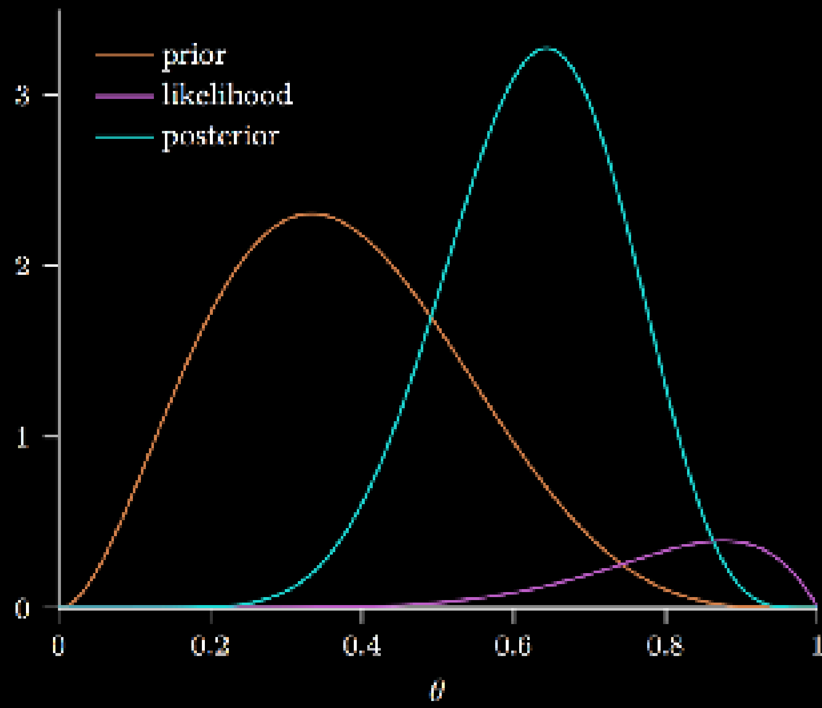

Maximum a Posteriori Probability Estimation (MAP)

For example, we can choose $\hat{\theta}$ to be the most likely θ given the data.
MAP Principle: Find $\hat{\theta}$ that maximizes the posterior distribution $P(\theta \mid D)$:

$$\begin{aligned}\hat{\theta}_{MAP} &= argmax_{\theta}\, P(\theta \mid D) \\ &= argmax_{\theta}\, logP(D \mid \theta) + logP(\theta)\end{aligned}$$

For out coin flipping scenario, we get:

$$\begin{aligned}\hat{\theta}_{MAP} &= argmax_{\theta}\; P(\theta|Data) \\ &= argmax_{\theta}\; \frac{P(Data|\theta)P(\theta)}{P(Data)} && \text{(By Bayes rule)} \\ &= argmax_{\theta}\; log(P(Data|\theta)) + log(P(\theta)) \\ &= argmax_{\theta}\; n_H \cdot log(\theta) + n_T \cdot log(1-\theta) + (\alpha - 1) \cdot log(\theta) + (\beta - 1) \cdot log(1-\theta) \\ &= argmax_{\theta}\; (n_H + \alpha - 1) \cdot log(\theta) + (n_T + \beta - 1) \cdot log(1-\theta) \\ &\Longrightarrow \hat{\theta}_{MAP} = \frac{n_H + \alpha - 1}{n_H + n_T + \beta + \alpha - 2}\end{aligned}$$

- As $n \rightarrow \infty$, $\hat{\theta}_{MAP} \rightarrow \hat{\theta}_{MLE}$.
- MAP is a great estimator if prior belief exists and is accurate.
- If n is small, it can be very wrong if prior belief is wrong!

"True" Bayesian approach

Note that MAP is only one way to get an estimator for θ. There is much more information in $P(\theta \mid D)$. So, instead of the maximum as we did with MAP, we can use the posterior mean (end even its variance).

$$\hat{\theta}_{post\_mean} = E[\theta, D] = \int_{\theta} \theta P(\theta \mid D) d\theta$$

For coin flipping, this can be computed as $\hat{\theta}_{post\_mean} = \frac{n_H + \alpha}{n_H + \alpha + n_T + \beta}$.