HW 2

1. Identifying and matching kernels.

Part I: Refresher: kernel = taking out constants that don't involve variable

a) $f(\theta|y) = ye^{-\theta y} \; ; \; \theta > 0$

$\Rightarrow$ kernel: $e^{-\theta y}$

b) $\frac{2^y}{(y-1)!}$ = scalar $\Rightarrow$ kernel: $\theta^{y-1} e^{-2\theta}$

c) scalar = 3, kernel = θ^2

Part II: Refresher: $\text{Beta}(a,b) = \theta^{a-1}(1-\theta)^{b-1}$

a) $f(\theta|y) \propto \text{Beta}(3, 12)$

b) $f(\theta|y) \propto \theta^{12-1}(1-\theta)^{3-1} = \text{Beta}(12, 3)$

c) $f(\theta|y) \propto \theta^{1-1}(1-\theta)^{1-1} = \text{Beta}(1, 1)$

2. Bernoulli.

a) The joint density of $y_1, y_2, \ldots, y_n$ is the product of all the y_i since they are iid.

$\Rightarrow p(y_1, y_2, \ldots, y_n|\theta) = p(y_1|\theta) \cdot p(y_2|\theta) \cdot \ldots \cdot p(y_n|\theta)$

for $p(y_i|\theta) = \theta^{y_i}(1-\theta)^{1-y_i}$

$\Rightarrow p(y_1, y_2, \dots, y_n \mid \theta) = \prod_{i=1}^{n} (\theta^{y_i} \cdot (1-\theta)^{1-y_i})$

b) Prior: Beta$(a,b) \Rightarrow p(\theta) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}$ → beta function

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

Using Bayes' Rule:

$$p(\theta \mid y_1, y_2, \dots, y_n) = \frac{p(y_1, y_2, \dots, y_n \mid \theta) \cdot p(\theta)}{p(y_1, y_2, \dots, y_n)}$$

$$p(y_1, y_2, y_3, \dots, y_n) = \sum_{\theta} p(y_1, y_2, \dots, y_n \mid \theta) \cdot p(\theta)$$

$$= \int_0^1 \prod_{i=1}^{n} (\theta^{y_i} \cdot (1-\theta)^{1-y_i}) \cdot \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} \, d\theta$$

$$= \frac{1}{B(a,b)} \sum_{\theta} \text{everything else}$$

$$\Rightarrow p(\theta \mid y_1, y_2, \dots, y_n) = \frac{\prod_{i=1}^{n} (\theta^{y_i} \cdot (1-\theta)^{1-y_i}) \cdot \theta^{a-1}(1-\theta)^{b-1}}{B(a,b) \cdot \sum_{\theta} \prod_{i=1}^{n} (\theta^{y_i} \cdot (1-\theta)^{1-y_i})}$$

$$= \frac{\theta^{(\sum_{i=1}^{n} y_i) + (a-1)} \cdot (1-\theta)^{(n - \sum_{i=1}^{n} y_i) + (b-1)}}{\sum_{\theta} \theta^{\sum_{i=1}^{n} y_i} \cdot (1-\theta)^{n - \sum_{i=1}^{n} y_i}} \quad (\star)$$

Let $k = \sum_{i=1}^{n} y_i$. Also consider the fact that the sum of independent Bernoulli variable follows a binomial distribution:

$p(x) = \binom{n}{k} p^{k} \cdot (1-p)^{n-k}$ (for n, p, k unrelated to any values in our problem. This is just the formula). Consider again:

$$p(\theta \mid y_1, y_2, \dots, y_n) = \frac{\text{numerator } (\star)}{\sum_{\theta} \binom{n}{\sum y_i} \theta^{\sum y_i} \cdot (1-\theta)^{n - \sum y_i}}$$

this is just a scalar so it won't affect the posterior distribution's form

$$= \frac{\text{numerator } (\star)}{\sum_{\theta} \binom{n}{k} \theta^{k} \cdot (1-\theta)^{n-k}}$$

$$= \text{numerator } (\star)$$

$$= \theta^{(\sum y_i + a) - 1} \cdot (1-\theta)^{(n - \sum y_i + b) - 1}$$

$$= \text{Beta}\left(\sum_{i=1}^{n} y_i + a, \; n - \sum_{i=1}^{n} y_i + b\right)$$

3. Bayes' Rule: $p(\theta|y) = \frac{p(y|\theta) \cdot p(\theta)}{p(y)}$

fixed calculations errors thanks to peer reviews

+) $p(y|\theta)$: $Y|\theta \sim Bin(6, \theta)$

$\Rightarrow p(y|\theta) = \binom{6}{y} \theta^y (1-\theta)^{6-y}$, Kasparov won 1 game

$\Rightarrow p(1|\theta) = \binom{6}{1} \theta^1 (1-\theta)^{6-1} = 6\theta(1-\theta)^5$

+) Calculating posteriors:

- $\theta = 0.2 \Rightarrow p(\theta = 0.2 | y=1) \propto p(1|0.2) \cdot p(0.2)$
  $= 6 \times 0.2 (1-0.2)^5 \cdot 0.1 \approx 0.039$
- $\theta = 0.5 \propto 6 \times 0.5 \times (1-0.5)^5 \times 0.25$
  ≈ 0.023
- $\theta = 0.8 \propto 6 \times 0.8 \times (1-0.8)^5 \times 0.65$
  ≈ 0.000998

