The equal-tail credible interval approach is ideal when the posterior distribution is symmetric.

But what if \( \pi(\theta|x) \) is skewed?

Picture:
Note that values of $\theta$ around 1 have much higher posterior probability than values around 7.5.

Yet 7.5 is in the equal-tails interval and 1 is not!

A better approach here is to create our interval of $\theta$-values having the **Highest Posterior Density**.
**Defn:** A $100(1 - \alpha)\%$ HPD region for $\theta$ is a subset $C \in \Theta$ defined by

$$C = \{\theta : \pi(\theta|x) \geq k\}$$

where $k$ is the **largest** number such that

$$\int_{\theta : \pi(\theta|x) \geq k} \pi(\theta|x) \, d\theta = 1 - \alpha.$$  

- The value $k$ can be thought of as a horizontal line placed over the posterior density whose intersection(s) with the posterior define regions with probability $1 - \alpha$. 
HPD Intervals / Regions

Picture: (95% HPD Interval)

\[ P\{\theta^*_L < \theta < \theta^*_U\} = 0.95. \]

The values between \( \theta^*_L \) and \( \theta^*_U \) here have the highest posterior density.
The HPD region will be an interval when the posterior is unimodal.

If the posterior is multimodal, the HPD region might be a discontiguous set.

The set \( \{ \theta : \theta \in (1.5, 3.9) \cup (5.8, 7.1) \} \) is the HPD region for \( \theta \) here.
Example 1 Revisited: HPD Interval

- See course web page for finding an HPD interval in R for the cabinet duration data example.

- Also note the hpd function in TeachingDemos package in R.
- See code for Example 2 (coin-flipping data) in R.
Conjugate Priors

- A prior $p(\theta)$ for a sampling model is called a **conjugate prior** if the resulting posterior $\pi(\theta|X)$ is in the **same distributional family** as the prior.

- For example, in Example 2, note that the Uniform(0,1) prior is simply a beta(1,1) prior.
  So: Prior is beta and likelihood is binomial
  $\Rightarrow$ Posterior is beta (with different parameter values!)

- Therefore this was a conjugate prior.
Suppose we observe $n$ independent Bernoulli($p$) r.v.’s $X_1, \ldots, X_n$. We wish to estimate the “success probability” $p$ via the Bayesian approach.

We will use a $\text{beta}(a, b)$ prior for $p$ and show this is a conjugate prior.

Consider the r.v. $Y = \sum_{i=1}^{n} X_i$. This has a binomial($n, p$) distribution.

We first write the joint density of $Y$ and $p$ (using $f(\cdot)$ to denote densities, not $p(\cdot)$, to avoid confusion with the parameter $p$).
Derivation of Beta/Binomial Model

\[ f(y, p) = f(y|p)f(p) = \left[ \binom{n}{y} p^y (1-p)^{n-y} \right] \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \]

\[ = \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{y+a-1}(1-p)^{n-y+b-1} \]
Derivation of Beta/Binomial Model

Although it is not really necessary, let's derive the marginal density of $Y$:

$$f(y) = \int_0^1 f(y, p) \, dp$$

$$= \frac{\Gamma(n + 1)}{\Gamma(y + 1)\Gamma(n - y + 1)} \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{y+a-1}(1 - p)^{n-y+b-1} \, dp$$

$$= \frac{\Gamma(n + 1)\Gamma(a + b)}{\Gamma(y + 1)\Gamma(n - y + 1)\Gamma(a)\Gamma(b)} \frac{\Gamma(y + a)\Gamma(n - y + b)}{\Gamma(n + a + b)}$$

$$\times \int_0^1 \frac{\Gamma(n + a + b)}{\Gamma(y + a)\Gamma(n - y + b)} p^{y+a-1}(1 - p)^{n-y+b-1} \, dp$$

$$= \frac{\Gamma(n + 1)\Gamma(a + b)}{\Gamma(y + 1)\Gamma(n - y + 1)\Gamma(a)\Gamma(b)} \frac{\Gamma(y + a)\Gamma(n - y + b)}{\Gamma(n + a + b)}$$
Derivation of Beta/Binomial Model

Then the posterior $\pi(p|y) = f(p|y)$ is

$$
\frac{f(y, p)}{f(y)} = \frac{\Gamma(n+1) \Gamma(a+b) p^{y+a-1}(1-p)^{n-y+b-1}}{\Gamma(y+1)\Gamma(n-y+1) \Gamma(a)\Gamma(b) \Gamma(n+a+b)}
$$

$$
= \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)} p^{y+a-1}(1-p)^{n-y+b-1}, \quad 0 \leq p \leq 1.
$$

Clearly this posterior is a beta($y + a, n - y + b$) distribution.
As an interval estimate for $p$, we could use a (quantile-based or HPD) credible interval based on this posterior.

As a point estimator of $p$, we could use:

1. The posterior mean $E[p|Y]$ (the usual Bayes estimator)
2. The posterior median
3. The posterior mode
Consider letting $\hat{p}_B$ = the posterior mean.

The mean of the (posterior) beta distribution is:

$$\hat{p}_B = \frac{y + a}{y + a + n - y + b} = \frac{y + a}{a + b + n}$$

Note $\hat{p}_B = \frac{y}{a + b + n} + \frac{a}{a + b + n}$

$$= \left[ \frac{n}{a + b + n} \right] \left( \frac{y}{n} \right) + \left[ \frac{a + b}{a + b + n} \right] \left( \frac{a}{a + b} \right)$$
So the Bayes estimator $\hat{p}_B$ is a weighted average of the usual frequentist estimator (sample mean) and the prior mean.

As $n \uparrow$, the sample data are weighted more heavily and the prior information less heavily.

In general, with Bayesian estimation, as the sample size increases, the likelihood dominates the prior.

Example with anthropology data.