Chapter 6: Inferring Binomial probability

Bernoulli distribution

- The bernoulli distribution can be written as

$$p(y|\theta) = \theta^y * (1 - \theta)^{(1-y)}$$

- $p(y|\theta)$ is called the **likelihood function** for θ
- Likelihood function DOES NOT integrate to 1 i.e while it represents probabilities for a given value of y if does not add up to 1.
- $p(y|\theta)$ can be called one of the following
  - likelihood function for θ
  - Probability for datum y
  - Bernoulli distribution **IF** θ is fixed and y is a variable

Beta distribution

- To find posterior distribution numerically, we need the distribution for all values of θ in the range $[0, 1]$ i.e the prior distribution value.

Conjugate distribution

- While selecting the prior distribution, the following will make our computation easier
  - If the product of $p(y|\theta)$ and $p(\theta)$ results in a function with the same form as $p(\theta)$. In this case, prior and posterior distribution will be of the same form as $p(\theta)$

    - We need to solve the denominator of the Bayes rule analytically. When $p(\theta)$ has the same form as $p(y|\theta)$, then the posterior will be of the same form as $p(\theta)$. Hence, $p(\theta)$ will be called a **conjugate prior** for p(y|\theta).
- For the bernoulli distribution, we need a prior of the form

$$\theta^a * (1-\theta)^b$$

- When we multiply the above distribution with the bernoulli distribution, we get

$$\theta^{y+a} * (1-\theta)^{1-y+b}$$

- The above prior distribution is called a **beta distribution**. It has the form:

$$p(\theta|a,b) = \beta(\theta|a,b)$$

$$p(\theta|a,b) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}$$

- $B(a,b)$ is the normalizing constant that ensures the area for the beta function integrates to 1.

$$B(a,b) = \int_0^1 d\theta\ \theta^{(a-1)}(1-\theta)^{(b-1)}$$

- a and b must be positive
- Beta distribution is only defined for interval (0,1)
    - As the value of a increases, the distribution will move towards **higher** values of θ
    - As the value of b increases, the distribution will move towards **lower** values of θ
    - As the values of a and b increase together, the distribution becomes **narrower**.
    -

- a and b are called **shape parameters**

Specifying a beta prior

- The easy way to select values of a and b is to directly map them from the given data. For eg, if we have seen 4 heads and 4 tails from 8 total flips

$$n = a + b$$

- **Mean** of beta distribution

$$\mu = \frac{a}{a + b}$$

- **Mode**

$$\omega = \frac{a - 1}{a + b - 2}$$

- When a = b, mean = mode
- When a > b, mean and mode > 0.5
- When a < b, mean and mode < 0.5
- **Spread**

$$\varkappa = a + b$$

- Solving the above formulas for a and b, we get

$$a = \mu\varkappa$$

$$b = (1 - \mu)\varkappa$$

$$a = \omega(\varkappa - 2) + 1$$

$$b = (1 - \omega)(\varkappa - 2) + 1$$

- For a beta distribution with mean \mu and standard deviation σ

$$a = \mu(\frac{\mu(1-\mu)}{\sigma^2} - 1)$$

$$b = (1-\mu)(\frac{\mu(1-\mu)}{\sigma^2} - 1)$$

Posterior Beta

- Substituing the prior in the bayes formula with the bernoulli likelihood function, we can get the posterior distribution as

$$p(\theta|z,N) = \frac{p(z|\theta,N) * p(\theta)}{p(z,N)}$$
$$= \frac{\theta^{((z+a)-1)}(1-\theta)^{((N-z+b)-1)}}{B(z+a, N-z+b)}$$

- Check section 6.3.1 again for representing the posterior as a compromise between the prior and the liklelihood
- **NOTE**: The choice of prior n (which equals a + b) should represent the size of the new data set that would sway us away from our prior toward the data proportion.
-