HW 3 Problem 1

1.1 MLE: $\theta = \text{argmax}_\theta (1-\theta)^{m-h}\theta^h$

↑ Bernoulli's Distribution

m = total data points
h = total heads in data

$\log\theta = (m-h)\log(1-\theta) + h\log\theta$

$\frac{\partial}{\partial\theta}\log\theta = -\frac{m-h}{1-\theta} + \frac{h}{\theta} = 0$

$\frac{h}{\theta} = \frac{m-h}{1-\theta}$

$h - h\theta = m\theta - h\theta$

$h = m\theta$

$\theta_{MLE} = \frac{h}{m} = \frac{\text{heads in dataset}}{\text{total pts in dataset}}$

1.2 MAP:

$\theta_{MAP} = \text{argmax}_\theta P(\theta|D)$

using Bayes Theorem

$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$

↑ ignore b/c no θ term

$P(\theta) = \theta^{a-1}(1-\theta)^{b-1}$ (Beta distribution)

$\theta_{MAP} = \text{argmax}_\theta P(D|\theta)P(\theta)$

$= \text{argmax}_\theta \theta^h(1-\theta)^{m-h}(\theta^{a-1})(1-\theta)^{b-1}$

$= (\theta^{a+h-1})(1-\theta)^{b+m-h-1}$

$\log\theta_{MAP} = (a+h-1)\log\theta + (b+m-h-1)\log(1-\theta)$

$\frac{\partial}{\partial\theta}\log\theta_{MAP} = \frac{a+h-1}{\theta} + -\frac{b+m-h-1}{1-\theta} = 0$

$\frac{a+h-1}{\theta} = \frac{b+m-h-1}{1-\theta}$

$\theta(b+m-h-1) = (a+h-1)(1-\theta)$

$\theta b + \theta m - \theta h - \theta = a+h-1 - \theta a - \theta h + \theta$

$\theta a + \theta b + \theta m - 2\theta = a+h-1$

$\theta_{MAP} = \frac{a+h-1}{a+b+m-2}$ => if $a=b=1$, $\theta_{MAP} = \frac{1+h-1}{1+1+m-2} = \frac{h}{m} = \theta_{MLE}$