=) Consider a Bernoulli distribution as our likelihood function:

$$\mathcal{L}(\theta;X) = p(X|\theta) \quad , \theta \to \text{prob. of heads}$$

$$\stackrel{iid}{=} \prod_{n=1}^{n} p(x_n|\theta)$$

$$\mathcal{L}(\theta;X) = \theta^{k}(1-\theta)^{n-k} \qquad k \to \text{no. of heads}$$

$n \to$ total.

$$\log \mathcal{L}(\theta;X) = k\log\theta + (n-k)\log(1-\theta)$$

i.

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \log \mathcal{L}(\theta;X)$$

$$\Rightarrow \frac{\partial \log \mathcal{L}}{\partial \theta} = 0$$

$$\Rightarrow \frac{k}{\hat{\theta}} + \frac{(n-k)}{(1-\hat{\theta})} \times (-1) = 0$$

$$\frac{k}{\hat{\theta}} = \frac{(n-k)}{1-\hat{\theta}}$$

$$\Rightarrow k - k\hat{\theta} = n\hat{\theta} - k\hat{\theta}$$

$$\boxed{\Rightarrow \hat{\theta} = \frac{k}{n}} \quad \text{MLE.}$$

(ii) Within a Bayesian framework, we try to find the posterior.

i.e.

$$p(\theta|x) = \frac{\overbrace{p(x|\theta)}^{\text{likelihood}} \cdot \overbrace{p(\theta)}^{\text{Prior}}}{\underbrace{p(x)}_{\text{Normalising const.}}}$$

$$p(\theta|x) \propto p(x|\theta) \cdot p(\theta)$$

for our model,

$$p(\theta|x) \propto \theta^{k}(1-\theta)^{n-k} \cdot U(0,1)$$

$$\propto \theta^{k}(1-\theta)^{n-k} \times \frac{1}{1-0}$$

$$p(\theta|x) \propto \theta^{k}(1-\theta)^{n-k}$$

ii)

$$\mathbb{E}(\theta) = \int_0^1 \theta \, p(\theta|x) \, d\theta$$

$$= \int_0^1 \theta \left\{ \theta^k (1-\theta)^{n-k} \right\} d\theta$$

$$= \int_0^1 \theta^{k+1} (1-\theta)^{n-k+1} \, d\theta$$

Using $\text{Beta}(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} \, dt$

$$\mathbb{E}(\theta) = \text{Beta}(k+2, n-k+2)$$

$$\boxed{\mathbb{E}(\theta) = \frac{\Gamma(k+2)\,\Gamma(n-k+2)}{\Gamma(n+4)}}$$

iii)

$$\theta_{MAP} = \underset{\theta}{\arg\max} \; p(\theta|x)$$

$$= \underset{\theta}{\arg\max} \; \theta^k (1-\theta)^{n-k}.$$

again, we can write $\frac{\partial \mathcal{L}}{\partial \theta} = 0$

and get $\boxed{\hat{\theta}_{MAP} = \frac{k}{n}}$ which is intuitive because we're using Uniform Prior.