6

$$\frac{d}{d\theta}\left(\theta^t(1-\theta)^h\right) = t\,\theta^{t-1}(1-\theta)^h - \theta^t h(1-\theta)^{h-1}$$
$$= \theta^{t-1}(1-\theta)^{h-1}\left[t(1-\theta) - h\theta\right]$$

$$\frac{d^2}{d\theta^2}\left(\theta^{t-1}\cdot(1-\theta)^{h-1}\cdot\left[t\cdot(1-\theta) - h\cdot\theta\right]\right)$$
$$= \left[(t-1)\cdot\theta^{t-2}\cdot(1-\theta)^{h-1} - \theta^{t-1}\cdot(h-1)\cdot(1-\theta)^{h-2}\right]\cdot\left[t(1-\theta) - h\cdot\theta\right] - \left[\theta^{t-1}\cdot(1-\theta)^{h-1}\right]\cdot\left[t+h\right]$$
$$= t(t-1)\cdot\theta^{t-2}\cdot(1-\theta)^h - t(h-1)\cdot\theta^{t-1}\cdot(1-\theta)^{h-1} - h(t-1)\cdot\theta^{t-1}\cdot(1-\theta)^{h-1} - h\,\theta^t\cdot(h-1)\cdot(1-\theta)^{h-2} - \theta^{t-1}\cdot(1-\theta)^{h-1}\cdot(t+h)$$
$$= \theta^{t-2}\cdot(1-\theta)^h\cdot\left(t(t-1)\right) + \theta^{t-1}\cdot(1-\theta)^{h-1}\cdot\left[\underbrace{-t\cdot(h-1) - h(t-1) + (t+h)}_{\substack{-th+t-th+h+t+h \\ =2[(t+h)-th]}}\right] + \theta^t\cdot(1-\theta)^{h-2}\cdot\left(-h(h-1)\right)$$

$$= \theta^{t-2}\cdot(1-\theta)^h\cdot\left[t(t-1)\right] + \theta^{t-1}\cdot(1-\theta)^{h-1}\cdot\left[2\left((t+h)-th\right)\right] + \theta^t\cdot(1-\theta)^{h-2}\cdot\left(h(1-h)\right)$$

$$d/d\theta\left(\log\left(\theta^t(1-\theta)^h\right)\right) = d/d\theta\left(t\log(\theta) + h\log(1-\theta)\right)$$
$$= \frac{t}{\theta} - \frac{h}{1-\theta}$$

$$d/d\theta\left(t/\theta - h/(1-\theta)\right) = -t/\theta^2 + h/(1-\theta)^2 = \frac{h}{(1-\theta)^2} - \frac{t}{\theta^2}$$

7.

$$\Rightarrow \left(\log(f(x))\right)' = \frac{f'(x)}{f(x)} \Rightarrow \left(\log(f(x))\right)' = 0 \iff f'(x) = 0$$

$$\Rightarrow \left(\log(f(x))\right)'' = \frac{f''(x)f(x) - f(x)'f(x)'}{f(x)^2} = \frac{f''(x)}{f(x)} - \frac{f'(x)^2}{f(x)^2}$$

for f to have a maximum at an arbitrary point x_k we need $\begin{cases} f'(x_k) = 0 \\ f''(x_k) < 0 \end{cases}$

as for $\log(f(x)) = g(x)$, we have $\begin{cases} g'(x) = 0 \Leftrightarrow f'(x) = 0 \\ g''(x) = \frac{f''(x)}{f(x)} - \frac{f'(x)^2}{f(x)^2} \end{cases}$ $\rightarrow$ in regards to x_k where $f'(x_k) = 0$

$$g'(x_k) = f'(x_k) = 0$$

$$g''(x_k) = \frac{f''(x_k)}{f(x)} - \frac{f'(x_k)^2}{f(x)^2} = \frac{f''(x_k)}{f(x)} \quad f(x) > 0$$

$\Rightarrow$ $g''(x)$ has the same sign as $f''(x)$ in f's critical points

$\Rightarrow$ when f has a local maximum we have

$f'(x) = 0 \wedge f''(x) < 0 \Leftrightarrow g'(x) = 0 \wedge g''(x) < 0$

so any local maximum in f is also a local maximum of g

$\Rightarrow$ concluding, we have that the computation of the log-likelihood is trivial compared to the standard likelihood, and the logarithm preserves the monotony of the original function. Implying that we can compute the log-likelihood when trying to solve likelihood optimization problems.

8

- find posterior distribution $\longrightarrow$ $p(\theta | D) = \frac{p(D|\theta)\, p(\theta)}{p(D)}$
- prior is $Beta(a, b) = p(\theta)$

- so we know that $p(D|\theta)$ is modeled as a binomial distribution with param. θ
- since the random variable originates from a Bernoulli experiment we can assume that the variable is i.i.d

$\Rightarrow$ likelihood $= P(D|\theta) = \theta^m (1-\theta)^{N-m}$ $\rightarrow$ leading constant has no impact in maximizing likelihood

$$\Rightarrow p(\theta|D) = \left[\theta^m (1-\theta)^{N-m} \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \right] / P(D)$$

$$\propto \theta^{m+a-1} (1-\theta)^{N-m-b-1} \longrightarrow \text{similar to Beta distribution}$$

$$\sim Beta(m+a, N-m-b) = \underbrace{\frac{\Gamma(N+a-b)}{\Gamma(m+a)\,\Gamma(N-m-b)}}_{1/P(D)} \theta^{m+a-1} (1-\theta)^{N-m-b-1}$$

$\Rightarrow$ so now we know that $P(\theta | D) = \text{Beta}(m+a, N-m-b)$

$\Rightarrow$ and $E[\theta | D]$ is known for a Beta distribution $\longrightarrow \dfrac{\alpha}{\alpha+\beta} \longrightarrow \dfrac{m+a}{N+a-b}$

$\Rightarrow$ the prior distribution was also a Beta dist. with $E[\theta] = \dfrac{a}{a+b}$

$\Rightarrow$ we need to show that $\dfrac{m+a}{N+a-b} = \lambda \left(\dfrac{a}{a+b}\right) + (1-\lambda)\theta_{MLE}$

$$\Rightarrow \theta_{MLE} = \arg\max_\theta p(D|\theta) = \arg\max_\theta \theta^m (1-\theta)^{N-m} = \arg\max_\theta \log(\theta^m (1-\theta)^{N-m})$$

$$= \arg\max_\theta m \log\theta + (N-m)\log(1-\theta)$$

$$\Rightarrow (m \log(\theta) + (N-m)\log(1-\theta))' = 0 \Leftrightarrow$$

$$\Leftrightarrow \frac{m}{\theta} - \frac{(N-m)}{1-\theta} = 0 \Leftrightarrow \frac{1-\theta}{\theta} = \frac{N-m}{m} \Leftrightarrow \frac{1}{\theta} = \frac{N-m-m}{m} \Leftrightarrow \theta = \frac{m}{N-2m}$$

$$\Leftrightarrow \frac{m+a}{N+a-b} = \lambda\left(\frac{a}{a+b}\right) + (1-\lambda)\left(\frac{m}{N-2m}\right) = \lambda\left(\frac{a}{a+b}\right) + \frac{m}{N-2m} - \frac{\lambda m}{N+2m} \Leftrightarrow$$

$$\Leftrightarrow \frac{m+a}{N+a-b} - \frac{m}{N-2m} = \lambda\left(\frac{a}{a+b} - \frac{m}{N-2m}\right) \Leftrightarrow \lambda = \frac{\frac{m+a}{N+a-b} - \frac{m}{N-2m}}{\frac{a}{a+b} - \frac{m}{N-2m}}$$

$\Rightarrow$ this can be solved for λ and theoretically (assuming no mistakes) should result in a value in $[0,1]$

(3) $\lambda_{MAP} = \arg\max_\lambda p(\lambda | x)$

$$p(\lambda|x) \propto p(x|\lambda)p(\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$\propto \lambda^x e^{-\lambda} \cdot \lambda^{a-1} e^{-b\lambda} = \lambda^{x+a-1} e^{-\lambda(b+1)}$$

$\Rightarrow$ we are looking for $\arg\max_\lambda \left(\lambda^{x+a-1} e^{-\lambda(b+1)}\right)$

$\Rightarrow$ taking the logarithm (preserving monotony)

$$\log\left(\lambda^{x+a-1} e^{-\lambda(b+1)}\right) = (x+a-1)\log(\lambda) - \lambda(b+1)\underbrace{\log(e)}_{1}$$

$$\left(\log(p(\lambda|x))\right)' = \frac{x+a-1}{\lambda} - (b+1)$$

$$= 0 \Leftrightarrow \frac{x+a-1}{b+1} = \lambda$$

$$\left(\log(p(\lambda|x))\right)'' = \frac{-(x+a-1)}{\lambda^2}$$

$$\rightarrow -\frac{(x+a-1)}{\frac{(x+a-1)^2}{(b+1)^2}} = -\frac{\overbrace{(b+1)^2}^{>0}}{\underbrace{x+a-1}_{>0}} < 0 \;//$$

$\Rightarrow$ so $\lambda_{MAP} = \arg\max_{\lambda}(p(\lambda|x)) = \dfrac{x+a-1}{b+1}$