ADVANCED BAYESIAN MODELING

Posterior Predictive Checking in General

Replicate Data

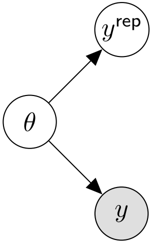
Consider posterior predictive checking more formally ...

y = observed data set $\qquad$ y^{rep} = hypothetical replicate data set

Under model, y and y^{rep} have the same distribution given parameter θ, but are otherwise independent. (The same value of θ defines their common distribution.)

Since θ is unobserved, work with the posterior predictive distribution, having density

$$p(y^{\text{rep}} \mid y) \;=\; \int p(y^{\text{rep}} \mid \theta)\, p(\theta \mid y)\, d\theta$$



Structure implies, for example,

$$p(y^{\mathsf{rep}} \mid \theta, y) \quad = \quad p(y^{\mathsf{rep}} \mid \theta)$$

We can simulate y^{rep}, but it exists only to represent the posterior predictive distribution of the data, so is regarded as unobserved.

Test Quantities

Model and data will be compared using scalar **test quantity**

$$T(y, \theta)$$

also called a **discrepancy measure** when larger values indicate greater disparity.

If it doesn't depend on θ, it is a **test statistic** $T(y)$.

Allowing T to depend on θ may allow more direct definitions of discrepancies.

In that case, T cannot be computed on the data, but can still be used in expressions that are averaged over the posterior.

Classical p-values

When discrepancy measure T is a test statistic, the classical p-value is

$$p_C \quad = \quad \Pr\big(T(y^{\text{rep}}) \geq T(y) \mid \theta\big)$$

Small p-values indicate evidence against the model – the data have a larger T than expected.

To evaluate, must either

- Choose T so that the probability does not depend on unknown θ, or
- Substitute a value for θ (null or estimate)

Posterior Predictive p-values

For general discrepancy measure T, the Bayesian **posterior predictive p-value** is

$$p_B \;=\; \Pr\big(T(y^{\text{rep}},\theta) \;\geq\; T(y,\theta) \mid y\big)$$

The probability is over the joint posterior (predictive) distribution of (y^{rep},θ).

In practice, approximate p_B by jointly simulating (y^{rep},θ) (given y) many times and finding how often

$$T(y^{\text{rep}},\theta) \;\geq\; T(y,\theta)$$

Unlike classical p-value, posterior predictive p-value depends on prior.

Advantages of posterior predictive p-value:

- Allows T to depend on θ.
- Doesn't require knowing exact sampling distribution of T (i.e. knowing θ).
- Also assesses prior, not just sampling distribution.

Example: Flint Data (continued)

Consider alternative test quantity for asymmetry (not a statistic):

$$T(y, \mu) \quad = \quad |\hat{q}_{0.9} - \mu| \; - \; |\hat{q}_{0.1} - \mu|$$

```
> Tysim <- numeric(1000)
> Tyrepsim <- numeric(1000)
> for(s in 1:1000){
+   Tysim[s] <- abs(quantile(log(Flintdata$FirstDraw),0.9) - post.mu.sim[s]) -
+               abs(quantile(log(Flintdata$FirstDraw),0.1) - post.mu.sim[s])
+   Tyrepsim[s] <- abs(quantile(yreps[s,],0.9) - post.mu.sim[s]) -
+                  abs(quantile(yreps[s,],0.1) - post.mu.sim[s])
+ }
```

The posterior predictive p-value:

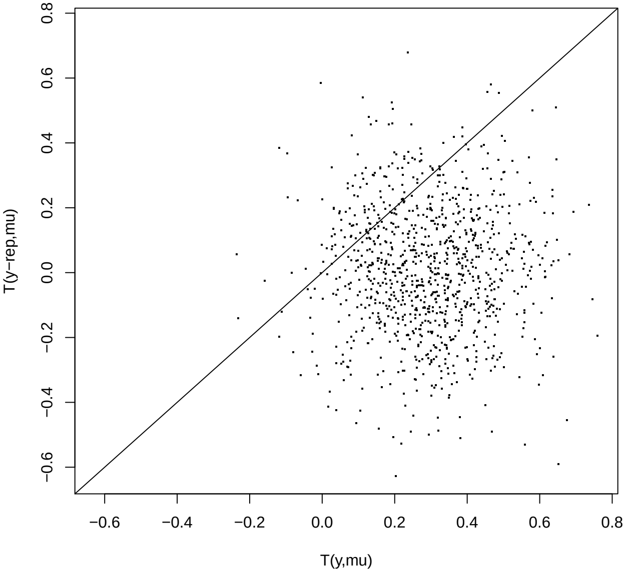
```
> mean(abs(Tyrepsim) >= abs(Tysim))
[1] 0.222
```

Note: The discrepancy measure is actually $|T(y, \mu)|$.

Unlike earlier analysis, no asymmetry is detected.

Occasionally useful to plot $T(y^{\text{rep}}, \theta)$ versus $T(y, \theta)$, with a line to indicate where they are equal:

```
> plot(Tysim, Tyrepsim, pch=".", cex=2,
+       xlim=c(min(Tysim, Tyrepsim), max(Tysim, Tyrepsim)),
+       ylim=c(min(Tysim, Tyrepsim), max(Tysim, Tyrepsim)),
+       xlab="T(y,mu)", ylab="T(y-rep,mu)")
> abline(a=0,b=1)
```



0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
T(y−rep,mu)
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8
T(y,mu)

Comments:

- BDA3 suggests using two-sided test quantities, which indicate discrepancy if either too large or too small.

  In that case, evidence against the model would be indicated if p_B is either too close to 0 or too close to 1.

- Practical importance of a discrepancy should be considered, not just significance level.

  Even a wrong model can be useful.