Chapter 1.5

The posterior prediction distribution

Bayesian prediction

- Often the objective is to predict a future event
- Example: Last spring we planted $n = 10$ seedlings and $Y = 2$ survived the winter, if we plant n again this year what is the probability at least one will survive the winter?
- Let Y^* be the predicted value and θ be the true survival probability
- If the parameters were known then we would predict

$$Y^*|\theta \sim \text{Binomial}(10, \theta)$$

  and thus $\text{Prob}(Y > 0) = 1 - (1 - \theta)^{10}$

- Of course, if we knew the parameters we would be doing probability and not statistics

Two types of uncertainty

- **Sampling**: Even if we knew θ, we could not predict Y exactly because there is inherent randomness in which plants survive. This is quantified using the likelihood distribution, $Y|\theta$.

- **Parametric**: We can never know θ exactly. This uncertainty is quantified by its prior and posterior distributions.

- How to resolve each type of uncertainty?

Plug-in approach

- One approach to prediction is the "plug-in" approach
- That is, if $\hat{\theta}$ is an estimate, then use prediction distribution

$$Y^* \sim f(Y|\hat{\theta})$$

- Example: $\hat{\theta} = 2/10$ then predict

$$\text{Prob}(Y > 0) = 1 - (1 - 0.2)^{10}$$

- If θ has small uncertainty this is fine
- Otherwise, this underestimates uncertainty in Y^*

The PPD

- For the sake of prediction, the parameters are not of interest
- They are vehicles by which the data inform about the predictive model
- The **Posterior Predictive Distribution** (PPD) averages over their posterior uncertainty

$$f(Y^*|Y) = \int f(Y^*|\theta)p(\theta|Y)d\theta$$

- This properly accounts for parametric uncertainty
- The input is data, the output is a prediction distribution

Computing the PPD

- Monte Carlo sampling approximates the PPD
- Say $\theta^{(1)}, ..., \theta^{(S)}$ are samples from the posterior
- If we make a sample for Y^* for each $\theta^{(s)}$,

$$Y^{*(s)} \sim f(Y|\theta^{(s)})$$

  then the $Y^{*(s)}$ are samples from the PPD
- The posterior predictive mean is approximated by the sample mean of the $Y^{*(s)}$
- The probability that $Y^* > 0$ is approximated by the sample proportion of the $Y^{*(s)}$ that are non-zero

Bayesian prediction

```
> # Data
> Y <- 2; n <- 10

> # The posterior is theta|Y~Beta(A,B)
> A <- Y+1; B <- n-Y+1
>
> # Plug in estimate of P(Ystar>0)
> 1-dbinom(0,10,.2)
[1] 0.8926258
>
> # Approximate the PPD using MC sampling
> theta <- rbeta(100000,A,B)
> Ystar <- rbinom(100000,10,theta)
> mean(Ystar>0)
[1] 0.87454
```

Bayesian prediction