Generalization of Conditions to Starting Point Cutoff Phenomenon

Tianyi Pan

Abstract

In this report I generalize the result from [3] to show that for a fixed discrete-time regular Markov chain, cutoff phenomenon exists for unbounded steps under certain distribution requirements. I then gives several concrete distributional examples and discussion further directions.

1 Introduction

In the study of stochastic process, cutoff phenomenon represents the abrupt change of the total variation distance between the transition kernel and its stationary distribution for some Markov chains. Theorem 2.6 in [3] describes the equivalent condition of this phenomenon to the concentration of stopping time on some sets. In this report I generalize to unbounded steps in the transition of Markov chains, which establish more conclusions and tools to use in order to show the cutoff for Markov chains.

In Section 2 I introduce the concepts and notations. In Section 3, Theorem 3.2 is established through Lemma 3.2 and more refined conditions on $\mathbb{R}$ are built through Proposition 3.3 and Remark 3.4. Examples of common distributions with cutoff are shown in the end of this section. In Section 4 a counter-example is illustrated and a potential further direction is discussed.

2 Definitions

This report adapts notations from [3]. Let P be a ϕ-irreducible, aperiodic Markov chain with stationary distribution π on state space $(\mathcal{X}, \mathcal{B})$. For any $S \in \mathcal{B}$ define the stopping time τ_S on S

$$\tau_S = \inf \{t \geq 0 \mid X_t \in S\}$$

I also assume that P is regular, i.e. $\forall A \in \{A \in \mathcal{B} \mid \pi(A) > 0\}, \forall x \in \mathcal{X}, \mathbb{E}_x[\tau_A] < \infty$.

Definition 2.1: For any $x \in \mathcal{X}$ and $t \in \mathbb{R}^+$,

$$d_x(t) = \left\| P^{\lfloor t \rfloor}(x, \cdot) - \pi \right\|_{\mathrm{TV}}$$

which is the total variation distance between the transition kernel at $\lfloor t \rfloor$ time (starting from x) and the stationary distribution π.

Definition 2.2: Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of starting points in $\mathcal{X}$ and let $\{t_n\}_{n\in\mathbb{N}}$ be increasing sequences of positive real numbers s.t. $\lim_{n\to\infty} t_n = \infty$. P has **starting point cutoff** at time t_n starting from x_n if

(i) $\forall c \in (0, 1)$,

$$\lim_{n\to\infty} d_{x_n}(ct_n) = 1$$

(ii) $\forall c > 1$,

$$\lim_{n\to\infty} d_{x_n}(ct_n) = 0$$

Remark 2.3: According to Theorem 2.6 in [3], the existence of cutoff phenomenon is equivalent to the concentration of stopping time to any regular set S. That is,

$$\lim_{n\to\infty} \mathbb{P}_{x_n}\left(\left| \frac{\tau_S}{t_n} - 1 \right| \leq \varepsilon \right) = 1, \forall \varepsilon > 0$$

Therefore, an analysis of the concentration of τ_S would be sufficient to conclude the presence of cutoff.

In this report, I will explore unbounded steps, which are special types of random variables that are though unbounded, more concentrated around its mean.

Definition 2.3: A centered random variable W is **sub-Exponential** if there exists $K_5 > 0$ s.t.

$$\mathbb{E}[e^{\lambda W}] \leq e^{\lambda^2 K_5^2}, \forall |\lambda| \leq \frac{1}{K_5}$$

and denote $\mathcal{E}(K_5)$ as the set of random variables that satisfy the above condition.

The equivalences of conditions for random variables being sub-Exponential are discussed in [7] and will be applied later in Section 3 to derive the desired result.

3 Conditions for Cutoff Phenomenon

In Theorem 5.3 in [3], martingale difference method is applied on the martingale $\{M_t\}_{t=0}^T$ defined by

$$M_t = \begin{cases} \mathbb{E}_{X_t}[\tau_S] + t & \text{if } t < \tau_S \\ \tau_S & \text{if } t \geq \tau_S \end{cases}$$

and $T = \lfloor (1+\varepsilon) E_{x_n}[\tau_S] \rfloor$.

