STA 250 Lecture Notes(Oct. 21)

Olivia Lee

Metropolis-Within-Gibbs

- Allows you to sample from a high-dimensional distribution using a sequence of lower dimensional distributions.
- Generalizes Gibbs sampler in two ways
  1. If we can't sample directly/exactly, then we can use MH.
  2. Sample sub-blocks of parameters, not necessary full conditionals.
     (ie) $P(\theta_1, \theta_2, \phi|\overrightarrow{y})$
     $\rightarrow$ sample$P(\theta_1, \theta_2|\phi, \overrightarrow{y})$ (exactly)
     $\rightarrow$ sample$P(\phi|\theta_1, \theta_2 \overrightarrow{y})$ (using MH)

Idea of Homework:
Two Strategies

1. Sample from $P(\overrightarrow{\beta}|\overrightarrow{y})$ using MH and a multivariate proposal for $\overrightarrow{\beta}$
2. For $j = 1, ..., p$ sample from $P(\beta_j|\beta_{[-j]}, \overrightarrow{y})$

$\star$ Recommend to do both in homework!

Q:When to use non-symmetric proposals?
A:When θ has compact support then proposal can respect the boundaries.
Ex: $X_i \sim Posi(\lambda) i = 1, ..., n.$ Prior $\lambda \sim t_v I_{\{(0,\infty)\}}$
Use MH, proposal $\theta^\star \sim TN(\theta^{(t)}, v^2, [0, \infty)) \Rightarrow$ No longer symmetric!

Checking your MCMC code
For one dataset, how to know we reach convergence?

1. By eye, using traceplots.
2. Use effective sample size to gauge roughly "how well" converged
3. Other diagnostics : run multiple Markov Chains then compare if they are the same(Gelmen-Rubin)

$\star$ If pass these tests, it doesn't necessary mean convergence. But if fail to pass, no converge.

Fortunately, we can use **simulation studies** to check if everything is working!

Idea:

1. Simulation $\theta_{(j)}$ from the prior $P(\theta)$

2. Simulate a dataset $\overrightarrow{y}_{(j)}$ from the model $P(y|\theta_{(j)})$
3. Sample from posterior for dataset $\overrightarrow{y}_{(j)}$
4. Find $100(1-\alpha)\%$ central credible interval for θ
5. Record yes or no that the interval contained $\theta_{(j)}$
6. Check roughly $100(1-\alpha)\%$ of intervals contained their specific $\theta_{(j)}$

<u>Why does this work?</u>

$$\begin{aligned}
&\int P(\theta)\int I_{\{\theta\in S^{1-\alpha}(\overrightarrow{y})\}}P(\overrightarrow{y}|\theta)dyd\theta\\
=&\int\int I_{\{\theta\in S^{1-\alpha}(\overrightarrow{y})\}}P(\theta)P(\overrightarrow{y}|\theta)dyd\theta\\
=&\int P(y)[\int I_{\{\theta\in S^{1-\alpha}(\overrightarrow{y})\}}P(\overrightarrow{y}|\theta)d\theta]dy\\
=&(1-\alpha)\int P(y)dy\\
=&1-\alpha
\end{aligned}$$

Comparison

<u>Bayes:</u> $\int P(\theta)\int I_{\{\theta\in S^{1-\alpha}(\overrightarrow{y})\}}P(\overrightarrow{y}|\theta)dyd\theta = 1-\alpha$

<u>Frequentist:</u> $\int I_{\{\theta\in C^{1-\alpha}(\overrightarrow{y})\}}P(y|\theta)dy = 1-\alpha \quad \forall\theta$

