STA 250 Class Note - Oct. 21

Chih-Hsin Hsueh

I. Metropolis-Within-Gibbs

- Allows you to sample from a high-dimensional distribution using a sequence of lower dimensional distributions.
- Generalize Gibbs Sampler in two ways:
  - If we cannot sample directly (exactly), then you can use MH.
  - Sample sub-blocks of parameters, not necessarily full conditionals.
    i.e $P(\theta_1, \theta_2, \phi \mid \vec{y}) \rightarrow$ sample $P(\theta_1, \theta_2 \mid \phi, \vec{y}) \rightarrow$ sample $P(\phi \mid \theta_1, \theta_2, \vec{y})$

  [**Note**: Idea for homework: 2 strategies – (a) Sampling from $P(\vec{\beta} \mid \vec{y})$ using MH and a multivariate proposal for $\vec{\beta}$. (b) For $j = 1, \ldots, p$ sample β_j from $P(\beta_j \mid \beta_{[-j]}, \vec{y})$.]
- When to use non-symmetric proposals?
  - If θ has compact support, then a non-symmetric proposal can be used to respect the boundaries.
  - example:
    $X \sim_{iid} Poi(\lambda),\ i = 1, \ldots, n,$ prior $\lambda\ t_\nu \sim I_{[0,\infty]}$.
    Use MH proposal $\theta^\star \sim TN(\theta^{(t)}, \nu^2, [0, \infty)) \Longrightarrow$ No longer symmetric!!

II. Checking Your MCMC Codes

For one dataset, how to know if we reach convergence?

(i) Trace plot, "by eye".

(ii) Use ESS (effective sample size) to gauge roughly "how well" a chain converges (i.e., the equivalent number of independent samples).

(iii) Other diagnostics.

- Gelman-Rubin (Multiple chains)
- Heidelberger
- Geweke

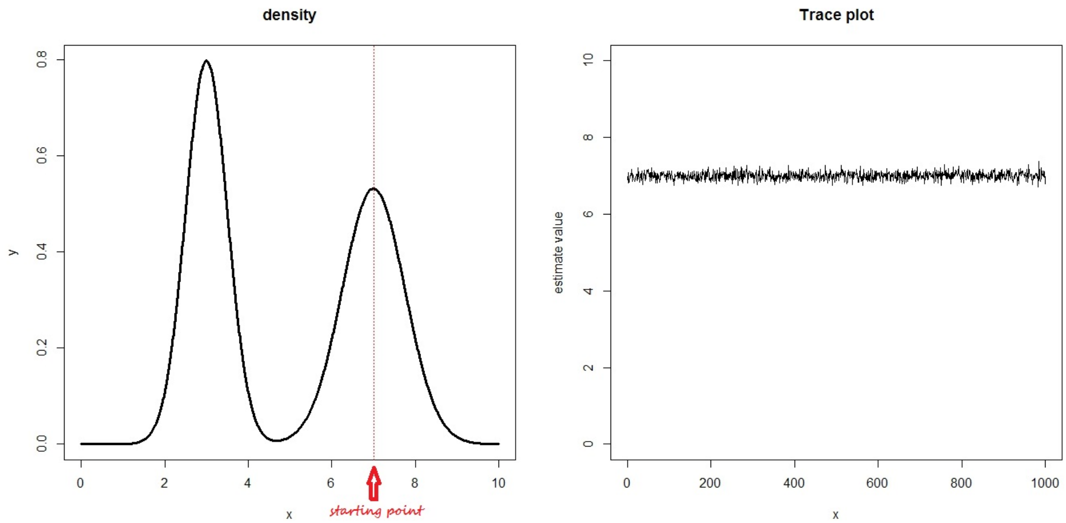
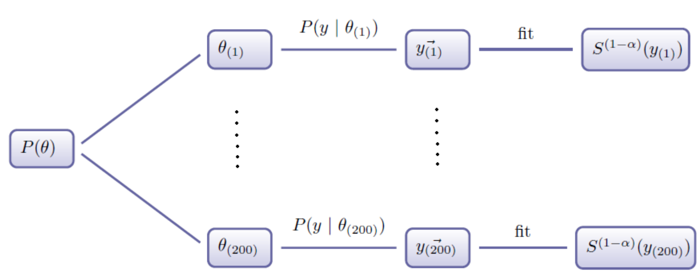


Figure 1: If our starting point is $x = 7$, then our Markov Chain will walk around $x = 7$ with beautiful trace plot. In this case, although we pass three tests above, we can not conclude our MCMC converges.

The other choice for checking convergence: Fortunately, we can use **simulation studies** to check everything is working.

<u>Idea:</u>

1. Simulate $\theta_{(j)}$ from the prior $P(\theta)$.
2. Simulate a dataset $\vec{y_{(j)}}$ from the model $P(y \mid \theta_{(j)})$.
3. Sample from the posterior for data set $\vec{y_{(j)}}$.
4. Find $100(1-\alpha)\%$ central credible interval for θ from dataset $\vec{y_{(j)}}$.



