Significance and Outliers in Markov Chains

Steven R. Dunbar

<2024-09-17 Tue>

Why Significance for Markov Chains?

Gerrymandering – Drawing political subdivision lines to advantage one party or group over another

Gerrymandered political divisions are "weird" or "unusual":

- Shape (Polsby-Popper measure)
- Political Party Distribution (efficiency gap, partisan symmetry, mean-median, others)
- Communities of interest (racial divisions)

Significance for Legislative Maps

To detect "weirdness" or "unusualness":

- Large sample of legal political subdivisions,
- Compare legislated subdivision to sample on some measure,
- Gerrymandered maps *should* be statistical outliers.

To generate a large sample, use Markov chain Monte Carlo.

Work of Chikina, Frieze, Mattingly and Pegden (2020).

Value Function

$\mathcal{X}$ is the state space of the chain.

Definition

A **value function**, is a function $v : \mathcal{X} \to \mathbb{R}$.

The value function has auxiliary information about the states of the chain and has no relationship to the transition probabilities of the Markov chain.

For statistical significance tests the relevant feature of the value function is the ranking on the elements of $\mathcal{X}$.

Outliers

Definition

A real number α_0 is an ϵ-**outlier** among $\alpha_0, \alpha_1, \dots \alpha_n$ (not necessarily in increasing order and with repetitions possible) if there are, at most, $\epsilon(n+1)$ indices i for which $\alpha_i \le \alpha_0$.

In other words $\operatorname{card}\left(\{\nu : \alpha_\nu \le \alpha_0\}\right) \le \epsilon(n+1)$.

Alternatively, α_0 is in the smallest ϵ-fraction of the $n+1$ ranked values of $\alpha_0, \alpha_1, \dots \alpha_n$.

Example

For example,

$$\alpha_0 = 3, \alpha_1 = 1, \alpha_2 = 4, \alpha_3 = 1, \alpha_4 = 5, \alpha_5 = 9,$$
$$\alpha_6 = 2, \alpha_7 = 6, \alpha_8 = 5, \alpha_9 = 3, \alpha_{10} = 5.$$

Then $\alpha_0 = 3$ is *not* an 0.1-outlier since $5 = \operatorname{card}\left(\{\nu : \alpha_\nu \le \alpha_0\}\right) \ge \epsilon(n+1) = 1.1$.

$\alpha_0 = 3$ *is not* a 0.2-outlier (the cardinality is 5, greater than 2.2) or even a 0.3-outlier.

This example says $\alpha_0 = 3$ has lesser/equal values more common than small fractions of the ordered values.

Among-ℓ smallest

Definition

For $1 \le \ell \le n+1$ say α_j is **among-ℓ-smallest** in $\alpha_0, \dots \alpha_n$ if there are at most ℓ indices from $0, \dots, n$ such that α_ν is at most as large as α_j. That is, α_j is among-ℓ-smallest in $\alpha_0, \dots \alpha_n$ if

$$\mathrm{card}\left(\{\nu : \alpha_\nu \le \alpha_j\}\right) \le \ell.$$

α_j is among-1-smallest in $\alpha_0, \dots \alpha_n$ when its value is the *unique minimum value*.

Connections

- If α_0 is an ϵ-outlier, then α_0 is among-$\lfloor \epsilon(n+1) \rfloor$-smallest.
- If α_0 is among-ℓ-smallest, then α_0 is an $\frac{\ell}{n+1}$-outlier.
- "Among-ℓ-smallest" is convenient for proofs.
- "ϵ-outlier" is convenient for statistical testing.

Median, Percentile and Rank

- ϵ-outlier is a generalization (or variant) of "median", "quartile", "percentile" and also "rank"
- With the median, repetitions of median *do not* matter. However, with the $1/2$-outlier, repetitions or equality *do* matter.
- To be an outlier, values $\leq \alpha_0$ appear infrequently. Repetitions of α_0 work against outlier status.
- For a set with duplicates (or ties), rank is not well-defined since there are several methods of ranking the ties.

Notation for Proof

Definition

Let n be fixed. Let $X_1, X_2, \dots, X_n$ be a sample of the Markov chain starting from $X_0 = \sigma_0$. Suppose the state space has a value function $v : \mathcal{X} \to \mathbb{R}$. For $0 \le j \le n$ define

$$\rho(j; \ell, n) = \mathbb{P}\left[v(X_j) \text{ is among-}\ell\text{-smallest in } v(X_0), \dots, v(X_n)\right]$$

and

$$\rho(j; \ell, n \mid \sigma) = \mathbb{P}\left[v(X_j) \text{ is among-}\ell\text{-smallest in } v(X_0), \dots, v(X_n) \mid X_j = \sigma\right].$$

Special Case

$\rho(j; 1, n)$ is the probability X_j has the *unique minimum value* in the Markov chain of length n.

