Credible Sets

Statistics · Module 6 — Bayesian Inference

Node 6.4

Position in Knowledge Graph

System: Statistics

Structural Domain: Bayesian Inference

Node: 6.4

Credible sets are the Bayesian analogue of confidence sets, providing posterior probability statements about parameter regions. This node depends on the posterior distribution (Node 6.1) and the frequentist confidence set framework (Node 4.1), which it reinterprets from the Bayesian perspective. It enables posterior predictive model checking (Node 6.5) and connects to asymptotic frequentist-Bayesian reconciliation via the Bernstein–von Mises theorem (Node 6.7).

Formal Definition

Let $\pi(\theta \mid x)$ be the posterior distribution on $\Theta \subseteq \mathbb{R}^k$.

Credible Set

A set $C(x) \subseteq \Theta$ is a $(1-\alpha)$-**credible set** if

$$\int_{C(x)} \pi(\theta \mid x)\, d\theta \geq 1 - \alpha.$$

The probability is with respect to the posterior. The parameter θ is the random quantity; the data x are fixed.

Highest Posterior Density (HPD) Region

The **HPD region** at level $1-\alpha$ is

$$C^{\mathrm{HPD}}(x) = \big\{\theta \in \Theta : \pi(\theta \mid x) \geq k_\alpha\big\},$$

where k_α is the largest constant such that

$$\int_{C^{\mathrm{HPD}}(x)} \pi(\theta \mid x)\, d\theta = 1 - \alpha.$$

Equal-Tailed Credible Interval

For scalar θ, the **equal-tailed** $(1-\alpha)$**-credible interval** is

$$C^{\mathrm{ET}}(x) = \big[Q_{\alpha/2},\ Q_{1-\alpha/2}\big],$$

where $Q_p = Q_p(\pi(\cdot \mid x))$ denotes the p-quantile of the posterior distribution.

Intuition

A credible set directly answers: given the observed data and the model, what region contains θ with posterior probability at least $1-\alpha$?

The HPD region collects the most probable parameter values first, so it is the shortest interval (smallest volume region) achieving the desired posterior coverage. For symmetric unimodal posteriors, the HPD and equal-tailed intervals coincide. For skewed posteriors, the HPD region is shorter but asymmetric, while the equal-tailed interval is symmetric in probability but longer.

The credible set interpretation is conceptually simpler than the frequentist confidence set: it conditions on the observed data and makes a probability statement about the parameter. The price is dependence on the prior.

Structural Role

Credible sets are the canonical Bayesian uncertainty quantification tool for parameters. They require the posterior (Node 6.1) and parallel the frequentist confidence sets (Node 4.1).

The HPD region connects to Bayesian point estimation (Node 6.3): the MAP estimator is always contained in the HPD region, and as $\alpha \to 1$ the HPD region shrinks to the MAP.

The Bernstein–von Mises theorem (Node 6.7) establishes that in regular parametric models, $(1-\alpha)$-credible sets have asymptotic frequentist coverage $1-\alpha$, reconciling the two frameworks.

Mathematical Development

Proof: HPD is the Shortest Credible Interval

Theorem. Among all $(1-\alpha)$-credible sets in $\mathbb{R}$, the HPD region has the smallest Lebesgue measure.

Proof. Let C be any $(1-\alpha)$-credible set with $\int_C \pi(\theta \mid x)\, d\theta \geq 1-\alpha$. Let C^{HPD} be the HPD region with threshold k_α.

Consider the symmetric difference. For any $\theta_1 \in C^{\mathrm{HPD}} \setminus C$ and $\theta_2 \in C \setminus C^{\mathrm{HPD}}$, we have $\pi(\theta_1 \mid x) \geq k_\alpha > \pi(\theta_2 \mid x)$.