The author shows that

$$\mathbb{P}(|\tau_S - \mathbb{E}_x[\tau_S]| > \varepsilon E_x[\tau_S]) \leq \mathbb{P}(|M_T - M_0| > \varepsilon E_x[\tau_S])$$

and bounding the concentration of the martingale differences through Azuma-Hoeffding inequality is sufficient to make the conclusion. Here I adapt the same approach.

Lemma 3.1: Suppose there exists K_5, for all $x \in \mathcal{X} \setminus S$,

$$|\mathbb{E}_{X_1}[\tau_S] - \mathbb{E}_x[\tau_S]| \in \mathcal{E}(K_5)$$

Then for any $x \in \mathcal{X}$ and $\varepsilon > 0$,

$$\mathbb{P}_x(|\tau_S - \mathbb{E}_x[\tau_S]| \geq \varepsilon \mathbb{E}_x[\tau_S]) \leq \min\left\{2\exp\left(-\frac{\varepsilon^2 \mathbb{E}_x[\tau_S]}{2(1+\varepsilon)K_6^2}\right), 2\exp\left(-K_7 \varepsilon \mathbb{E}_x[\tau_S]\right)\right\}$$

for some constant K_6, K_7.

Proof: For the martingale $\{M_t\}$ defined above, if $X_{t-1} \in S$, then $M_t - M_{t-1} = 0$; else,

$$|M_t - M_{t-1}| = |\mathbb{E}_{X_1}[\tau_S] - \mathbb{E}_x[\tau_S] + 1| \leq 1 + |\mathbb{E}_{X_1}[\tau_S] - \mathbb{E}_x[\tau_S]|$$

Let $W = |\mathbb{E}_{X_1}[\tau_S] - \mathbb{E}_x[\tau_S]| \in \mathcal{E}(K_5)$, then

$$\begin{aligned} \|1+W\|_{L_p} &= \mathbb{E}[|1+W|^p]^{1/p} \\ &\leq \mathbb{E}[|1|^p]^{1/p} + \mathbb{E}[|W|^p]^{1/p} && \text{(Minkowski's Inequality)} \\ &\leq 1 + K_5 p && \text{(Proposition 2.7.1 in [7])} \\ &\leq (K_5+1)p && (p \geq 1) \end{aligned}$$

which implies $1+W$ is sub-Exponential. The monotonicity between the martingale difference and $1+W$, with one equivalent definition of sub-Exponential random variable in [7], shows that it is also sub-Exponential. More precisely, $\forall t \geq 0$

$$\begin{aligned} \mathbb{P}(|M_t - M_{t-1}| \geq t) &\leq P(1+W \geq t) \\ &\leq 2\exp(-t/K_1) \end{aligned}$$

By the footnote 6 in Proposition 2.7.1 in [7], with some additional efforts added in its proof, one can derive that all K's in the equivalent definitions of sub-Exponential random variables differ by at most a constant factor. In this sense,

$$\mathbb{E}[e^{\lambda(M_t - M_{t-1})} \mid \mathcal{F}_{t-1}] \leq e^{\lambda^2 K_6^2}, \forall |\lambda| \leq \frac{1}{K_7}$$

Therefore, by Theorem 6.8 from [5], which provides a sub-Exponential version of Azuma's inequality, one derives the desired result. □

Theorem 3.2: Suppose there exists K_5, for all $x \in \mathcal{X} \setminus S$,

$$|\mathbb{E}_{X_1}[\tau_S] - \mathbb{E}_x[\tau_S]| \in \mathcal{E}(K_5)$$

Then for any $\{x_n\}_{n\in\mathbb{N}}$ s.t.

$$\lim_{n\to\infty} \mathbb{E}_{x_n}[\tau_S] = \infty$$

there is cutoff phenomenon with $t_n = \mathbb{E}_{x_n}[\tau_S]$.

Proof: From the result in Lemma 3.1, takes $n \to \infty$, then $\mathbb{E}_{x_n}[\tau_S] \to \infty$ and the probability is squeezed to 0 in either entry in the minimum operator. Then Remark 2.3 gives the desired formulation of the cutoff. □

Here are some concrete realizations of the relatively abstract conditions in Theorem 3.2 under a more restricted setting. Let $\{W_t\}_{t\in\mathbb{N}}$ be i.i.d. $\mathbb{R}$-valued sub-Exponential random variables. Let the Markov chain be random walks with each step distribution W_t. Let $S = [-\infty, 0)$.