Likewise, $\rho(j; 1, n \mid \sigma)$ is the probability $X_j = \sigma$ has the *unique minimum value* in the Markov chain of length n given that the Markov chain passes through σ at step j.

Proposition

1. Fix n.
2. Suppose X_0 is chosen from the stationary distribution π of the reversible Markov chain X.
3. Suppose $X_1, X_2, \dots, X_n$ is a sample of the Markov chain starting from X_0.
4. Suppose the state space has a value function $v : \mathcal{X} \to \mathbb{R}$.

Then

$$\mathbb{P}\left[v(X_0) \text{ is among-}\ell\text{-smallest in } v(X_0), \dots, v(X_n)\right]$$

is

$$\rho(0; \ell, n) \le \sqrt{\frac{2\ell - 1}{n + 1}}.$$

Example: Ehrenfest urn Model

Two urns, labeled A and B, contain a total of N balls. At each step a ball is selected at random with all selections equally likely. Then an urn is selected, urn A with probability $p = \frac{1}{2}$ and urn B with probability $q = 1 - p = \frac{1}{2}$ and the ball is moved to that urn. The state at each step is the number of balls in the urn A, from 0 to N. Start from state σ selected at random from the stationary (binomial) distribution.

For this simple example, let the value function be the number of balls in urn A, so that the value function is identical with the state number. (The simplicity of the alternative Ehrenfest urn model limits the possibilities for value functions.)

Example: Ehrenfest urn model

Note: 0 is a "weird" or "unusual" value (or state).

With $N = 21$ balls, Markov chain runs of 399 steps in 400 trials the proportion showing 0 is an $.01$-outlier is about 0.92.

With longer Markov chain runs of 800 or 1600 steps in 1000 trials, the proportion showing 0 is an 0.01-outlier is better than 0.97.

With $N = 51$ balls, the probability of showing state 0 is unusual in long runs is virtually certain.

Corollary: The $\sqrt{2\epsilon}$ Test

Corollary (The $\sqrt{2\epsilon}$ Test)

1. *Fix* n.
2. *Suppose* X_0 *is chosen from the stationary distribution* π *of the reversible Markov chain* X.
3. *Suppose* $X_1, X_2, \dots, X_n$ *is a sample of the Markov chain starting from* X_0.
4. *Suppose the state space has a value function* $v : \mathcal{X} \to \mathbb{R}$.

Then the event $v(X_0)$ *is an* ϵ*-outlier among* $v(x_0), \dots v(x_n)$ *with probability* $p = \sqrt{2\epsilon}$.

Application

In the general setting of a reversible Markov chain, the Corollary leads to a simple quantitative procedure for asserting rigorously σ_0 is atypical with respect to π without knowing the mixing time of X: simply observe a random trajectory $\sigma_0 = X_0, X_1, X_2, \dots, X_n$ from σ_0 for any fixed n. If $v(\sigma_0)$ is an ϵ-outlier among $v(X_0), v(X_1), \dots, v(X_n)$, then the probability of *erroneously* concluding X_0 is *not unusual*, when in fact the null hypothesis that X_0 is not unusual is not true, is less than $\sqrt{2\epsilon}$.

Improvements

- With serial Markov chain sampling, $O(\sqrt{\epsilon})$ significance is the best possible.
- With *parallel* Markov chains sampling, significance at 2ϵ is possible.
- With *randomized parallel* Markov chains sampling, statistical power and large-deviation-style results are possible.

Besag and Clifford, 1989

Proof Sketch (part 1)

For simplicity, omit the value function in the proof and use the $\ell = 1$ case, so the unique minimum value is the focus.

