# Bayesian Inference: Posterior Intervals

- Simple values like the posterior mean $E[\boldsymbol{\theta}|\mathbf{X}]$ and posterior variance $var[\boldsymbol{\theta}|\mathbf{X}]$ can be useful in learning about $\boldsymbol{\theta}$.
- Quantiles of $\pi(\boldsymbol{\theta}|\mathbf{X})$ (especially the posterior median) can also be a useful summary of $\boldsymbol{\theta}$.
- The ideal summary of $\boldsymbol{\theta}$ is an interval (or region) with a certain probability of containing $\boldsymbol{\theta}$.
- Note that a classical (frequentist) **confidence interval** does not exactly have this interpretation.

# Definitions of Coverage

- **Defn.**: A random interval $\big(L(\mathbf{X}), U(\mathbf{X})\big)$ has $100(1-\alpha)\%$ **frequentist coverage** for $\theta$ if, **before** the data are gathered,

$$P[L(\mathbf{X}) < \theta < U(\mathbf{X})|\theta] = 1 - \alpha.$$

(**Pre-experimental** $1-\alpha$ coverage)

- Note that if we observe $\mathbf{X} = \mathbf{x}$ and plug $\mathbf{x}$ into our confidence interval formula,

$$P[L(\mathbf{x}) < \theta < U(\mathbf{x})|\theta] = \begin{cases} 0 & \text{if } \theta \notin \big(L(\mathbf{x}), U(\mathbf{x})\big) \\ 1 & \text{if } \theta \in \big(L(\mathbf{x}), U(\mathbf{x})\big) \end{cases}$$

(**NOT** Post-experimental $1-\alpha$ coverage)

# Definitions of Coverage

- **Defn.**: An interval $\big(L(\mathbf{x}), U(\mathbf{x})\big)$, **based on the observed data** $\mathbf{X} = \mathbf{x}$, has $100(1-\alpha)\%$ **Bayesian coverage** for $\theta$ if

$$P[L(\mathbf{x}) < \theta < U(\mathbf{x}) | \mathbf{X} = \mathbf{x}] = 1 - \alpha.$$

(**Post-experimental** $1-\alpha$ coverage)

- The frequentist interpretation is less desirable if we are performing inference about $\theta$ based on a **single** interval.

# Frequentist Coverage for Bayesian Intervals

- Hartigan (1966) showed that for standard posterior intervals, an interval with $100(1-\alpha)\%$ **Bayesian coverage** will have

$$P[L(\mathbf{X}) < \theta < U(\mathbf{X})|\theta] = (1-\alpha) + \epsilon_n,$$

  where $|\epsilon_n| < a/n$ for some constant $a$.
  $\Rightarrow$ **Frequentist** coverage $\rightarrow 1-\alpha$ as $n \rightarrow \infty$.
- Note that many classical CI methods only achieve $100(1-\alpha)\%$ frequentist coverage asymptotically, as well.

# Bayesian Credible Intervals

- A **credible interval** (or in general, a **credible set**) is the Bayesian analogue of a confidence interval.
- A $100(1-\alpha)\%$ credible set $\mathcal{C}$ is a subset of $\Theta$ such that

$$\int_{\mathcal{C}} \pi(\boldsymbol{\theta}|\mathbf{X})\, d\boldsymbol{\theta} = 1 - \alpha.$$

- If the parameter space $\Theta$ is discrete, a sum replaces the integral.

# Quantile-Based Intervals

- If $\theta_L^*$ is the $\alpha/2$ posterior quantile for $\theta$, and $\theta_U^*$ is the $1-\alpha/2$ posterior quantile for $\theta$, then $(\theta_L^*, \theta_U^*)$ is a $100(1-\alpha)\%$ credible interval for $\theta$.

Note: $P[\theta < \theta_L^* | \mathbf{X}] = \alpha/2$ and $P[\theta > \theta_U^* | \mathbf{X}] = \alpha/2$.

$$
\begin{aligned}
&\Rightarrow P\{\theta \in (\theta_L^*, \theta_U^*) | \mathbf{X}\} \\
&= 1 - P\{\theta \notin (\theta_L^*, \theta_U^*) | \mathbf{X}\} \\
&= 1 - \left( P[\theta < \theta_L^* | \mathbf{X}] + P[\theta > \theta_U^* | \mathbf{X}] \right) \\
&= 1 - \alpha.
\end{aligned}
$$

# Quantile-Based Intervals

Picture:

# Example: Quantile-Based Interval

- Suppose $X_1, \ldots, X_n$ are the durations of cabinets for a sample of cabinets from Western European countries.
- We assume the $X_i$'s follow an exponential distribution.

$$p(X_i|\theta) = \theta e^{-\theta X_i}, \;\; X_i > 0$$
$$\Rightarrow L(\theta|\mathbf{X}) = \theta^n e^{-\theta \sum_{i=1}^n x_i}$$

Suppose our prior distribution for $\theta$ is

$$p(\theta) \propto 1/\theta, \;\; \theta > 0.$$

$\Rightarrow$ Larger values of $\theta$ are less likely **a priori**.