Proposition 3.3: Assume $W_t \overset{i.i.d}{\sim} W$, and for all $y > 0$,

$$\mathbb{E}[W \mid W < -y] + y \geq -C \tag{$*$}$$

for some constant C, then this chain exhibits starting point cutoff for $\{x_n\}_{n\in\mathbb{N}}$ as starting point.

Proof: Let $S_t = \sum_{i=0}^{t} W_i$. Let $\mu = \mathbb{E}[W]$. Wald's equation gives

$$\mathbb{E}_x[\tau] = \frac{\mathbb{E}_x[S_\tau]}{\mu}$$

If $\mathbb{E}[W \mid W < -y] + y \geq -C$, then

$$\begin{aligned}
\mathbb{E}_x[x + S_\tau] &= \mathbb{E}_x\left[\mathbb{E}[x + S_\tau \mid S_{\tau-1} = y]\right] \\
&= \mathbb{E}_x\left[\mathbb{E}[y + W_1 \mid \tau = 1]\right] && \text{(Strong Markov Property)} \\
&= \mathbb{E}_x\left[\mathbb{E}[W \mid W < -y] + y\right] \\
&\geq -C
\end{aligned}$$

Then

$$-x - C \leq \mathbb{E}_x[S_\tau] \leq -x$$

Therefore, after some algebraic manipulations combined with Wald's equation, one can get

$$|\mathbb{E}_{X_1}[\tau_S] - \mathbb{E}_x[\tau_S]| \leq \frac{C}{|\mu|} + \frac{|W_1|}{|\mu|}$$

Since translation and scaling keeps sub-Exponential property and taking absolute value keeps the random variable sub-Exponential by satisfying the first definition in Theorem 2.7.1 in [7], one get that $|\mathbb{E}_{X_1}[\tau_S] - \mathbb{E}_x[\tau_S]|$ is sub-Exponential. Then Theorem 3.2 provides the conditions needed for cutoff.

Remark 3.4: For equation $(*)$, note that for translation of distribution $W' = W - c, c \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}[W' \mid W' < -y] + y &= \mathbb{E}[W - c \mid W - c < -y] + y \\
&= \mathbb{E}[W \mid W < -(y - c)] + (y - c) \\
&\geq -C
\end{aligned}$$

for $y \geq c$, and is bounded for $y \in [0, c]$, so should be bounded on $y \in \mathbb{R}^+$.

Also, for scaling of distribution $W' = aW$,

$$\begin{aligned}
\mathbb{E}[W' \mid W' < -y] + y &= \mathbb{E}[aW \mid aW < -y] + y \\
&= a\mathbb{E}[W \mid W < -y/a] + a(y/a) \\
&\geq -aC
\end{aligned}$$

Hence condition ($*$), or cutoff phenomenon, is satisfied as long as one of the distributions in the "translation and scaling family" is satisfied.

Remark 3.5: Let F denote the Cumulative Density Function (cdf) of step W. Then

$$\begin{aligned}
\mathbb{E}[W' \mid W' < -y] + y &= \frac{\int_{-\infty}^{-y} w dF}{\int_{-\infty}^{-y} w dF} - y \\
&= \frac{\int_{-\infty}^{-y} (w+y)) dF}{\int_{-\infty}^{-y} w dF} \\
&= \frac{F(w)(w+y) \mid_{-\infty}^{-y} - \int_{-\infty}^{-y} F(w) dw}{F(-y)} \\
&= -\frac{\int_{-\infty}^{-y} F(w) dw}{F(-y)} \\
&= -\frac{\int_{y}^{\infty} F(-w) dw}{F(-y)} \qquad (**)
\end{aligned}$$

Hence the analysis of equation ($*$) can be done through the analysis on equation ($**$), i.e. on the behavior of tail bounds.

