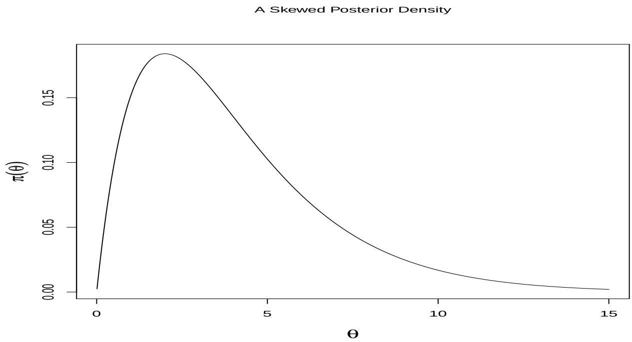
HPD Intervals / Regions

- The equal-tail credible interval approach is ideal when the posterior distribution is symmetric.
- But what if $\pi(\theta|x)$ is skewed?

Picture:



HPD Intervals / Regions

- Note that values of θ 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 θ-values having the **Highest Posterior Density**.

HPD Intervals / Regions

Defn: A $100(1-\alpha)\%$ HPD region for θ is a subset $\mathcal{C} \in \Theta$ defined by

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

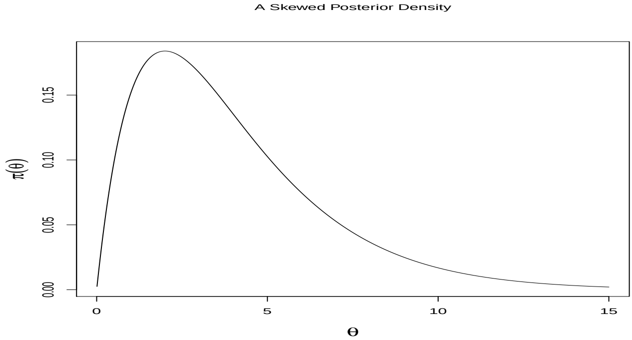
where k is the **largest** number such that

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

- 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)



$\Rightarrow P\{\theta_L^* < \theta < \theta_U^*\} = 0.95.$
The values between θ_L^* and θ_U^* here have the **highest posterior density**.

HPD Intervals / Regions

- 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**.
  Picture:
- The set $\{\theta : \theta \in (1.5, 3.9) \cup (5.8, 7.1)\}$ is the HPD region for θ 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 $p(\boldsymbol{\theta})$ for a sampling model is called a **conjugate prior** if the resulting posterior $\pi(\boldsymbol{\theta}|\mathbf{X})$ 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.

Complete Derivation of Beta/Binomial Bayesian Model

- 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 *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

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

Derivation of Beta/Binomial Model

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

$$
\begin{aligned}
f(y) &= \int_0^1 f(y,p)\,\mathrm{d}p \\
&= \frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{y+a-1}(1-p)^{n-y+b-1}\,\mathrm{d}p \\
&= \frac{\Gamma(n+1)\Gamma(a+b)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(a)\Gamma(b)} \frac{\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(n+a+b)} \\
&\times \int_0^1 \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)} p^{y+a-1}(1-p)^{n-y+b-1}\,\mathrm{d}p \\
&= \frac{\Gamma(n+1)\Gamma(a+b)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(a)\Gamma(b)} \frac{\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(n+a+b)}
\end{aligned}
$$

Derivation of Beta/Binomial Model

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

$$\begin{aligned}
\frac{f(y,p)}{f(y)} &= \frac{\frac{\Gamma(n+1)}{\Gamma(y+1)\Gamma(n-y+1)}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{y+a-1}(1-p)^{n-y+b-1}}{\frac{\Gamma(n+1)\Gamma(a+b)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(a)\Gamma(b)}\frac{\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(n+a+b)}} \\
&= \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)}p^{y+a-1}(1-p)^{n-y+b-1}, \;\; 0 \le p \le 1.
\end{aligned}$$

Clearly this posterior is a beta$(y+a, n-y+b)$ distribution.

Inference with Beta/Binomial Model

- 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

Inference with Beta/Binomial Model

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

$$= \left[\frac{n}{a+b+n}\right]\left(\frac{y}{n}\right) + \left[\frac{a+b}{a+b+n}\right]\left(\frac{a}{a+b}\right)$$

Inference with Beta/Binomial Model

- 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.