5. Record Yes/No whether the interval contained $\theta_{(j)}$.
6. Check roughly $100(1-\alpha)\%$ of intervals contain their own specific $\theta_{(j)}$.

Why does it work?

$$\begin{aligned}
\int P(\theta) \int I_{\{\theta \in S^{1-\alpha}(y)\}} P(\vec{y} \mid \theta) dy d\theta &= \int \int I_{\{\theta \in S^{1-\alpha}(y)\}} P(\theta) P(\vec{y} \mid \theta) dy d\theta \\
&= \int \int I_{\{\theta \in S^{1-\alpha}(y)\}} P(\vec{y}) P(\theta \mid \vec{y}) d\theta dy \\
&= \int P(\vec{y}) \left(\int I_{\{\theta \in S^{1-\alpha}(y)\}} P(\theta \mid \vec{y}) d\theta \right) dy \\
&= \int P(\vec{y})(1-\alpha) dy \\
&= 1 - \alpha
\end{aligned}$$

$$\begin{cases} \textbf{Bayes:} & \int P(\theta) \int I_{\{\theta \in S^{1-\alpha}(y)\}} P(\vec{y} \mid \theta) dy \; d\theta = 1 - \alpha \\ \textbf{Frequentist:} & \int I_{\{\theta \in C^{1-\alpha}(y)\}} P(\vec{y} \mid \theta) dy = 1 - \alpha, \quad \forall \theta \end{cases}$$

Example

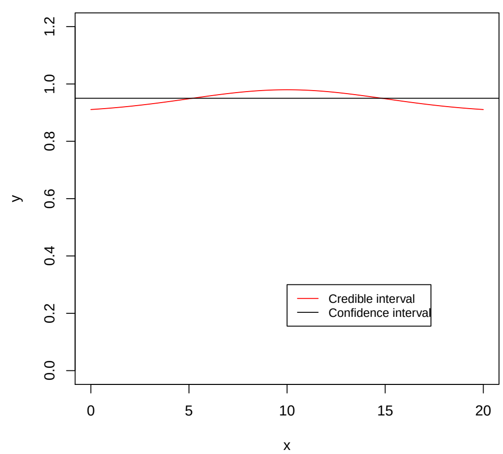


Figure 2: 95% credible interval v.s 95% C.I.

III. Posterior Predictive Checking

Validation simulation checks you can sample from the posterior under your model. 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 simulated from

$P(y \mid \theta)$, the simulated data should look "similar" to the real data.

Formally:

$$P(\tilde{y} \mid \vec{y}) = \int P(\tilde{y}, \theta \mid \vec{y}) d\theta$$
$$= \int P(\tilde{y} \mid \theta) P(\theta \mid \vec{y}) d\theta$$

Recipe:

- Sample $\theta^{(t)}$ from the posterior $P(\theta \mid \vec{y})$.
- For each $\theta^{(t)}$, sample a new dataset from $P(y \mid \theta^{(t)})$.
- We now have is m predictive dataset and one real dataset.
- We can take univariate summary statistics of each dataset and the real dataset and compare. [**Note**: We could use min, mean, median, max.]

Example:

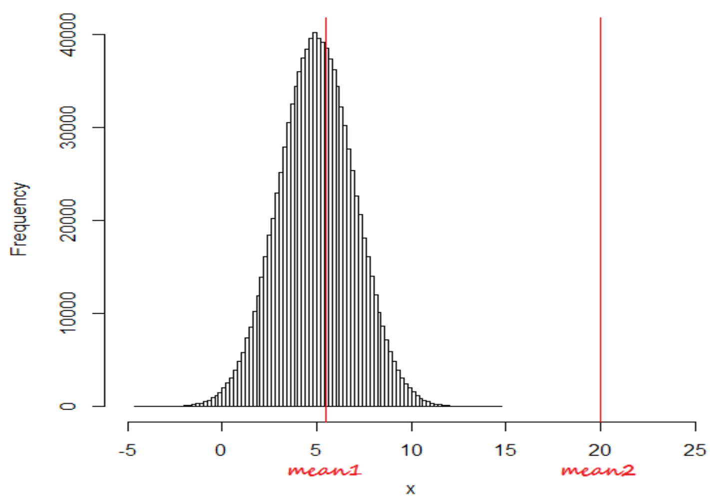


Figure 3: If the real mean equals mean1, then we will conclude we have a good fit. If the real mean equals mean2, then we will conclude it's not a good fit.

Posterior predictive p-value:

$$2 \min\{P(T(y^{\star}) > T(\vec{y})), P(T(y^{\star}) < T(\vec{y}))\}$$

Heuristically, if P is small, then your model is not a good fit to the data.