Remark 3.6: For random walks on $\mathbb{R}^n$ one can pick any direction that is feasible and easier to work with, then project the walk on the 1-dimensional subspace. For example, multidimensional normal distribution random walk also have cutoff phenomenon, since one can always project in any subspace and it keeps the normality, and then follow the derivation of standard normal distribution in Example 3.7. Accordingly, one would generalize the definition of sub-Exponential property to n-dimensions. One would apply the n-dimensional version of moment generating function to prove the Azuma's inequality with the same metric embedded in λ-space defined for sub-Exponential random variables.

Example 3.7:

(i) For $F(-w) = w^k e^{-w}, k \in \mathbb{N}$, condition ($**$) is satisfied by doing integration by parts directly on the numerator and note that the expression will converge to 1 as $y \to \infty$:

$$\begin{aligned}
-\frac{\int_{-\infty}^{-y} w^k e^{-w} dw}{w^k e^{-w}} &= -\frac{e^{-w}(w^k + P_{k-1}(w))}{w^k e^{-w}} \qquad (\deg P_{k-1} = k-1) \\
&\to -1
\end{aligned}$$

as $w \to \infty$.

(ii) For $F(-w) = P(w)e^{-w}$, where $P(w)$ is a polynomial, note that the highest degree term will dominate when y large, so the expression in condition ($**$) will also be bounded.

(iii) For $F(-w) = \Phi(-w)$, where Φ is the cdf of standard normal distribution, since the normal distribution is symmetric, I can consider the right tail. Let $f(y) = \mathbb{E}(W|W > y) - y$, then

$$f'(y) = \frac{\phi}{(1-\Phi)^2}\left[\phi - y(1-\Phi)\right] - 1$$

where ϕ denote the probability density function (pdf) of standard normal distribution. Note that using integration by parts with $u = \frac{1}{w}$ and $dv = we^{-w^2/2}dw$ (the conclusion draws from [6])

$$\begin{aligned}
\int_y^\infty e^{-w^2/2} dw &= -\frac{e^{-w^2/2}}{w}\Big|_y^\infty - \int_y^\infty \frac{e^{-w^2/2}}{w^2} dw \qquad \text{(Integration by Parts)} \\
&= \frac{e^{-y^2/2}}{y} - \int_y^\infty \frac{e^{-w^2/2}}{w^2} dw
\end{aligned}$$

Again applying integration by parts with $u = \frac{1}{w^3}$ and $dv = we^{-w^2/2}dw$, one gets

$$\begin{aligned}\int_y^\infty \frac{e^{-w^2/2}}{w^2}dw &= -\frac{e^{-w^2/2}}{w^3}\Big|_y^\infty - \int_y^\infty -e^{-w^2/2}\frac{-3}{w^4}dw \\ &= \frac{e^{-y^2/2}}{y^3} - \int_y^\infty \frac{3e^{-w^2/2}}{w^4}dw\end{aligned}$$

Hence, after scaling $\frac{1}{\sqrt{2\pi}}$, I get

$$1 - \Phi \geq \frac{\phi}{x}(1 - \frac{1}{x^2})$$

Further,

$$\begin{aligned}f'(y) &\leq \phi \cdot \frac{y^2}{\phi^2}(1 - \frac{1}{y^2})^{-2} \cdot (\phi - y \cdot \frac{\phi}{y}(1 - \frac{1}{y^2})) - 1 \\ &= (1 - \frac{1}{y^2})^{-2} - 1 \\ &= \frac{2y^2 - 1}{(y^2 - 1)^2} \\ &\leq \frac{3}{y^2} \qquad \text{(for } y \text{ large)}\end{aligned}$$

Then one can integrate f' by the Fundamental Theorem of Calculus and show that condition f is bounded, thus condition $(*)$ is satisfied.