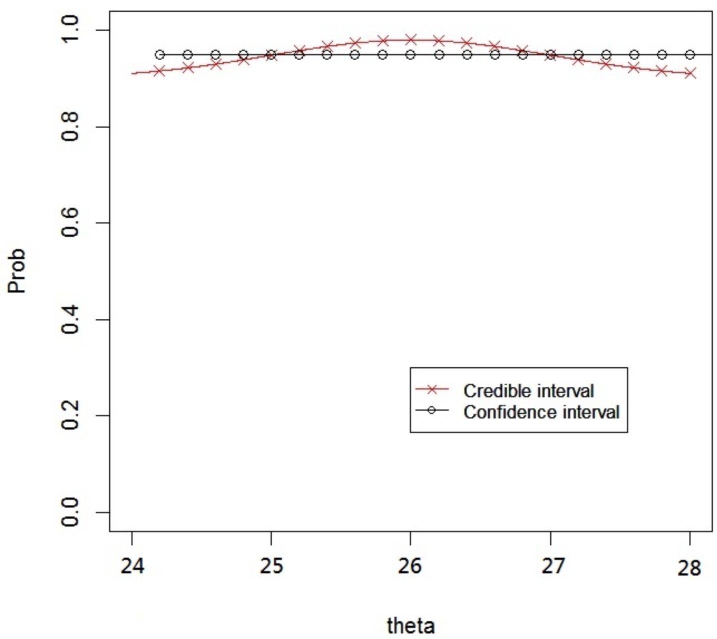


Figure 1: 95% Credoble Interval vs 95% Confidence Interval

Posterior Predictive Checking

• Validation simulation checks you can sample from the posterior under your model. But it doesn't tell you if your model is a good fit to the data.

Idea:

Having fit your model to the data, you know roughly what θ is, so if you simulate from $P(y|\theta)$ the simulated data should look "similar" to the real data.

Formally: $P(\tilde{y}|\overrightarrow{y}) = \int P(\tilde{y}, \theta|\overrightarrow{y})d\theta = \int P(\tilde{y}|\theta)P(\theta|\overrightarrow{y})d\theta$

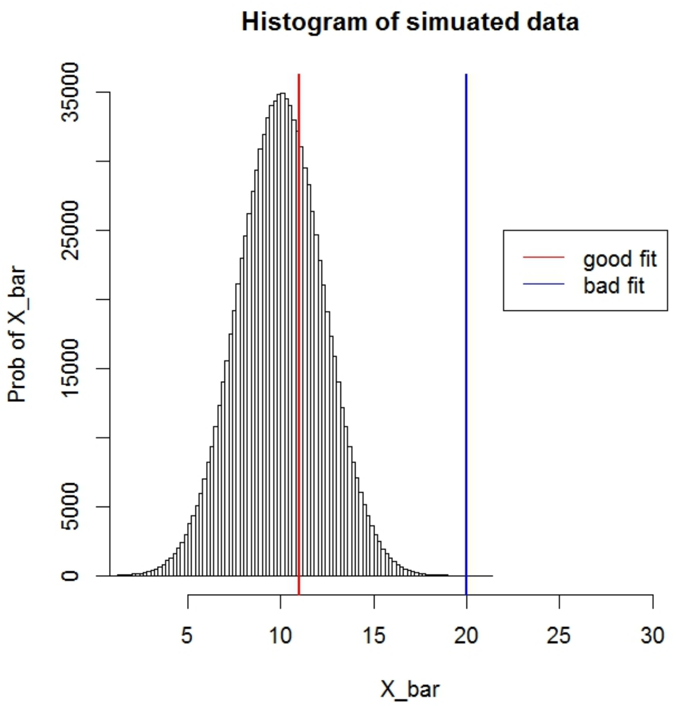
Recipe:

•Sample $\theta^{(t)}$ from the posterior $P(\theta|\overrightarrow{y})$

•For each$\theta^{(t)}$, sample a new dataset from $P(\overrightarrow{y}|\theta^{(t)})$

•We now have is m predictive datasets and 1 real dataset.

•We can take univariate summary statistics of each dataset and compare to the real dataset



⋆ Could use mean, median, min, max.

⋆ Compare the posterior predictive p-value.

Posterior Predictive P-value:

$$p = 2 * min\{P(T(y^*) > T(\overrightarrow{y})), P(T(y^*) < T(\overrightarrow{y}))\}$$

Heuristically, if p is small then your model is not a good fit of the data.