The posterior mass of $C \setminus C^{\mathrm{HPD}}$ must be at least that of $C^{\mathrm{HPD}} \setminus C$ (since both C and C^{HPD} achieve coverage $1-\alpha$). But every point in $C \setminus C^{\mathrm{HPD}}$ has posterior density strictly less than every point in $C^{\mathrm{HPD}} \setminus C$. Therefore

$$\lambda(C \setminus C^{\mathrm{HPD}}) \geq \frac{\int_{C^{\mathrm{HPD}} \setminus C} \pi(\theta \mid x)\, d\theta}{k'_\alpha} \geq \lambda(C^{\mathrm{HPD}} \setminus C),$$

where $k'_\alpha \leq k_\alpha$ bounds the density on $C \setminus C^{\mathrm{HPD}}$ from above. Thus $\lambda(C) = \lambda(C \cap C^{\mathrm{HPD}}) + \lambda(C \setminus C^{\mathrm{HPD}}) \geq \lambda(C^{\mathrm{HPD}})$. $\square$

Formal Distinction from Confidence Sets

A frequentist $(1-\alpha)$-confidence set $C_{\text{freq}}(X)$ satisfies

$$\mathbb{P}_\theta\big(\theta \in C_{\text{freq}}(X)\big) \geq 1-\alpha \quad \text{for all } \theta \in \Theta.$$

Here θ is fixed and X varies. The probability is over repetitions of the experiment.

A Bayesian $(1-\alpha)$-credible set $C(x)$ satisfies

$$\pi\big(\theta \in C(x) \mid X = x\big) \geq 1-\alpha.$$

Here x is fixed and θ varies under the posterior. The probability is conditional on the observed data.

These are fundamentally different probability statements. A credible set with good posterior coverage need not have good frequentist coverage, and vice versa.

When Credible Sets Have Frequentist Coverage

Under the Bernstein–von Mises theorem (Node 6.7), for regular parametric models with fixed dimension:

$$\pi(\theta \mid X_1, \ldots, X_n) \approx \mathcal{N}\left(\hat{\theta}_n,\ \frac{I(\theta_0)^{-1}}{n}\right)$$

in total variation. The HPD credible region based on this posterior is asymptotically

$$C_n^{\text{HPD}} \approx \big\{\theta : n(\theta - \hat{\theta}_n)^\top I(\theta_0)(\theta - \hat{\theta}_n) \leq \chi^2_{k,1-\alpha}\big\},$$

which coincides with the Wald confidence ellipsoid. Therefore $\mathbb{P}_{\theta_0}(\theta_0 \in C_n^{\text{HPD}}) \to 1-\alpha$.

HPD for Common Posterior Families

Normal posterior: If $\theta \mid x \sim \mathcal{N}(\mu_n, \sigma_n^2)$, the HPD interval is symmetric about μ_n:

$$C^{\text{HPD}} = \big[\mu_n - z_{1-\alpha/2}\,\sigma_n,\ \mu_n + z_{1-\alpha/2}\,\sigma_n\big].$$

This coincides with the equal-tailed interval by symmetry.

Beta posterior: If $\theta \mid x \sim \text{Beta}(\alpha_n, \beta_n)$ with $\alpha_n \neq \beta_n$, the HPD interval is shorter than the equal-tailed interval. It must be computed numerically by finding k_α such that the density level set has posterior mass $1-\alpha$.

Gamma posterior: Similarly asymmetric; the HPD interval shifts toward smaller values relative to the equal-tailed interval due to right skewness.

Multivariate HPD Regions

For $\theta \in \mathbb{R}^k$, the HPD region is a level set of the joint posterior density. For a multivariate normal posterior $\theta \mid x \sim \mathcal{N}_k(\mu_n, \Sigma_n)$, the HPD region is the ellipsoid

$$C^{\text{HPD}} = \big\{\theta : (\theta - \mu_n)^\top \Sigma_n^{-1}(\theta - \mu_n) \leq \chi^2_{k,1-\alpha}\big\}.$$

Worked Examples

Example 1: HPD Interval for a Beta Posterior

Data: $n = 30$ Bernoulli trials, $s = 21$ successes. Prior: $\mathrm{Beta}(1, 1)$ (uniform).

Posterior: $\mathrm{Beta}(22, 10)$. Posterior mean: $22/32 = 0.6875$. Posterior mode: $21/30 = 0.7$.

Equal-tailed 95% interval: Using Beta quantiles, $[Q_{0.025}, Q_{0.975}] \approx [0.518, 0.832]$. Length: 0.314.

HPD 95% interval: The density level set $\{\theta : f(\theta) \geq k\}$ must be found numerically. For $\mathrm{Beta}(22, 10)$: $C^{\mathrm{HPD}} \approx [0.530, 0.840]$. Length: 0.310.

The HPD interval is shifted slightly toward the mode and is shorter than the equal-tailed interval due to the right skewness of $\mathrm{Beta}(22, 10)$.

