Posterior Predictive Distribution in Regression

Example 3: In the regression setting, we have shown that the posterior predictive distribution for a new response vector $\mathbf{y}^*$ is multivariate-t.

- To check model fit, we can generate samples from the posterior predictive distribution (letting $\mathbf{X}^* =$ the observed sample $\mathbf{X}$) and plot the values against the y-values from the original sample.
- If an observed y_i falls far from the center of the posterior predictive distribution, this i-th observation is an outlier.
- If this occurs for many y-values, we would doubt the adequacy of the model.
- See `R` example (small automobile data set).

Posterior Predictive Distribution in Regression

- We can also make predictions and "prediction intervals" for new responses with specified predictor values.
- For example, consider a new observation with predictor variable values in the vector $\mathbf{x}^* = (1, x_1^*, x_2^*, \ldots, x_{k-1}^*)$ (or the predictor values for several new observations could be contained in the matrix $\mathbf{X}^*$).
- We can generate the posterior predictive distribution with $\mathbf{X}^*$ and compute the posterior median (for a point prediction) or posterior quantiles (for a prediction interval).
- See `R` example.

CHAPTER 7 SLIDES START HERE

Issues with Classical Hypothesis Testing

- Recall that classical hypothesis testing emphasizes the **p-value**: The probability (under H_0) that a test statistic would take a value as (or more) favorable to H_a as the observed value of this test statistic.
- For example, given iid data $\mathbf{x} = x_1, \ldots, x_n$ from $f(x|\theta)$, where $-\infty < \theta < \infty$, we might test $H_0 : \theta \leq 0$ vs. $H_a : \theta > 0$ using some test statistic $T(\mathbf{X})$ (a function of the data).
- Then if we calculated $T(\mathbf{x}) = T^*$ for our observed data $\mathbf{x}$, the p-value would be:

$$\text{p-value} = P[T(\mathbf{X}) \geq T^* | \theta = 0] = \int_{T^*}^{\infty} f_T(t|\theta = 0)\, dt$$

where $f_T(t|\theta)$ is the distribution (density) of $T(\mathbf{X})$.

Issues with Classical Hypothesis Testing

- This p-value is an average over T values (and thus sample values) that **have not occurred** and are **unlikely to occur**.
- Since the inference is based on "hypothetical" data rather than **only** the **observed** data, it violates the Likelihood Principle.
- Also, the idea of conducting many repeated tests that motivate "Type I error" and "Type II error" probabilities is not sensible in situations where our study is not repeatable.

The Bayesian Approach

- A simple approach to testing finds the posterior probabilities that θ falls in the null and alternative regions.
- We first consider one-sided tests about θ of the form:

$$H_0 : \theta \leq c \quad \text{vs.} \quad H_a : \theta > c$$

  for some constant c, where $-\infty < \theta < \infty$.
- We may specify prior probabilities for θ such that

$$p_0 = P[-\infty < \theta \leq c] = P[\theta \in \Theta_0]$$

  and

$$p_1 = 1 - p_0 = P[c < \theta < \infty] = P[\theta \notin \Theta_0]$$

  where Θ_0 is the set of θ-values such that H_0 is true.

The Bayesian Approach

- Then the **posterior probability** that H_0 is true is:

$$P[\theta \in \Theta_0|\mathbf{x}] = \int_{-\infty}^{c} p(\theta|\mathbf{x})\,\mathrm{d}\theta$$

$$= \frac{\int_{-\infty}^{c} p(\mathbf{x}|\theta)p_0\,\mathrm{d}\theta}{\int_{-\infty}^{c} p(\mathbf{x}|\theta)p_0\,\mathrm{d}\theta + \int_{c}^{\infty} p(\mathbf{x}|\theta)p_1\,\mathrm{d}\theta}$$

by Bayes' Law (note the denominator is the marginal distribution of $\mathbf{X}$).

The Bayesian Approach

- Commonly, we might choose an uninformative prior specification in which $p_0 = p_1 = 1/2$, in which case $P[\theta \in \Theta_0|\mathbf{x}]$ simplifies to

$$\frac{\int_{-\infty}^{c} p(\mathbf{x}|\theta) p_0 \, \mathrm{d}\theta}{\int_{-\infty}^{\infty} p(\mathbf{x}|\theta) p_0 \, \mathrm{d}\theta} = \frac{\int_{-\infty}^{c} p(\mathbf{x}|\theta) \, \mathrm{d}\theta}{\int_{-\infty}^{\infty} p(\mathbf{x}|\theta) \, \mathrm{d}\theta}$$

Hypothesis Testing Example

- **Example 1** (Coal mining strike data): Let Y = number of strikes in a sequence of strikes before the cessation of the series.
- Gill lists $Y_1, \dots, Y_{11}$ for 11 such sequences in France.
- The Poisson model would be natural, but for these data, the variance greatly exceeds the mean.
- We choose a geometric(θ) model

$$f(y|\theta) = \theta(1-\theta)^y$$

  where θ is the probability of cessation of the strike sequence, and y_i= number of strikes before cessation.
- **Exercise:** Show that the Jeffreys prior for θ is $p(\theta) = \theta^{-1}(1-\theta)^{-1/2}$. We will use this as our prior.

Hypothesis Testing Example

- So the posterior is:

$$\begin{aligned}\pi(\theta|\mathbf{y}) &\propto L(\theta|\mathbf{y})p(\theta) \\ &= \theta^{n}(1-\theta)^{\sum y_i}\theta^{-1}(1-\theta)^{-1/2} \\ &= \theta^{n-1}(1-\theta)^{\sum y_i - 1/2}\end{aligned}$$

  which is a beta$(n, \sum y_i + 1/2)$ distribution.
- We will test $H_0 : \theta \leq 0.05$ vs. $H_a : \theta > 0.05$.
- Then $P[\theta \leq 0.05|\mathbf{y}] = \int_0^{0.05} \pi(\theta|\mathbf{y})\, \mathrm{d}\theta$, which is the area to the left of 0.05 in the beta$(n, \sum y_i + 1/2)$ density.
- This can be found directly (or via Monte Carlo methods).
- See `R` example with coal mining strike data.