STAT 3006: Statistical Computing
Lecture 10*

19 March

7.5 Convergence Diagnostics

When can we stop MCMC iterations? In this subsection, we provide a way to check whether or not the Markov chain constructed by MCMC attains stationary. For a parameter θ of interest, we start with $\frac{m}{2}$ chains ($\frac{m}{2}$ different initial values for θ) and each chain has $4n$ iterations for MCMC. m is even and n is an integer. For each chain, we only keep last half of iterations, which results in $\frac{m}{2}$ chains and each chain has $2n$ iterations. Moreover, we further cut each chain into two chains with equal length. Finally, we obtain $m = \frac{m}{2} \times 2$ chains and each chain has $n = \frac{2n}{2}$ iterations. Assume that for chain j ($j = 1, \ldots, m$), it has iterations $\{\theta_{ij} : i = 1, \ldots, n\}$. We define the following notations:

$$\begin{aligned}
\bar{\theta}_{\cdot j} &= \frac{1}{n}\sum_{i=1}^{n}\theta_{ij} \\
\bar{\theta}_{\cdot\cdot} &= \frac{1}{m}\sum_{j=1}^{m}\bar{\theta}_{\cdot j} \\
s_j^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(\theta_{ij} - \bar{\theta}_{\cdot j})^2 \\
B &= \frac{n}{m-1}\sum_{j=1}^{m}(\bar{\theta}_{\cdot j} - \bar{\theta}_{\cdot\cdot})^2 \\
W &= \frac{1}{m}\sum_{j=1}^{m}s_j^2 \quad .
\end{aligned}$$

B describes the variance between $\frac{m}{2}$ chains, and W summarizes within-chain variances. Therefore, we call B between-chain variance and call W within-chain variance. We can think that when multiple chains attains stationary and mix well with each other, B would be very close to W. Therefore, we introduce the *potential scale reduction* factor for assessing convergence of θ is

$$\hat{R} = \sqrt{\frac{n-1}{n} + \frac{1}{n}\frac{B}{W}}.$$

*If you have any question about the note, please send an email to xyluo@link.cuhk.edu.hk

When we sample from $f(\theta)$ using the MCMC algorithm and start with multiple initial values, we calculate the $\hat{R}$ for θ. If $\hat{R}$ is less than a value greater than and close to 1, say 1.1, stop the MCMC algorithm, and then collect last half of samples.

8 Permutation and Bootstrap

In this section, we investigate statistical problems under the frequentist framework.

8.1 Permutation Test

We have two groups of samples. The first group consist of $\{x_1, \ldots, x_m\}$ and samples in the second group are $\{y_1, \ldots, y_n\}$. Assume x_i $(i = 1, \ldots, m) \sim F$ and y_i $(i = 1, \ldots, n) \sim G$. μ_F and μ_G represent the means of F and G, respectively. We are interested in whether or not $\mu_F = \mu_G$ without specifying the parametric forms of F and G. Define the null hypothesis as $H_0 : \mu_F = \mu_G$. We will construct a process to obtain the p value for the hypothesis testing problem.

1 Calculate the observed mean difference *statistic* $T_{obs} = \frac{1}{m}\sum_{i=1}^{m} x_i - \frac{1}{n}\sum_{i=1}^{n} y_i$.

2 Redefine $\{x_1, \ldots, x_m, y_1, \ldots, y_n\}$ as $\{z_1, \ldots, z_m, z_{m+1}, \ldots, z_{m+n}\}$. That is to say, $x_i = z_i$ for $i = 1, \ldots, m$ and $y_i = z_{m+i}$ for $i = 1, \ldots, n$.

3 Given a permutation ρ of $\{1, 2, \ldots, m, m+1, \ldots, m+n\}$, we obtain a permutation set $\{z_{\rho(1)}, \ldots, z_{\rho(m)}, z_{\rho(m+1)}, \ldots, z_{\rho(m+n)}\}$, and calculate a new *statistic*: $T^* = \frac{1}{m}\sum_{i=1}^{m} z_{\rho(i)} - \frac{1}{n}\sum_{i=1}^{n} z_{\rho(m+i)}$.

4 Repeat the step 3 R times, resulting in $\{T_1^*, \ldots, T_R^*\}$. Calculate the p value as $\frac{1}{R}\sum_{j=1}^{R} I(|T_j^*| \geq |T_{obs}|)$.

Assume the significance level of the hypothesis testing is α. If $p < \alpha$, we reject the null hypothesis $H_0 : \mu_F = \mu_G$; otherwise, we accept H_0.

8.2 Bootstrap

We have a collection of samples $\{x_1, x_2, \ldots, x_m\}$. Assume x_i *i.i.d.* $\sim F(\cdot|\theta)$ and a known estimator $T(x_1, x_2, \ldots, x_m)$ can be used to approximate θ. We are interested in the characteristics of the estimator $T(x_1, x_2, \ldots, x_m)$, such as T's variance and bias. In the following, we give the procedures (called bootstrap) to approximate T's variance and bias.

1 Calculate the estimator $T(x_1, \ldots, x_m)$ based on the original dataset $\{x_1, x_2, \ldots, x_m\}$.

2 Randomly draw m samples from $\{x_1, x_2, \ldots, x_m\}$ with replacement, and we obtain $\{x_1^*, \ldots, x_m^*\}$.

3 Based on the bootstrap dataset $\{x_1^*, \ldots, x_m^*\}$, we calculate the $T^* := T(x_1^*, \ldots, x_m^*)$.

4 Repeat steps 2 and 3 B times, resulting in $T_1^*, \ldots, T_B^*$.

5 We estimate $Var(T(x_1, \ldots, x_m))$ by $\frac{1}{B-1} \sum_{i=1}^B (T_i^* - \frac{1}{B} \sum_{i=1}^B T_i^*)^2$, and estimate the bias of $T(x_1, \ldots, x_m)$ by $\frac{1}{B} \sum_{i=1}^B T_i^* - T(x_1, \ldots, x_m)$.