+) $p(y) = \sum_\theta p(y|\theta) \cdot p(\theta)$
$= p(1|0.2) \cdot p(0.2) + p(1|0.5) \cdot p(0.5) + p(1|0.8) \cdot p(0.8)$
≈ 0.088

$\Rightarrow p(\theta = 0.2 | y=1) \approx \frac{0.039}{0.088} \approx \boxed{0.44}$

$p(\theta = 0.5 | y=1) \approx \frac{0.023}{0.088} \approx \boxed{0.26}$

$p(\theta = 0.8 | y=1) \approx \frac{0.000998}{0.088} \approx \boxed{0.016}$

4. $E[L(\theta, \hat{\theta})|y] = \int L(\theta, \hat{\theta}) \cdot p(\theta|y) \, d\theta$ (because the expectation is taken with respect to $\theta|y$)

refresher: $E[\theta|y] = \int \theta \cdot p(\theta|y) d\theta$

$= \int (\theta^2 - 2\theta\hat{\theta} + \hat{\theta}^2) \cdot p(\theta|y) \, d\theta$

$= \int \theta^2 \cdot p(\theta|y) - 2\theta\hat{\theta} \cdot p(\theta|y) + \hat{\theta}^2 p(\theta|y) \, d\theta$

$= \left(\int \theta^2 \cdot p(\theta|y) \, d\theta\right) - \left(\int 2\theta\hat{\theta} \cdot p(\theta|y) \, d\theta\right) + \left(\int \hat{\theta}^2 \cdot p(\theta|y) \, d\theta\right)$

$= E[\theta^2|y] - E[2\theta\hat{\theta}|y] + E[\hat{\theta}^2|y]$

(but $\hat{\theta}$ is a constant) $= E[\theta^2|y] - 2\hat{\theta} \cdot E[\theta|y] + \hat{\theta}^2$

↳ this is minimized when

derivative with respect to $\hat{\theta}$, which is not a constant here since we need to find a value for it

$$\frac{d}{d\hat{\theta}}\left(E[\theta^2|y] - 2\hat{\theta}E[\theta|y] + \hat{\theta}^2\right) = 0$$

$$\Rightarrow -2 \cdot E[\theta|y] + 2\hat{\theta} = 0 \Rightarrow \hat{\theta} = E[\theta|y]$$ ■

5. a) General a and b choosing guidelines:

$E(\theta) = \frac{a}{a+b}$. Our mean: 15%

Higher a and b means lower variability. Our range: 10% - 25%

Based on these, we can use the model $\theta \sim Beta(2, 11)$

b) $a_{post} = a_{prior} + \#Successes = 2 + 30 = 32$

$b_{post} = b_{prior} + \#fails = 11 + 60 = 71$

$\Rightarrow \theta|y \sim Beta(32, 71) \Rightarrow y = 30, n = 90$

c) $E[\theta|y] = \frac{a+y}{a+b+n} = \frac{2+30}{2+11+90} \approx \boxed{0.31}$

Posterior mode $= \frac{a_{post} - 1}{a_{post} + b_{post} - 2} = \frac{31}{101} \approx \boxed{0.307}$

Posterior SD $= \sqrt{\frac{a_{post} \times b_{post}}{(a_{post} + b_{post})^2 (a_{post} + b_{post} + 1)}} = \sqrt{\frac{32 \times 71}{(32+71)^2 (32+71+1)}}$

$\approx \boxed{0.045}$

d) The model reflects the data more:

- Posterior mean ~ 31%, which is close to the proportion of success in the data.
- The prior model sample is much smaller than the data sample, so the posterior mean is calculated with more weight on the data success rate.

6. $p(\lambda | y_1, y_2, \dots, y_n) \propto p(y_1, y_2, \dots, y_n | \lambda) \cdot p(\lambda)$

Refresher: $p(y|\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$ for Poisson

$(y_1, y_2, \dots, y_n$ iid)

$\Rightarrow p(y_1, y_2, \dots, y_n | \lambda) = p(y_1|\lambda) \cdot p(y_2|\lambda) \cdots p(y_n|\lambda)$

$= \prod_{i=1}^{n} \left(\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}\right) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}$

$$\Rightarrow p(\lambda | y_1, y_2, \dots, y_n) \propto \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i}^{n} y_i!} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$\propto \frac{b^a}{\Gamma(a)} e^{-n\lambda} \lambda^{\sum_{i}^{n} y_i + a - 1} \cdot e^{-(b+1)\lambda}$$

Refresher: $f(x | k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$ (gamma PDF)

$\Rightarrow p(\lambda | y_1, y_2, \dots, y_n) = \text{Gamma}\left(\sum_{i=1}^{n} y_i + a, b + 1\right)$

because: λ: variable

$\Rightarrow \lambda^{\sum y_i + a - 1}$ gives $\sum y_i + a$ for the shape

$\Rightarrow e^{-(b+1)\lambda} = e^{-\frac{x}{\theta}} \Rightarrow -(b+1)\lambda = -\frac{x}{\theta}$

$\Rightarrow \theta = b + 1$

7. In our context, λ is the random variable

- Parameter a: decides the shape. The bigger a is, the smaller the interval on the curve (Beta is also like this). This also implies that the prior is strong and might weigh more in calculating the posterior. Bigger a also indicates less variability. The reverse is true for a small parameter a. In a Poisson context, a represents the number of occurrences in an interval
- Parameter b: this is the exponent of e, - The larger b is, the more concentrated the distribution is. This means lower variability in the random variable λ.