Interval Estimation

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

02/8/2021

Credible sets

- Often wish to describe full posterior with a *credible set*
- A $100 \times (1 - \alpha)\%$ credible set for θ is a subset $C \subseteq \Theta$ with

$$1 - \alpha \leq P(C \mid \mathbf{y}) = \int_C p(\theta \mid \mathbf{y}) d\theta$$

- "The probability that θ is in C is at least $(1 - \alpha)$"
- If $\theta \in \mathbb{R}$ and $C = [a, b]$ for $a, b \in \mathbb{R}$, C may be called a *credible interval* or *Bayesian confidence interval*

Credible sets: HPD

- A *highest posterior density* (HPD) includes only the "most likely" θ values:

$$C = \{\theta \in \Theta : p(\theta \mid \mathbf{y}) \geq K(\alpha)\}$$

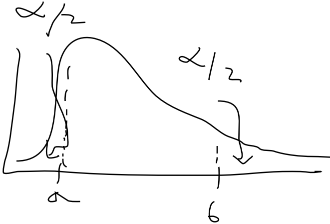
where $K(\alpha)$ is the largest constant giving

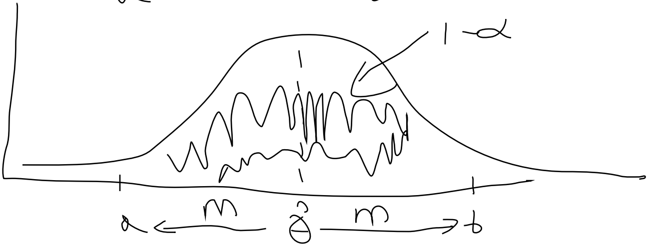
$$P(C \mid \mathbf{y}) \geq 1 - \alpha.$$

- The smallest possible credible set
- May not be an interval
- Visualized as horizontal line intersecting density curve.


p(θ|y)
1-α
k(α)
a
b


α/2
1-α
α/2
a
b


α/2
α/2
a
b


1-α
m
m
a ← θ̂ → b

Credible sets: other intervals

- A *equal tail* interval excludes equal probability in the upper and lower tail
  - $C = [a, b]$ where $a = \alpha/2$ quantile and $b = (1 - \alpha/2)$ quantile of $p(\theta \mid \mathbf{y})$
  - Symmetric – includes "the middle" of the posterior
  - Invariant under monotone transformations

- Alternatively, construct interval so that point estimate $\hat{\theta}$ is central:

$$C = [\hat{\theta} - m, \hat{\theta} + m]$$

Example: Emergency room

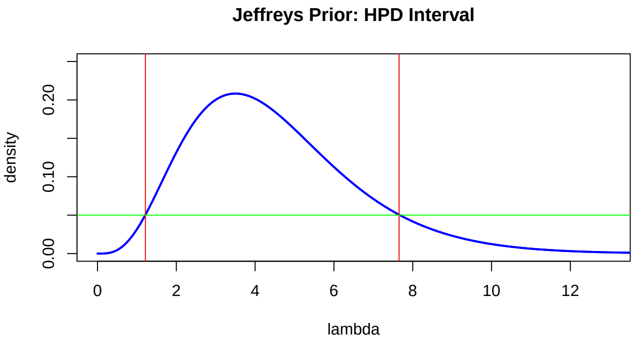
- An emergency room receives $y = 4$ collapsed lung (Pneumothorax) patients in one week.
- Model as Poisson distribution with weekly rate λ
- Use (improper) Jeffries prior for λ:

$$p(\lambda) \propto \sqrt{1/\lambda} \propto \text{Gamma}(0.5, 0)$$

- Posterior is $p(\lambda \mid y = 4) = \text{Gamma}(4.5, 1)$
- Create 90% credible interval for λ

Example: Emergency room

- HPD interval: $K(0.10) \approx 0.05$

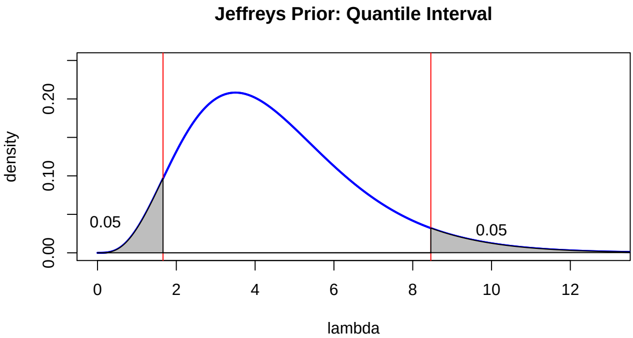


-

- $C = [1.21, 7.65]$

Example: Emergency room

- Find 0.05 and 0.95 quantiles for Gamma$(4.5, 1)$

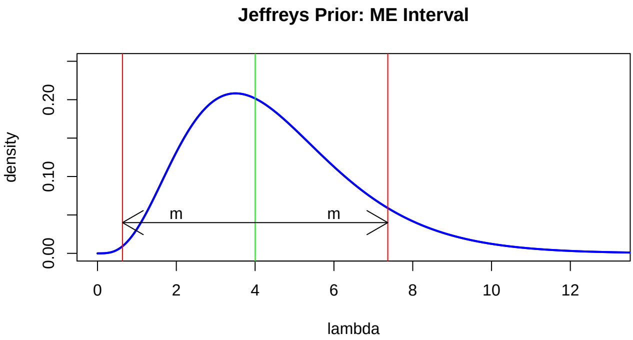


- http://www.ericfrazerlock.com/Interval_estimation_Rcode1.r

- $C = [1.66, 8.46]$

Example: Emergency room

- Create interval centered at MLE $\hat{\lambda} = 4$
  - $m \approx 3.37$



- http://www.ericfrazerlock.com/Interval_estimation_Rcode1.r

- $C = [0.63, 7.37]$

Decision Theory for Credible Sets

- Jim Berger, 1985, in *Statistical decision theory and Bayesian analysis*
  - "we do not view credible sets as having a clear decision-theoretic role, and are therefore leery of 'optimality' approaches to selection of a credible set"
- Nevertheless...
- Action space could involve all subsets of parameter space