4 Further Discussions

4.1 A Counter-Example to Condition $(**)$ for Sub-Exponential Random Variables

Consider A cdf $F(-w)$ that values εe^{-y_0} for $\varepsilon << 1$ at y_0 and keeps constant as much as it can. Assume it is sub-Exponential and symmetric, thus $F(-w) \leq e^{-w}$. Then the constant value can "sustain" for $-\ln \varepsilon$ until it hits the bound of exponential. Then considering condition $(**)$, we find that

$$\begin{aligned}-\frac{\int_{y_0}^\infty F(-w)dw}{F(-y_0)} &\leq -\frac{\int_{y_0}^{y_0 - \ln \varepsilon} \varepsilon e^{-y_0} dw}{\varepsilon e^{-y_0}} \\ &= -\ln \varepsilon\end{aligned}$$

Since ε can be taken arbitrarily small, $\ln \varepsilon$ will go to infinity and thus there is not a uniform lower bound. Hence, there exists sub-Exponential random variables that fails to satisfy condition $(**)$.

4.2 Drift Theorems in Proving the Cutoff

The key in deriving the cutoff for sub-Exponential martingale difference is to show that the stopping state, $\mathbb{E}[x + S_\tau]$, is bounded from below. Condition $(*)$ use iterated expectation to realize this condition. One can consider a more macro view, i.e. apply variants of optional stopping theorem on sub-/super-martingale to bound the expectation. Drift theorems are used to analyze the expectation and concentration of stopping time. By Wald's equation, one can convert the problem in space domain into a problem in time domain.

By theorem 3.2 (i) in [4], one can construct non-negative function g to restrict the general stochastic process to a non-negative chain, thus able to apply Optional Stopping Theorem on sub-martingale. However, it turns out that $g(x) = \max\{0, x\}$ turns the condition

$$\mathbb{E}[g(X_t) - g(X_{t+1}) - \alpha_u; X_t > 0 \mid \mathcal{F}_t] \geq 0, \forall t \in \mathbb{N}, \alpha_u > 0$$

exactly into condition $(**)$.

On the other hand, if one restricts the space domain to $[-C, \infty)$ and apply Theorem 7 in [2], which requires the new chain to satisfy

$$X_t - \mathbb{E}[X_{t+1} \mid \mathcal{F}_t] \geq \delta$$

For random walks, $\delta = -\mathbb{E}[W \mid W \geq -C] < -\mathbb{E}[W]$, so for unbounded distribution W, δ converges to what we want (i.e. $-\mathbb{E}[W]$) but never touches it. If we consider $\mathbb{E}_x[\tau]$ as a function of $x \in \mathbb{R}$, then it is under a set of lines whose slope is arbitrarily close to the desired value. However, from there we cannot make any conclusion: For example, the function $f(x) = x + \sqrt{x}$ has the tangent line slope close to 1 when x large, but it is not under any $x + C$ for C constant since $\sqrt{x}$ goes to infinity.

Furthermore, Theorem 5.5 in [1] gives concentration upper bounds for the stopping time. However, it only applies when stopping time is larger than $2n/\delta$, where the desired cutoff would be n/δ and would like to apply the tail sum formula of expectation to bound the stopping time.

However, applying drift theorems is a feasible direction to work on. To do this, one would develop more general versions that allow unbounded state space from below. Also, one could draw ideas from multiplicative and variable drifts to fit into the context of this problem. **??**

References

[1] Timo Kötzing. Theory of stochastic drift. *arXiv preprint arXiv:2406.14589*, 2024.

[2] Timo Kötzing and Martin S Krejca. First-hitting times under drift. *Theoretical Computer Science*, 796:51–69, 2019.

[3] David Ledvinka and Jeffrey S. Rosenthal. Equivalence of starting point cutoff and the concentration of hitting times on a general state space, 2023. `https://probability.ca/jeff/ftpdir/Cutoff.pdf` [Accessed: 2024-8-28].

[4] Per Kristian Lehre and Carsten Witt. Tail bounds on hitting times of randomized search heuristics using variable drift analysis. *Combinatorics, Probability and Computing*, 30(4):550–569, 2021.

[5] Alessandro Rinaldo. Advanced statistical theory lecture 6: Sept 19, 2018. `https://www.stat.cmu.edu/~arinaldo/Teaching/36710/F18/Scribed_Lectures/Sep19.pdf` [Accessed: 2024-8-29].

[6] Mike Spivey. Tail bounds of the normal distribution, 2011. `https://mikespivey.wordpress.com/2011/10/21/normaltails/` [Accessed: 2024-8-30].

[7] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.