A New Proof of Peskun's and Tierney's Theorems using Matrix Method

Longhai Li

Department of Statistics

University of Toronto

Asymptotic Variance of an i.i.d. Monte Carlo Estimator

Suppose we want to evaluate the expectation μ of function $f(x)$ w.r.t. the distribution with p.d.f. $\pi(x)$. If we can draw i.i.d. samples $X_1, \cdots, X_n$ from $\pi(x)$, the Monte Carlo estimator of μ is given by

$$\hat{\mu}_n = \frac{\sum_{i=1}^{n} f(X_i)}{n}$$

The Central Limit Theorem says

$$\sqrt{n}(\hat{\mu}_n - \mu) \rightarrow^d N(0, \sigma^2(f))$$

where

$$\sigma^2(f) = \mathrm{Var}(f(X_i)) = \lim_{n \rightarrow +\infty} n\mathrm{Var}(\hat{\mu}_n)$$

Asymptotic Variance of an MCMC Estimator

When we can not draw i.i.d. samples, instead we obtain $X_1, \cdots, X_n$ from a Markov chain whose transition probability T has $\pi(x)$ as its unique stationary distribution. Under some regularity conditions, the Central Limit Theorem still holds, but the asymptotic variance is no longer σ^2. It is defined by

$$\sigma_T^2(f) \quad = \quad \lim_{n \to +\infty} n\mathrm{Var}(\hat{\mu}_n) = \sigma^2 + 2\sum_{i=1}^{+\infty} \mathrm{Cov}_T(f(X_0, X_i))$$

Different Markov chain samplers can be compared based on this asymptotic variance. T_1 is said to be not less efficient than T_2 if

$$\sigma_{T_1}^2(f) \leq \sigma_{T_2}^2(f), \text{ for all } f$$

Let denote above ordering regarding T_1 and T_2 by $T_1 \geq^e T_2$

Peskun's Theorem(1973)

The state space $\mathcal{X}$ is finite. Let T_1 and T_2 are two reversible, aperiodic, irreducible transition matrix w.r.t. π.

If

$$T_1(i,j) \geq T_2(i,j), \text{ for all } i \neq j \in \mathcal{X}$$

then $T_1 \geq^e T_2$

Tierney's Theorem(1998)

Tierney generalized Peskun's theorem to general state space. Also a weaker ordering was proposed:

Let T_1 and T_2 are two reversible, aperiodic, irreducible transition kernels w.r.t. π.

If

$$\mathrm{Cov}_{T_1}(f(X_0), f(X_1)) \leq \mathrm{Cov}_{T_2}(f(X_0), f(X_1)) \text{ for all } f$$

then $T_1 \geq^e T_2$.

This theorem says that for reversible Markov chains, that T_1 has smaller lag-1 correlation than T_2 is sufficient to imply that it has smaller asymptotic variance than T_2.

Existing Proofs

Peskun's method: Differentiate the asymptotic variance that is expressed in matrix form due to Kenney w.r.t. $T(i,j)$, then show that it is a decreasing function of $T(i,j)$.

Tierney's method: In general space, a reversible Markov chain corresponds to a self-joint operator. He resorted to the spectral decomposition of self-joint operators to proof the result.

Neal's method: Neal(2004) shows the Peskun's theorem with an intuitive and creative way by defining a block transition using the Delta difference of the element (i,j) between T_1 and T_2.

However, all of above methods are lengthy and make the theorems seem mysterious. Next I will show that, when the space is finite, the Tierney's theorem is actually just a simple corollary of a well-known fact regarding positive-definite matrix. Also I will show that Peskun's theorem is very easy to obtain from Tieney's theorem. The proof is shorter and more elementary than Tierney's.

Matrix Expression of Asymptotic Variance

Let T be the transition matrix, and let $(\pi_1, \cdots, \pi_s)$ be the invariant probability, where s is the number of states in $\mathcal{X}$. Let

$$B = \begin{pmatrix} \pi_1 & & & \\ & \pi_2 & & \\ & & \ddots & \\ & & & \pi_s \end{pmatrix} \qquad A = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_s \\ \pi_1 & \pi_2 & \cdots & \pi_s \\ & \cdots & \cdots & \\ \pi_1 & \pi_2 & \cdots & \pi_s \end{pmatrix} \tag{1}$$

Kenney(1963) shows that the asymptotic variance is given by

$$\sigma_T^2(f) = f' \left(2B[I - (T - A)]^{-1} - B + BA\right) f$$

where $f = (f_1, \cdots, f_s)'$. The inverse in above expression is guaranteed existing for finite Markov chain. Peskun's proof starts from this expression.