$$\mathcal{A} = \{a : a \subseteq \Theta\}$$

- Alternatively, could just consider intervals. For univariate θ:

$$\mathcal{A} = \{a = [u, v] : u, v \in \mathbb{R}\}$$

Decision Theory for Credible Sets

- Possible loss function:

$$l(\theta, d(\mathbf{y})) = \mathbb{1}_{\{\theta \notin d(\mathbf{y})\}} + K \cdot \text{Volume}(d(\mathbf{y}))$$

- $\mathbb{1}_{\{\cdot\}}$ defines the indicator function $\mathbb{1}_{\{A\}} = \begin{cases} 1 & \text{if } A \\ 0 & \text{o.w.} \end{cases}$
- Volume$(d(\mathbf{y}))$ is the *volume* of $d(\mathbf{y})$

$$\int_{\theta \in d(\mathbf{y})} d\theta$$

  - For $d(\mathbf{y}) = [a, b]$, Volume$(d(\mathbf{y})) = b - a$

- Balances the coverage and size of a credible set.

Decision Theory for Credible Sets

- Leads to posterior risk $E 1_{\{\theta \notin d(y)\}} = P(\theta \notin d(y))$

$$\rho(p_\theta, d(\mathbf{y})) = P(\theta \notin d(\mathbf{y}) \mid \mathbf{y}) + K \cdot \text{Volume}(d(\mathbf{y}))$$

  - Minimized for $d(\mathbf{y}) = \{\theta : p(\theta \mid \mathbf{y}) \geq K\}$, an HPD interval

$$= 1 - P(\theta \in d(y)) + K \cdot \int_{d(y)} d\theta$$

$$= 1 - \int_{d(y)} p(\theta|y)\, d\theta + K \int_{d(y)} d\theta$$

$$= 1 + \int_{d(y)} (K - p(\theta|y))\, d\theta$$

minimized for

+ for $p(\theta|y) < K$

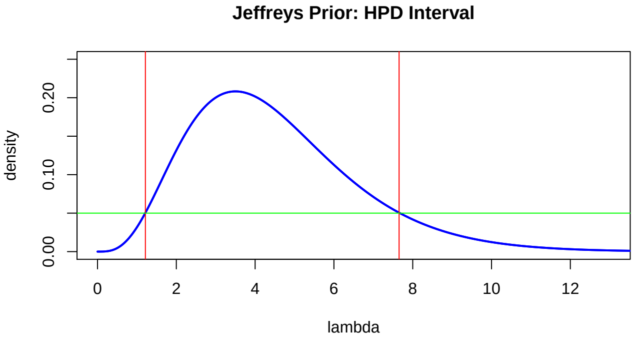
− for $p(\theta|y) > K$

Example: Emergency room (cont.)

- Decision theoretic framework:
  - *prior distribution:* $p(\lambda) = \text{Gamma}(1/2, 0)$, $\lambda \in \mathbb{R}^+$
  - *sampling distribution:* $y \sim \text{Poisson}(\lambda)$, $y \in \{0, 1, 2, \ldots\}$
  - *allowable actions:* $\mathcal{A} = \{a : a \subseteq \mathbb{R}^+\}$
  - *decision rule:* $d(y) = \{\lambda : p(\lambda \mid y) \geq 0.05\}$
  - *loss function:* $l(\lambda, d(y)) = \mathbb{1}_{\{\lambda \notin d(y)\}} + 0.05 \cdot \text{Volume}(d(y))$
- Recall $y = 4$, and so $p(\lambda \mid y) = \text{Gamma}(4.5, 1)$

Example: Emergency room (cont.)

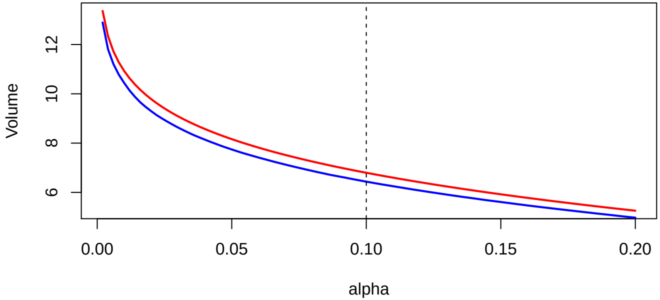
- Here $K = 0.05$ gives a HPD interval with $\alpha = 0.10$



- $d(y = 4) = [1.21, 7.65]$
  - an interval with volume $7.65 - 1.21 = 6.44$

Example: Emergency room (cont.)

- Volume/coverage plots for HPD and symmetric quantile intervals:



Code: http://www.ericfrazerlock.com/More_on_Interval_estimation_Rcode1.r