# Example: Quantile-Based Interval

Then

$$\begin{aligned}\pi(\theta|\mathbf{X}) &\propto p(\theta)L(\theta|\mathbf{X}) \\ &\propto \left(\frac{1}{\theta}\right)\theta^n e^{-\theta\sum x_i} \\ &= \theta^{n-1}e^{-\theta\sum x_i}\end{aligned}$$

- This is the **kernel** of a **gamma** distribution with "shape" parameter $n$ and "rate" parameter $\sum_{i=1}^n x_i$.
- So including the normalizing constant,

$$\pi(\theta|\mathbf{X}) = \frac{(\sum x_i)^n}{\Gamma(n)}\theta^{n-1}e^{-\theta\sum x_i}, \quad \theta > 0.$$

# Example: Quantile-Based Interval

- Now, given the observed data $x_1, \ldots, x_n$, we can calculate any quantiles of this gamma distribution.
- The 0.05 and 0.95 quantiles will give us a 90% credible interval for $\theta$.
- See `R` example with real data on course web page.

# Example: Quantile-Based Interval

- Suppose we feel $p(\theta) = 1/\theta$ is too subjective and favors small values of $\theta$ too much.
- Instead, let's consider the **noninformative** prior

$$p(\theta) = 1, \quad \theta > 0$$

  (favors all values of $\theta$ equally).
- Then our posterior is

$$\begin{aligned}\pi(\theta|\mathbf{X}) &\propto p(\theta)L(\theta|\mathbf{X}) \\ &= (1)\theta^n e^{-\theta \sum x_i} \\ &= \theta^{(n+1)-1} e^{-\theta \sum x_i}\end{aligned}$$

  $\Rightarrow$ This posterior is a gamma with parameters $(n+1)$ and $\sum x_i$.
- We can similarly find the equal-tail credible interval.

# Example 2: Quantile-Based Interval

- Consider 10 flips of a coin having $P\{\text{Heads}\} = \theta$.
- Suppose we observe 2 "heads".
- We model the count of heads as binomial:

$$p(X|\theta) = \binom{10}{X}\theta^X(1-\theta)^{10-X}, \;\; x = 0, 1, \ldots, 10.$$

- Let's use a uniform prior for $\theta$:

$$p(\theta) = 1, \;\; 0 \leq \theta \leq 1.$$

# Example 2: Quantile-Based Interval

- Then the posterior is:

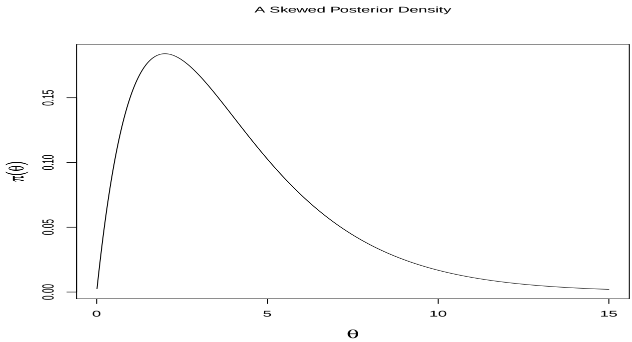
$$\begin{aligned}\pi(\theta|x) &\propto p(\theta)L(\theta|x)\\ &= (1)\binom{10}{x}\theta^x(1-\theta)^{10-x}\\ &\propto \theta^x(1-\theta)^{10-x}, \;\; 0 \le \theta \le 1.\end{aligned}$$

- This is a **beta** distribution for $\theta$ with parameters $x+1$ and $10-x+1$.
- Since $x = 2$ here, $\pi(\theta|x=2)$ is beta(3,9).
- The 0.025 and 0.975 quantiles of a beta(3,9) are (.0602, .5178), which is a 95% credible interval for $\theta$.

# 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 $\theta$ around 2.2 have **much** higher posterior probability than values around 11.5.
- Yet 11.5 is in the equal-tails interval and 2.2 is not!
- A better approach here is to create our interval of $\theta$-values having the **Highest Posterior Density**.

# HPD Intervals / Regions

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

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

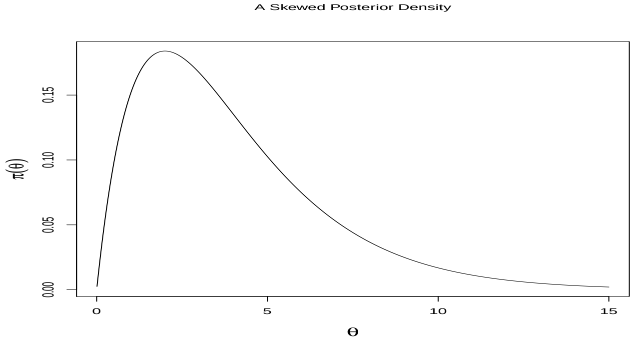
where $k$ is the **largest** number such that

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

- 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 $\theta_L^*$ and $\theta_U^*$ here have the **highest posterior density**.