The Joint Probability Matrix for the Successive States

$$BT = \begin{pmatrix} \pi_1 T(1,1) & \pi_1 T(1,2) & \cdots & \pi_1 T(1,s) \\ \pi_2 T(2,1) & \pi_2 T(2,2) & \cdots & \pi_2 T(2,s) \\ & \cdots & \cdots & \\ \pi_s T(s,1) & \pi_s T(s,2) & \cdots & \pi_s T(s,s) \end{pmatrix}$$

The element (i, j) of BT is the joint probability of $(X_0 = i, X_1 = j)$, i.e. $P(X_0 = i, X_1 = j)$, where $X_0 \sim \pi$ and X_1 is the next state of X_0 on the Markov chain governed by T, i.e. $X_1 \,|X_0 = i \sim T(i, \cdot)$

New Proof of Tierney's theorem

1) T is reversible w.r.t. $\pi \iff BT$ is a symmetric matrix. BA is also symmetric. Therefore $B(I - T + A)$ is symmetric.

2) Suppose $X_0 \sim \pi$ and X_1 be the next state of X_0 on the Chain governed by T, then

$$\begin{aligned} f'BTf &= \sum_{i,j \in \mathcal{X}} f_i f_j \pi_i T(i,j) = E[f(X_0) f(X_1)] \\ f'Bf &= E(f(X_1)^2) = E(f(X_0)^2) \\ f'BAf &= \sum_{i,j \in \mathcal{X}} f_i \pi_i \pi_j \pi_j = (E(f(X_0)))^2 \geq 0 \end{aligned}$$

So for $f \neq 0$,

$$\begin{aligned} 2f'B(I - T + A)f &= E(f(X_0)^2) + E(f(X_1)^2) - 2E[f(X_0)f(X_1)] + (E(f(X_0)))^2 \\ &= E((f(X_0) - f(X_1))^2) + (E(f(X_0)))^2 > 0 \end{aligned}$$

From 1) and 2), it follows that $B(I - T + A)$ is positive definite matrix.

New Proof of Tierney's Theorem (Cont.)

3) Let T_1 and T_2 is two reversible transition matrices,

$$
\begin{aligned}
& \sigma^2_{T_1}(f) \le \sigma^2_{T_2}(f) \text{ for all } f \in R^s - \{\underline{0}\} \\
\iff \quad & (Bf)'(B - BT_1 + BA)^{-1}(Bf) \le (Bf)'(B - BT_2 + BA)^{-1}(Bf) \\
\iff \quad & (B - BT_1 + BA)^{-1} \le (B - BT_2 + BA)^{-1} \\
\iff \quad & B - BT_1 + BA \ge B - BT_2 + BA \\
\iff \quad & f'BT_1f \le f'BT_2f \text{ for all } f \in R^s - \{\underline{0}\} \\
\iff \quad & \mathrm{Cov}_{T_1}(f(X_0), f(X_1)) \le \mathrm{Cov}_{T_2}(f(X_0), f(X_1)) \text{ for all } f \in R^s - \{\underline{0}\}
\end{aligned}
$$

In the fourth line of above, we use a simple fact regarding the positive definite matrix: For $A > 0, B > 0$, $A \ge B \iff A^{-1} \le B^{-1}$

Proof of Peskun's Theorem

Note that

$$\begin{aligned} E_T\left((f(X_0)-f(X_1))^2\right) &= E_T\left(((f(X_0)-\mu)-(f(X_1)-\mu))^2\right) \\ &= 2\mathrm{Var}(f(X_0)) - 2\mathrm{Cov}_T\left(f(X_0), f(X_1)\right) \end{aligned}$$

Hence,

$$\begin{aligned} & \mathrm{Cov}_{T_1}\left(f(X_0), f(X_1)\right) \leq \mathrm{Cov}_{T_2}\left(f(X_0), f(X_1)\right) \\ \Longleftrightarrow\quad & E_{T_1}\left((f(X_0)-f(X_1))^2\right) \geq E_{T_2}\left((f(X_0)-f(X_1))^2\right) \\ \Longleftrightarrow\quad & \sum_{i,j\in\mathcal{X}} (f_i - f_j)^2 \pi_i T_1(i,j) \geq \sum_{i,j\in\mathcal{X}} (f_i - f_j)^2 \pi_i T_2(i,j) \\ \Longleftarrow\quad & T_1(i,j) \geq T_2(i,j), \text{ for all } i \neq j \in \mathcal{X} \end{aligned}$$

Discussions

- This proof is the simplest to my knowledge for these theorems restricted in finite Markov chains. It makes these theorems look quite straigtforward.
- My original purpose is to show similar things for nonreversible chains. This proof points out that the essential step for proving Tierney's theorem is to show $A > B \Rightarrow A^{-1} < B^{-1}$. When A and B are symmetric, correspondingly the Markov chains are reversible, this is true. We present a new topic for mathematical people to work on: How to relax the condition of symmetry to obtain the above conclusion?
- There might be similar proofs for Markov chains on general space using the generalized concepts in operators theory, like inverse operator, inner product.