Example 2: Normal Posterior Credible Interval vs Confidence Interval

Data: $n = 25$, $\bar{x} = 5.2$, $\sigma^2 = 9$ (known). Prior: $\mu \sim \mathcal{N}(4, 4)$.

Posterior: $\mu \mid x \sim \mathcal{N}(\mu_n, \tau_n^2)$ with

$$\tau_n^2 = \frac{1}{1/4 + 25/9} = \frac{1}{0.25 + 2.778} = \frac{1}{3.028} \approx 0.330,$$

$$\mu_n = 0.330 \cdot (0.25 \cdot 4 + 2.778 \cdot 5.2) = 0.330 \cdot (1 + 14.444) = 0.330 \cdot 15.444 \approx 5.097.$$

95% credible interval: $5.097 \pm 1.96\sqrt{0.330} = 5.097 \pm 1.125 = [3.972, 6.222]$.

95% confidence interval (frequentist, no prior): $5.2 \pm 1.96 \cdot 3/5 = 5.2 \pm 1.176 = [4.024, 6.376]$.

The credible interval is shorter (width 2.25 vs 2.35) and shifted toward the prior mean of 4. As $n \to \infty$, the two intervals converge.

Cross-Links

- **Linear Algebra:** Multivariate HPD regions are ellipsoids defined by the posterior precision matrix. The eigenvectors of the precision matrix give the principal axes of the credible ellipsoid, connecting to PCA-like decompositions of posterior uncertainty.
- **AI:** Bayesian neural networks produce posterior distributions over weights; credible sets for predictions quantify epistemic uncertainty. Approximate HPD regions are computed via MCMC samples or variational posteriors.
- **Economics:** Bayesian credible intervals for structural parameters (elasticities, policy multipliers) directly quantify parameter uncertainty conditional on observed data, which is the natural object for policy decisions.
- **Quantum Mechanics:** Credible regions for quantum states (density matrices) arise in quantum state tomography. The HPD region over the space of density matrices provides the smallest set of states consistent with measurement data at a given posterior probability level.

Structural Compression

Invariant: $\pi(\theta \in C(x) \mid X = x) \geq 1 - \alpha$. The posterior mass of the credible set is at least $1 - \alpha$.

Mechanism: Integrate the posterior density over the candidate region $C(x)$. The HPD region selects the level set of the posterior density, which minimizes volume subject to the coverage constraint. This is a constrained optimization: minimize Lebesgue measure subject to posterior mass $\geq 1 - \alpha$.

Cross-System Echo: Credible sets are the Bayesian analogue of confidence sets. The Bernstein–von Mises theorem (Node 6.7) establishes their asymptotic equivalence in regular models, making credible sets the bridge between Bayesian and frequentist uncertainty quantification.

Exercises

1. For the posterior $\theta \mid x \sim \text{Gamma}(10, 2)$, compute the equal-tailed 95% credible interval and explain why the HPD interval will be shorter.
2. Prove that for a unimodal symmetric posterior, the HPD interval and the equal-tailed interval coincide.
3. Let $\theta \mid x \sim \text{Beta}(5, 15)$. Compute the posterior mean, posterior mode, and the equal-tailed 90% credible interval. Without computing it exactly, argue whether the HPD interval will shift left or right relative to the equal-tailed interval.
4. Construct an example where a 95% credible interval has frequentist coverage far from 95% for a specific θ_0. Specify the model, prior, and true parameter.
5. For the multivariate normal posterior $\theta \mid x \sim \mathcal{N}_2(\mu_n, \Sigma_n)$ with $\mu_n = (3, 1)^\top$ and $\Sigma_n = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, derive the equation of the 95% HPD ellipse and find its semi-axes.
6. Prove that the MAP estimator is always contained in the HPD region $C^{\text{HPD}}(x)$ for every $\alpha \in (0, 1)$, and that the HPD region shrinks to $\{\hat{\theta}^{\text{MAP}}\}$ as $\alpha \to 1$.
7. Argue, structurally, why the distinction between credible sets and confidence sets is sharpest in small samples with informative priors, and vanishes asymptotically under the conditions of the Bernstein–von Mises theorem: identify the role of the prior, the likelihood dominance rate, and the resulting posterior shape.

Statistics · Module 6 — Bayesian Inference · Node 6.4
Concepts: prior-posterior, bayesian-updating
Prerequisites: 6.1, 4.1
Enables: 6.5, 6.6

hbar.university — own.your.cognition