1. Let π denote the stationary distribution for X and suppose the initial state X_0 is distributed as $X_0 \sim \pi$, so that in fact, $X_\nu \sim \pi$ for all ν.
2. Because $X_s \sim \pi$ for all s, stationarity implies

$$\rho(j; 1, n \mid \sigma) = \mathbb{P}\left[X_{s+j} \text{ is among-1-smallest in } X_s, \ldots, X_{s+n} \mid X_j = \sigma\right]$$

Proof Sketch (part 2)

3. Because the chain is stationary and reversible, the probability of $(X_0, \dots, X_n) = (\sigma_0, \dots, \sigma_n)$ equals the probability $(X_0, \dots, X_n) = (\sigma_n, \dots, \sigma_0)$ for any fixed sequence $(\sigma_0, \dots, \sigma_n)$.
4. Thus, any sequence $(\sigma_0, \dots, \sigma_n)$ for which $\sigma_j = \sigma$ and σ_j is a among-1-smallest corresponds to an equiprobable sequence $(\sigma_n, \dots, \sigma_0)$, for which $\sigma_{n-j} = \sigma$ and σ_{n-j} is among-1-smallest.
5. Thus:

$$\rho(j; 1, n \mid \sigma) = \rho(n - j; 1, n \mid \sigma).$$

Proof Sketch (part 3)

6. The next claim is:

   $$\rho(j; 2\cdot 1 - 1, n \mid \sigma) \ge \rho(j; 1, j \mid \sigma) \cdot \rho(0; 1, n-j \mid \sigma).$$

   Proof: The first term $\rho(j; 1, j \mid \sigma)$ is the probability that X_j is among-1-smallest of the $j+1$ values from X_0 to X_j.
   The second term $\rho(0; 1, n-j \mid \sigma)$ is the probability that X_j is among-1-smallest of the $n-j+1$ values from X_j to X_n.

Proof Sketch (part 4)

7. Continuing proof that

$$\rho(j; 2\cdot 1 - 1, n \mid \sigma) \geq \rho(j; 1, j \mid \sigma) \cdot \rho(0; 1, n-j \mid \sigma).$$

These events are conditionally independent when conditioning on the value of $X_j = \sigma$, which allows taking the product. When both of these events happen, then X_j is among-$2 \cdot 1 - 1$-smallest in $X_0, \dots, X_n$, with $2 \cdot 1 - 1$ because X_j was counted in each event. Both events happening is a subevent of the desired event of X_j being among-$2 \cdot 1 - 1$-smallest from X_0 to X_n.

Proof Sketch (part 5)

8. Using the previous steps

$$\begin{aligned}\rho(j; 2\cdot 1 - 1, n \mid \sigma) &\geq \rho(j; 1, j \mid \sigma)\cdot\rho(0; 1, n-j \mid \sigma)\\ &= \rho(0; 1, j \mid \sigma)\cdot\rho(0; 1, n-j \mid \sigma)\\ &\geq \left(\rho(0; 1, n \mid \sigma)\right)^2 .\end{aligned}$$

where the first inequality comes from steps 6 and 7 The middle equality comes from step 5. The last inequality comes from the simple observation that $\rho(j; \ell, n \mid \sigma)$ is nonincreasing in n for fixed j, ℓ and σ, in particular for j fixed to be 0.

Proof Sketch (part 6)

9. Now $\rho(j;\ell,n) = \mathbb{E}\left[\rho(j;\ell,n \mid \sigma)\right]$, where the expectation is taken over the random choice of $\sigma \sim \pi$. Thus, taking expectations over the inequality in step

$$\begin{aligned} \rho(j;2\ell-1,n) &= \mathbb{E}\left[\rho(j;2\ell-1,n \mid \sigma)\right] \\ &\geq \mathbb{E}\left[\rho(j;\ell,n \mid \sigma)^2\right] \\ &\geq \left(\mathbb{E}\left[\rho(0,\ell,n \mid \sigma)\right]\right)^2 \\ &= \left(\rho(0;\ell,n)\right)^2 \end{aligned}$$

where the second of the two inequalities is the Cauchy-Schwartz inequality.

Proof Sketch (part 7)

10. To complete the proof, sum the left- and right-hand sides of the inequality in step 9

$$\sum_{j=0}^{n} \rho(j; 2\ell - 1, n) \geq (n+1)(\rho(0; \ell, n))^2.$$

11. Letting $\mathbf{1}_j$, $0 \leq i \leq n$ be the indicator variable that is 1 whenever X_j is among-$2\ell - 1$-smallest in $X_0, \dots, X_n$, then $\sum_{\nu=0}^{n} \mathbf{1}_\nu$ is the number of among-$2\ell - 1$-smallest terms, which is at most 2ℓ. Therefore, linearity of expectation gives

$$2\ell - 1 \geq (n+1)(\rho(0; l, n))^2$$

so

$$\rho(0; \ell, n) \leq \sqrt{\frac{2\ell - 1}{n+1}}.$$