)$$

Graphical posterior predictive checks

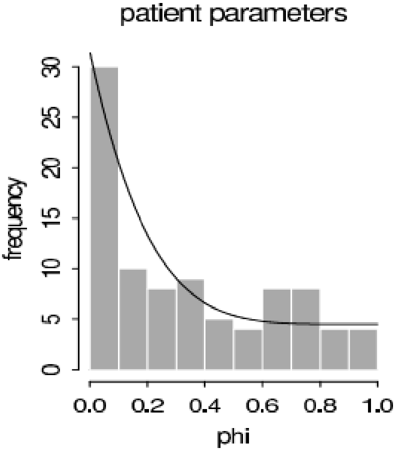


Table of Contents

Model checking for the educational testing example

- We illustrate the ideas of this chapter with the example from Section 5.5

1. Assumptions of the model
2. Comparing posterior inferneces to substantive knowledge
3. Posterior predictive checking
4. Sensitivity analysis

Model checking for the educational testing example

- **Assumption of the model**

  1. normality of the estimates y_j given θ_j and σ_j ,where σ_j are assumed known
  2. exchangeability of the prior distribution of the $\theta_j' s$
  3. normality of the prior distribution of each θ_j given μ and τ
  4. uniformity of the hyperprior distribution of (μ, τ)

Model checking for the educational testing example

- **Comparing posterior inferences**

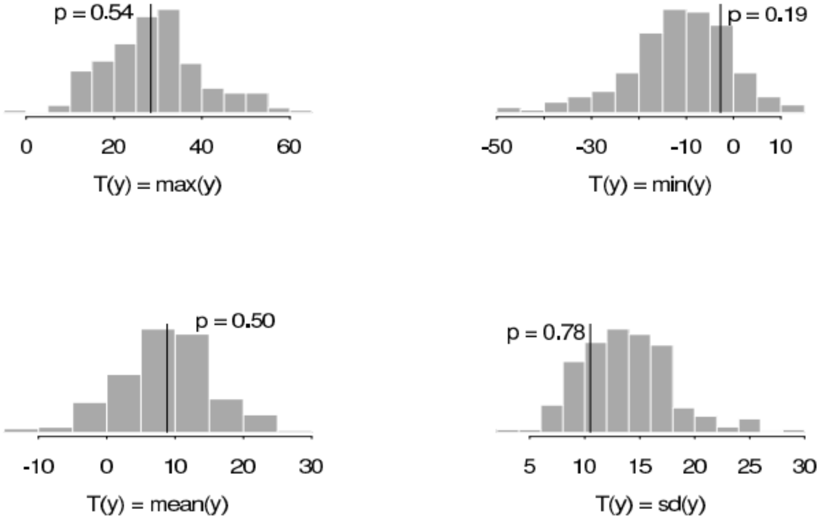
  1. Compare the posterior distribution of effects to our knowledge of educational testing
  2. Simulate the posterior predictive distribution of a hypothetical replication of the experiments

Model checking for the educational testing example

- **Posterior predictive checking**

1. $T_1 = max_j(y_i)$, $T_2 = min_j(y_j)$, $T_3 = mean(y_j)$, $T_4 = sd(y_j)$
2. Approximate the posterior predictive distribuiton of each statistic by the histogram of the values from the 200 simulations of the parameters and predictive data.
3. Compare each distribution to the observed value of the test quantity

Model checking for the educational testing example



Model checking for the educational testing example

- **Sensitiy analysis**
  - Other reasonable models might provide just as good a fit but lead to different conclusions
  - The uniform prior distribution for τ
  - The normal population distribution for the school effects
  - The normal likelihood

References i

A. G. et.al., *Bayesian Data Analysis 3rd*. CRC Press, 2013, ISBN: 9781439840955.