Automated Posterior Interval Evaluation for Inference in Probabilistic Programming

Edward Kao, Michael Yee | MIT Lincoln Laboratory

ABSTRACT
In probabilistic inference, credible intervals constructed from posterior samples provide ranges of likely values for continuous parameters of interest. Intuitively, an inference procedure is optimal if it produces the most precise posterior intervals that cover the true parameter value with the expected frequency in repeated experiments. We present new theories and methods for automating posterior interval evaluation of inference performance in probabilistic programming using two metrics: 1) truth coverage, and 2) ratio of the empirical over the ideal interval widths. Demonstrating with inference on popular regression and state-space models, we show how the metrics provide effective comparisons between different inference procedures, and capture the effects of collinearity and model misspecification. Overall, we claim such automated interval evaluation can accelerate the robust design and comparison of probabilistic inference programs by directly diagnosing how accurately and precisely they can estimate parameters of interest.

THEORY
Based on the statistical principle of evaluating Bayesian inference with frequentist properties (1,2), we compute two metrics for inference output in repeated experiments: 1) posterior credible interval coverage of the true parameter value (90% intervals should cover the truth 90% of the time), and 2) ratio of the empirical over the ideal interval widths (ratio of 1 indicates precise inference). The ideal interval width can be computed based on the asymptotic theorem:

The Bernstein-von Mises Theorem (3): For regular models, the posterior distribution of continuous parameters in finite dimensions converges asymptotically, with increasing data, to distribution Normal with mean at the true value $\theta^*$ and covariance equal to the inverse of the Fisher information matrix $I$ evaluated at $\theta^*$:

$$ \theta \sim \text{Normal}(\theta^*, -I(\theta^*)^{-1}) $$

The diagonal terms of the asymptotic covariance $I(\theta^*)$ provide the ideal interval width for each parameter in the model. In the non-asymptotic regime, the ideal interval width can be computed using the Laplace approximation where $I(\theta^*)$ is the Hessian of the log posterior distribution evaluated at $\theta^*$:

$$ \theta \sim \text{Normal}(\theta^*, -q(\theta^*)^{-1}) $$

AUTOMATION
Computing the proposed metrics based on posterior intervals can be automated for any probabilistic programming systems (4) that simulate data $D$ and parameters $\theta$ based on statistical models $M$ and priors, and infer the posterior distribution such that the likelihood function and the unnormalized posterior distribution are accessible. The Fisher information matrix can be computed via the hessian function on the log likelihood, through auto-differentiation. For non-asymptotic cases using the Laplace approximation, simply replace the likelihood with the unnormalized posterior distribution. We implement and demonstrate this automated evaluation in Gen (5).

FUTURE WORK
- Apply the proposed evaluation to real-world scenarios with a single data realization and unknown truth.
- Extend approach to general models with a mixture of regular and irregular parameters through conditioning and exploring generalizations of the Bernstein-von Mises Theorem.
- Explore simple and automated indicators for the adequacy of the normal assumption on the true posterior.

DEMONSTRATION ON BAYESIAN LINEAR REGRESSION
\[ y | x, \beta, \sigma^2 \sim \text{Normal}(\beta x, \sigma^2) \]
\[ y, \beta, \sigma^2 \sim \text{Normal}(X y, \sigma^2 I) \]

where $y$ and $\beta \in \mathbb{R}^d$, $y \in \mathbb{R}^N$ and $X$ is an $N \times d$ covariate matrix. Here, $d \geq 10$ and the covariates $X$ are generated independently with $\sigma = 0.6$ correlation.

For purpose of demonstration, we infer $\beta$ using Gibbs sampling, even though the posterior has a closed-form expression.

Evaluation cases: 1) Regular Gibbs sampling, 2) Covariates generated with collinearity, and 3) Prior location is misplaced.

DEMONSTRATION ON BAYESIAN LOGISTIC REGRESSION
\[ y | x, \beta, \pi \sim \text{Bernoulli}(\pi) \]
\[ y, \beta, \pi \sim \text{Bernoulli}(\text{sigmoid}(x \beta)) \]

We infer $\beta$ using Random Walk Metropolis-Hastings with two multivariate normal proposals: $w' \sim \text{Normal}(w, 0.2)$.

1. Scott proposal (approximation to asymptotically optimal proposal)
2. Naive proposal (diagonal covariance of $\Sigma = 0.2 I$)

DEMONSTRATION ON NONLINEAR STATE-SPACE MODEL
\[ x_t = x_{t-1} + 25 x_{t-1} - 8 \cos(x_{t-1}) + \delta_{t-1} \]
\[ y_t = x_{t-1} + \epsilon_t \]

A popular nonlinear state-space model in the literature. Here we generate 100 steps in time to capture both the periodic motion and nonlinear drift.

Evaluation quantifies how much faster the Scott proposal converges than the naive proposal, under collinearity.

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited.