Two Examples of COM Bounds using Spectral Gaps: Length of the LIS in a Random Permutation and Lipschitz Functions of 1d Markov Chains

Michael A. Frölich\textsuperscript{1} and Shannon Starr \textsuperscript{2}

\textsuperscript{1} Department of Anesthesiology and Perioperative Medicine, School of Medicine, University of Alabama at Birmingham (UAB)
\textsuperscript{2} Department of Applied Mathematics, UAB, Birmingham, AL 35294–1170

January 4, 2018

Abstract

We consider two examples for a well-known method for obtaining concentration of measure (COM) bounds for a given observable in a given measure. The method is to consider an auxiliary Markov chain for which the invariant distribution is the measure of interest. Then one obtains COM bounds involving two quantities. The first is the spectral gap of the Markov transition matrix. The second is an appropriate Lipschitz constant for the observable of interest with respect to 1 step of the Markov chain.

We consider two examples of the basic method. The first is to obtain rough COM bounds for the length of the longest increasing subsequence (LIS) in a uniform random permutation. The bounds are similar to well-known bounds of Talagrand using his isoperimetric inequality.

The second example is to consider a 1d Markov chain: $X_0, X_1, \ldots, X_n$. We assume the invariant measure for the chain $\mu$ is reversible, and let the initial distribution of $X_0$ be $\mu$. Then the observable of interest is any function $f(X_0, X_1, \ldots, X_n)$, which is Lipschitz with respect to replacement of single variables. One case of this is “target frequency analysis,” which is of interest in biostatistics.

The auxiliary Markov chain is Glauber dynamics which is gapped in 1d.

1 Statement of the General Method

This article is about obtaining concentration of measure (COM) bounds for certain observables in given measures.

For the present article, suppose that $\mathcal{X}$ is a fixed, finite set. Suppose we are interested in COM bounds for an observable, by which we mean a real-valued function $f$ on $\mathcal{X}$. And we are interested in COM bounds for the observable $f$ relative to a non-degenerate measure $\mu$ on $\mathcal{X}$. Let $\mathcal{M}_{+1}(\mathcal{X})$ denote the set of all probability measures $\mu$ on $\mathcal{X}$. (So we have $\forall x \in \mathcal{X}$ that $\mu(x) \geq 0$ and we also have $\sum_{x \in \mathcal{X}} \mu(x) = 1.$) Let $\mathcal{C}(\mathcal{X})$ denote the set of all real-valued functions $f : \mathcal{X} \to \mathbb{R}$.

Given a pair $(\mu, f)$ from $\mathcal{M}_{+1}(\mathcal{X}) \times \mathcal{C}(\mathcal{X})$, let us define the positive and negative fluctuations as

$$F^+_\mu(f; a) := \mu(\{x \in \mathcal{X} : f(x) - E_\mu[f] \geq a\}) \quad \text{and} \quad F^-_\mu(f; a) := F^+_\mu(-f; a),$$

for $a \in [0, \infty)$. We are interested in concentration of measure bounds which are bounds on $F^\mu_+(f; \cdot)$ for a particular pair $(\mu, f)$. In this article, we will sometimes derive different types of bounds for $F^\mu_+(f; \cdot)$ than for $F^-_\mu(f; \cdot)$. But if we have sufficiently good bounds for one, then that leads to (possibly weaker) bounds for the other one by Markov’s inequality, using the fact that $\int_0^\infty F^\mu_+(f; t) \, dt = \int_0^\infty F^-_\mu(f; t) \, dt$. 


In this setting, Aida and Stroock described a useful elementary method for obtaining such bounds, using Chebyshev’s inequality \([1]\). They did this along the way to proving even more sophisticated bounds, but for the present article we focus on their first, elementary method.

Suppose one has a Markov chain given by transition matrix \(P : \mathcal{X} \times \mathcal{X} \to \mathbb{R}\), having the property that \(\mu\) is stationary for \(P\). So

\[
\mu \cdot P = \mu, \text{ that is: } \forall x \in \mathcal{X}, \text{ we have } \mu(x) = \sum_{y \in \mathcal{X}} \mu(y)P(y, x).
\] (2)

Also, suppose that moreover \(\mu\) is reversible for \(P\), meaning

\[
\forall x, y \in \mathcal{X}, \text{ we have } \mu(x)P(x, y) = \mu(y)P(y, x).
\] (3)

(The condition (3) implies (2), of course, since \(P\) satisfies \(P(x, y) \geq 0\) for all \(x, y \in \mathcal{X}\) and \(\sum_{y \in \mathcal{X}} P(x, y) = 1\) for all \(x \in \mathcal{X}\).) Then Aida and Stroock used the Markov chain to derive COM bounds in the measure \(\mu\).

Denote the variance in the measure \(\mu\) as \(\text{Var}_\mu(f) = \sum_{x \in \Omega} (f(x) - E_\mu[f])^2 \mu(x)\), as usual. The Dirichlet form for \(\mu\), relative to \(\mu\), is denoted by \(\mathcal{E}_{P,\mu} : \mathcal{C}(\mathcal{X}) \times \mathcal{C}(\mathcal{X}) \to \mathbb{R}\), defined as

\[
\mathcal{E}_{P,\mu}(f, g) = \frac{1}{2} \sum_{x,y \in \mathcal{X}} (f(x) - f(y))(g(x) - g(y)) \mu(x)P(x, y).
\] (4)

It is well-known (see for example, Section 13.2 of \([6]\)) that

\[
\mathcal{E}_{P,\mu}(f, g) = \sum_{x \in \mathcal{X}} f(x) \cdot ((I - P)g)(x)\mu(x).
\] (5)

Since \(\mu\) is a reversible measure relative to the Markov transition matrix \(P\), the Dirichlet form \(\mathcal{E}_{P,\mu}\) is a positive semi-definite bilinear form on the vector space \(\mathcal{C}(\mathcal{X})\). Then the spectral gap of \(P\), relative to the reversible measure \(\mu\), is

\[
\Lambda^{(1)}(\mu) = \min \{\mathcal{E}_{P,\mu}(f, f) : f \in \mathcal{C}(\mathcal{X}), \text{Var}_\mu(f) \geq 1\} = \frac{1}{\sup \{\text{Var}_\mu(f) : \mathcal{E}_{P,\mu}(f, f) \leq 1\}},
\] (6)

which is well-defined as long as \(\mu\) is non-degenerate (\(\forall x \in \mathcal{X}, \text{ we have } \mu(x) > 0\) and \(|\mathcal{X}| \neq 1\). The spectral gap \(\Lambda^{(1)}(\mu)\) is strictly positive as long as \(P\) is irreducible (\(\forall x, y \in \mathcal{X}\), there is a \(T \in \{1, 2, \ldots\}\) and \(x_1, \ldots, x_{T-1} \in \mathcal{X}\) such that \(P(x_t, x_{t+1}) > 0\) for all \(t \in \{0, \ldots, T - 1\}\) where \(x_0 = x\) and \(x_T = y\).

See, for example, Levin and Peres for the notation (used in this article) related to finite state space Markov chains \([\mathbb{E}\,|\,X]\).

Then, also, let us denote the \(L^\infty-L^2\) Lipschitz constant of \(f \in \mathcal{C}(\mathcal{X})\) with respect to 1 step of the Markov chain

\[
\Phi_{P,\mu}^{\text{Lip}}(f) \overset{\text{def}}{=} \max_{x \in \mathcal{X}} \left(\sum_{y \in \mathcal{X}} P(x, y)|f(x) - f(y)|^2\right)^{1/2}
\] (7)

The Aida-Stroock theorem then gives exponential moment generating function bounds.

**Theorem 1.1** Define a function \(\Theta : [0, 1) \to [0, \infty)\) as

\[
\Theta(t) = \sum_{n=0}^{\infty} 2^n \ln \left(\frac{1}{1 - t \cdot 2^{-2n}}\right).
\] (8)

Then, for \(\lambda \geq 0\) satisfying \(\lambda < 2 \left(\Lambda^{(1)}(\mu)\right)^{1/2} / \Phi_{P,\mu}^{\text{Lip}}(f)\), it is true that

\[
\ln \left(\mathbb{E}_\mu[e^{\lambda f}]\right) \leq \lambda \mathbb{E}_\mu[f] + \Theta \left(\frac{\lambda \cdot \Phi_{P,\mu}^{\text{Lip}}(f)}{2 \left(\Lambda^{(1)}(\mu)\right)^{1/2}}\right).
\] (9)
This, in turn, gives bounds on $F_{\mu}^+ (f; \cdot)$ using Chebyshev’s inequality. Let us define

$$\Xi_P (\mu, f) \overset{\text{def}}{=} \frac{2 \left( \Lambda_P (1) (\mu) \right)^{1/2}}{\Phi_{\text{Lip}}^p (f)}.$$  \hfill (10)

Then we can deduce that the positive fluctuations actually an exponential decay bound almost at the rate of $\Xi_P (\mu, f)$:

**Corollary 1.2** For $a \geq 0$, the positive fluctuations obey the bound (which is written as a negative exponential value for the probability)

$$- \ln \left( F_{\mu}^+ (f; a) \right) \geq \max_{0 \leq \lambda \leq \Xi_P (\mu, f)} \left( \lambda (\lambda a - \Theta (\lambda \Xi_P (\mu, f))) \right).$$  \hfill (11)

By doing a small amount of calculus, this becomes clearer.

**Corollary 1.3** Defining a constant

$$\kappa = \sum_{n=1}^{\infty} 2^n \ln \left( \frac{1}{1 - 4^{-n}} \right) = \sum_{n=1}^{\infty} 2^n \sum_{m=1}^{\infty} \frac{4^{-mn}}{m} = \sum_{m=1}^{\infty} \frac{2}{m(4^m - 2)},$$  \hfill (12)

we have

$$\Theta (t) \leq \ln \left( \frac{1}{1 - t} \right) + \kappa. $$  \hfill (13)

And hence it follows (by using this bound and calculating the Legendre transform of the bounding function) that

$$\ln \left( F_{\mu}^+ (f; a) \right) \leq \kappa - \vartheta (a \cdot \Xi_P (\mu, f)),$$  \hfill (14)

where $\vartheta : [0, \infty) \to [0, \infty)$ is an asymptotically linear function

$$\vartheta (x) = x \cdot \varphi (x) + \ln \left( \frac{2}{x} \cdot \varphi (x) \right), \quad \text{for} \quad \varphi (x) = \sqrt{1 + \frac{1}{x^2} - \frac{1}{x}}.$$  \hfill (15)

The absolute constant $\kappa$ satisfies

$$1.084640 \leq \kappa \leq 1.084645.$$  \hfill (16)

An excellent reference for Theorem 1.1 and Corollary 1.2 is Ledoux’s monograph on the concentration of measure phenomenon [5]. As stated before, Aida and Stroock proved their result on the way to proving more sophisticated results. They did not include details of the proofs beyond the basic outline. But in Section 3.1 in Ledoux’s monograph he states the equivalent result as Theorem 3.3, and he gives complete details of the proof. In particular, when the details are laid out, it becomes apparent that one can obtain a slight improvement if one restricts attention to positive fluctuations. (So far, we have restricted attention to positive fluctuations.)

**Theorem 1.4** In place of (7) consider the asymmetric Lipschitz constant

$$\tilde{\Lambda}_{\text{asym}} (P; f) \overset{\text{def}}{=} \max_{x \in \mathcal{X}} \left( \sum_{y \in \mathcal{X}} P (x, y) |f(x) - f(y)|^2 \cdot 1_{(0, \infty)} (f(x) - f(y)) \right)^{1/2}. $$  \hfill (17)

Also, define the replacement of (10) as

$$\Xi_P^+ (\mu, f) \overset{\text{def}}{=} \frac{2 \left( \Lambda_P (1) (\mu) \right)^{1/2}}{\Phi_{\text{Lip}}^p (P; f)}. $$  \hfill (18)
Then, we also have, for $\lambda \geq 0$, satisfying $\lambda < \Xi_\mu^*(\mu, f)$, it is true that

$$\ln \left( \mathbb{E}_\mu [e^{\lambda f}] \right) \leq \lambda \mathbb{E}_\mu [f] + \Theta \left( \frac{\lambda}{\Xi_\mu^*(\mu, f)} \right). \quad (19)$$

And, hence, it is also true that for $a \geq 0$, the positive fluctuations obey the bound

$$- \ln \left( \mathcal{F}_\mu^+ (f; a) \right) \geq \max_{\lambda \leq \lambda \leq \Xi_\mu^*(\mu, f)} \left( \lambda a - \Theta \left( \frac{\lambda}{\Xi_\mu^*(\mu, f)} \right) \right). \quad (20)$$

We do not bother to re-state the calculus facts, but suffice it to say that in Corollary 1.2 and Corollary 1.3 the constant $\Xi_\mu^*(\mu, f)$ may be replaced by $\Xi_\mu^*(\mu, f)$.

The slight generalization of Theorem 1.4 will be used in our first example. We will also include a brief discussion to note that the asymmetric focus on fluctuations is natural and also is already well-established in some famous examples. Let us mention that the asymmetric Lipschitz constant defined in 1.4 satisfies good properties.

**Proposition 1.5** We have the following properties.

1. For all $f, g \in C(\mathcal{X})$, we have $\tilde{\Phi}_\text{asym}^\text{Lip} (P; f + g) \leq \tilde{\Phi}_\text{asym}^\text{Lip} (P; f) + \tilde{\Phi}_\text{asym}^\text{Lip} (P; g)$.

2. For all $f \in C(\mathcal{X})$ and all $c \in [0, \infty)$ we have $\tilde{\Phi}_\text{asym}^\text{Lip} (P; c \cdot f) = c \cdot \tilde{\Phi}_\text{asym}^\text{Lip} (P; f)$.

3. Moreover, if we assume that $P$ is irreducible, then we have: $\tilde{\Phi}_\text{asym}^\text{Lip} (P; f) = 0$ implies that $f$ is constant.

The third condition above is related to being Lipschitz. Another condition justifying that name is the following. Suppose that $X_0, X_1, X_2, \ldots$ are the states of a random realization of the Markov chain starting from $\mu$, so that

$$\forall t \in \mathbb{N}, \forall x_0, x_1, \ldots, x_t \in \mathcal{X}, \text{ we have } \mathbb{P}(\{X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t\}) = \mu(x_0)P(x_0, x_1) \cdots P(x_{t-1}, x_t). \quad (21)$$

Then

$$\left( \mathbb{E} \left[ (f(X_0) - f(X_t))^2 \cdot 1_{(0, \infty)}(f(X_0) - f(X_t)) \right] \right)^{1/2} \leq t \cdot \tilde{\Phi}_\text{asym}^\text{Lip} (P; f). \quad (22)$$

**2 First example of the general method: longest increasing sub-sequence of a uniform random permutation**

Given $m, n \in \mathbb{N} = \{1, 2, \ldots\}$, let $A_m = \{0, \frac{1}{m}, \frac{2}{m}, \ldots, 1\}$ and let $A_{m,n} = A_m^n$, which is the set of all $(x_1, \ldots, x_n)$ with $x_1, \ldots, x_n \in A_m$. The Markov chain transition matrix $P_{m,n} : A_{m,n} \rightarrow A_{m,n}$ is just replacement of one of the coordinates uniformly at random, where the replacement is by an element of $A_m$ chosen uniformly at random. Therefore, we have

$$P_{m,n}(\{x_1, \ldots, x_n\}, \{y_1, \ldots, y_n\}) = \frac{1}{mn} \sum_{i=1}^n \left( \prod_{j \in \{1, \ldots, n\} \setminus \{i\}} \delta(x_i, y_i) \right), \quad (23)$$

where $\delta(x, y) = 1$ if $x = y$, and equals 0 otherwise. Since the Kronecker $\delta$ function is symmetric, this is a reversible measure for the invariant measure $\mu_{m,n}$ which is the uniform measure

$$\mu_{m,n}(\{x_1, \ldots, x_n\}) = \frac{1}{m^n}, \text{ for every } (x_1, \ldots, x_n) \in A_m^n. \quad (24)$$
Now suppose that we have a set $J \subset \{1, \ldots, n\}$ and we choose for each $i \in J$ a function $f_i : A_m \to \mathbb{R}$ such that $\sum_{x \in A_m} f_i(x) = 0$, but $f_i$ is not identically zero. In other words, we just assume $f_i$ is orthogonal to the constant function. Then, letting $F : X_{m,n} \to \mathbb{R}$ be the function

$$F(x_1, \ldots, x_n) = \prod_{i \in J} f_i(x_i),$$

it is easy to see that $P_{m,n} F = \frac{n-|J|}{n} F$. (If one of the coordinate indices in $J$ is selected for replacement, then the function $f_i(y_i)$ after replacement has average equal to zero since $f_i$ is orthogonal to the constant function. If any other index is chosen, then the function $F$ is unchanged.) This suffices to determine a spanning set of eigenvectors. So the set of eigenvalues is $\{1, 1 - \frac{1}{n}, 1 - \frac{2}{n}, \ldots, 0\}$. In particular, using (5) and (6), we have the following.

**Lemma 2.1** For the replacement Markov chain we have been considering, the spectral gap is

$$\Lambda_{m,n}^{(1)} (P_{m,n}) = \frac{1}{n}.$$  

Now, for the observable, we take $f_{m,n} : X_{m,n} \to \mathbb{R}$ to be the length of the longest increasing subsequence. More precisely, let us define $f_n : [0,1]^n \to \mathbb{R}$ to be

$$f_n((x_1, \ldots, x_n)) = \max (\{ |J| : J \subset \{1, \ldots, n\}, \text{ and } \forall i,j \in J, \text{ we have } (i < j) \Rightarrow (x_i < x_j) \}).$$

Then we take $f_{m,n}$ to just be the restriction $f_n \upharpoonright A_m^n$.

Now let us try to calculate the Lipschitz constant. Suppose that for some $(x_1, \ldots, x_n) \in \mathbb{R}^n$, the set $J \subset \{1, \ldots, n\}$ is a set where $f_{m,n}((x_1, \ldots, x_n)) = |J|$, and such that

$$\forall i, j \in J, \text{ we have } (i < j) \Rightarrow (x_i < x_j).$$

Then, in one step of the Markov chain, updated by $P_{m,n}$, the only way to decrease the value of $f_{m,n}$ is to choose one of the indices in $J$ to replace by a uniform random sample. That has probability equal to $|J|/n$. In that case, it is still possible that the length of the longest increasing subsequence does not decrease. But if it does decrease, it only decreases by 1. Therefore, we have, defining

$$\Delta_{m,n}((x_1, \ldots, x_n)) \overset{\text{def}}{=} \sum_{(y_1, \ldots, y_n) \in X_{m,n}} P_{m,n}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) \cdot \left( f_{m,n}((x_1, \ldots, x_n)) - f_{m,n}((y_1, \ldots, y_n)) \right)^2,$$

it is the case that

$$\Delta_{m,n}((x_1, \ldots, x_n)) \leq \frac{|J|}{n}.\quad (27)$$

Now for the Lipschitz constant (asymmetric, semi-norm) we have to maximize over all choices of $(x_1, \ldots, x_n)$.

That would actually give us a much larger constant than if we restricted to the typical choice of $(x_1, \ldots, x_n)$. That is because the typical value of $|J|$ is approximately $2\sqrt{n}$, as determined by Vershik and Kerov [11] and Logan and Shepp [7]. The correct order is $\sqrt{n}$ from an even easier argument of Hammersley [2]. But if we had large deviation bounds, then we could use those to initialize a more refined bound. In view of all of this, we will just truncate, by-hand. Given any constant $K \in \mathbb{N}$, let us define

$$f_n^{(K)}((x_1, \ldots, x_n)) = \min (\{ K, f_n((x_1, \ldots, x_n)) \}).$$

Then we define $f_{m,n}^{(K)} = f_n^{(K)} \upharpoonright A_m^n$. Then the above calculations show that if $(x_1, \ldots, x_n)$ is such that $|J| \leq K$ then we have

$$\Delta_{m,n}((x_1, \ldots, x_n)) \leq \frac{K}{n}.\quad (31)$$

5
But if $|J| \geq K + 1$ then $f_n((x_1, \ldots, x_n)) = |J| > K$ and no matter what, we will also have $f_n^{(K)}((y_1, \ldots, y_n)) = K = f_n^{(K)}((y_1, \ldots, y_n))$. Therefore, from (17) we have

$$\Phi_{\text{asym}}(P; f_n^{(K)}) \leq \left( \frac{K}{n} \right)^{1/2}. \quad (32)$$

This is the bound which we wanted.

We will take $K$ to be a number depending on $n$, so that we really have a sequence $K_n$. And we will choose the sequence such that

$$\lim_{n \to \infty} n^{-1/2} K_n = u \in (2, \infty). \quad (33)$$

As stated before, to obtain un-restricted bounds, we would need to combine this with large deviation bounds. Implicitly, we are assuming that in more general applications, it would be easier to get large deviation bounds than concentration-of-measure bounds. So, from (18) we have

$$\Xi^+(\mu_{m,n}, f_{m,n}) \leq 2 \sqrt{n^{1/2} K_n}. \quad (34)$$

Then, using this with Corollary 1.3 using $\Xi^+(\mu, f)$ in place of $\Xi(\mu, f)$ as discussed at the end of the last section, we obtain the following.

**Corollary 2.2** For the truncation of the length of the longest increasing subsequence, truncated at the level $K = K_n$ such that $\lim_{n \to \infty} K_n/(uN^{1/2}) = 1$ for some $u \in (2, \infty)$, we have the bound

$$\ln \left( F_{\mu_{m,n}}^{+} \left( f_{m,n}^{(K_n)}, tn^{1/4} \right) \right) \leq \kappa - \vartheta \left( 2t \sqrt{\frac{n^{1/2}}{K_n}} \right). \quad (35)$$

So in particular, for a fixed $t \in (0, \infty)$ the right hand side converges to $\kappa - \vartheta(2t/u^{1/2})$.

Note that in the above, the best case for the right hand side would be if one could take $u$ close to 2 to get $\kappa - \vartheta(\sqrt{2}t)$. But one cannot do better than that with these methods. (On can only do that well if one uses good large deviation bounds that are sufficiently good even going down to the median.) These bounds essentially show that with this method one can determine that the fluctuations are no larger than order $n^{1/4}$. As shown by Baik, Deift and Johansson the true fluctuations are of order $n^{1/6}$. But that is bounded by order $n^{1/4}$, so that these bounds are not untrue. They just are not very sharp. But that is the situation also for Talagrand’s bounds from [10].

We note that the idea of developing asymmetric bounds for the positive and negative fluctuations is not an original idea. It is already advocated by Talagrand. The reason that he obtains bounds as good as he does for the length of the longest increasing subsequence is that the function $f_n((x_1, \ldots, x_n))$ only depends on $(x_1, \ldots, x_n)$ through the points whose indices are in $J \subset \{1, \ldots, n\}$. He called functions such as this “configuration functions.” Another good reference is Steele’s monograph [9].

We also note that bounds for the negative fluctuations are easily obtained from the bounds for the positive fluctuations, using Markov’s inequality. Of course, one will not obtain as sharp a result that way. Using

$$\int_0^{\infty} F_{\mu}^{-}(f; t) \, dt = \int_0^{\infty} F_{\mu}^{+}(f; t) \, dt,$$  \hspace{1cm} (36)

we may determine

$$F_{\mu}^{-}(f; t) \leq \frac{1}{t} \int_0^{\infty} F_{\mu}^{+}(f; t) \, dt. \quad (37)$$
If we have bounds showing that $F_{\mu_{m,n}}^+(f_{m,n}(K_n);t)\,dt$ decays exponentially, because we chose $K_n$ to be approximately $un^{1/2}$ for some $u \in (2, \infty)$, then we can see that
\[
\int_0^\infty F_{\mu_{m,n}}^+(f_{m,n}(K_n);t)\,dt \leq Cn^{1/4},
\]
for a constant $C$ that depends on $K_n$, or alternatively depends on $u$. That way, we would see that the negative fluctuations $F_{\mu_{m,n}}^-(f_{m,n}(K_n);t)$ are also decaying when $t$ is at the order of $n^{1/4}$. So, even though the positive and negative fluctuations have different types of bounds, the order of the size of the fluctuations that one obtains bounds for using this technique is the same for both positive and negative fluctuations. It is order $n^{1/4}$ in this case.

**Remark 2.3** If we take $n$ fixed and let $m$ go to $\infty$, then we obtain the analogous bounds when the points $x_1, \ldots, x_n$ are chosen uniformly on the continuous interval $[0,1]$, in an IID fashion. Since the function only depends on the permutation or relative order induced by the points, that is not a singular limit. Rather, for finite $m$, the probability that none of the components are equal is 1 minus a quantity which is on the order of $n^2/m$ by the Birthday problem. When none of the components are equal, conditioning on that event, we do have uniform random permutations, just as if $x_1, \ldots, x_n$ were distributed uniformly on the continuous interval $[0,1]$ in an IID fashion.

### 3 Second example of the general method: Lipschitz functions of 1d Markov chains

Suppose that $\mathcal{Y}$ is a finite state space, and consider a larger state space $\mathcal{X}_n = \mathcal{Y}^{n+1}$ for some $n \in \mathbb{N} = \{1, \ldots, n\}$. Suppose that $Q: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ is a Markov transition matrix which is irreducible and aperiodic, and suppose that there is a measure $\nu: \mathcal{Y} \rightarrow \mathbb{R}$ (satisfying $\nu(x) \geq 0$ for all $x \in \mathcal{Y}$ and $\sum_{x \in \mathcal{Y}} \nu(x) = 1$). And, suppose that $\nu$ is reversible with respect to $Q$:
\[
\forall x, y \in \mathcal{Y}, \quad \text{we have } \nu(x)Q(x,y) = \nu(y)Q(y,x).
\]

By irreducibility and aperiodicity, we know that $\Lambda_\nu^{(1)}(Q) > 0$. The Markov chain we will consider is on $\mathcal{X}_n$ instead of $\mathcal{Y}$. But this fact is potentially useful for proving lower bounds on the spectral gap of the chain on $\mathcal{X}_n$.

Before stating the Markov transition matrix for the chain on $\mathcal{X}_n$, let us define the measure we wish to be the invariant measure for the Markov chain. Let $\mu_n: \mathcal{X}_n \rightarrow \mathbb{R}$ be the measure defined as
\[
\mu_n((x_0, x_1, \ldots, x_n)) = \mathbb{P}_\nu\{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\} = \nu(x_0)P(x_0, x_1)\cdots P(x_{n-1}, x_n),
\]
for each $(x_0, x_1, \ldots, x_n) \in \mathcal{X}_n = \mathcal{Y}^{n+1}$. Here $X_0, X_1, \ldots, X_n$ is viewed as a sample of the 1d Markov chain, from times $t = 0$ to $t = n$, started at time 0 in the distribution $\nu$. By reversibility of $\nu$ with respect to $Q$, we also have
\[
\mu_n((x_0, x_1, \ldots, x_n)) = \nu(x_n)P(x_n, x_{n-1})\cdots P(x_1, x_0),
\]
and for $t \in \{1, \ldots, n-1\}$
\[
\mu_n((x_0, x_1, \ldots, x_n)) = \nu(x_t)\left(P(x_t, x_{t+1})\cdots P(x_{n-1}, x_n)\right)\left(P(x_t, x_{t-1})\cdots P(x_1, x_0)\right).
\]

These alternative formulations are potentially useful for proving reversibility for the Markov chain on $\mathcal{X}_n$, which is what we consider next.
We consider Glauber dynamics as the Markov chain on $\mathcal{X}_n = \mathcal{Y}^{n+1}$. In other words, we consider $P_n : \mathcal{X}_n \times \mathcal{X} \to \mathbb{R}$ to be the Markov transition matrix where

$$P_n((x_0, \ldots, x_n), (y_0, \ldots, y_n)) = \frac{1}{n+1} \sum_{t=0}^{n} P_{n,t}((x_0, \ldots, x_n), (y_0, \ldots, y_n)), \quad (43)$$

where for $t \in \{1, \ldots, n-1\}$ we have

$$P_{n,t}((x_0, \ldots, x_n), (y_0, \ldots, y_n)) = \frac{Q(x_{t-1}, y_t)Q(y_t, x_{t+1})}{\sum_{z \in \mathcal{Y}} Q(x_{t-1}, z)Q(z, x_{t+1})} \cdot \prod_{s \in \{0, \ldots, n\} \setminus \{t\}} \delta(x_s, y_s), \quad (44)$$

while

$$P_{n,0}((x_0, \ldots, x_n), (y_0, \ldots, y_n)) = \frac{\nu(y_0)Q(y_0, x_1)}{\sum_{z \in \mathcal{Y}} \mu(z)Q(z, x_1)} \cdot \prod_{s \in \{1, \ldots, n\}} \delta(x_s, y_s). \quad (45)$$

Note that by reversibility of $\nu$ with respect to $Q$, these can be written seemingly different but equivalent ways. For example,

$$P_{n,0}((x_0, \ldots, x_n), (y_0, \ldots, y_n)) \cdot \mu((x_0, x_1, \ldots, x_n)) = \frac{Q(x_1, y_0)}{\sum_{z \in \mathcal{Y}} Q(x_1, z)} \cdot \prod_{s \in \{1, \ldots, n\}} \delta(x_s, y_s), \quad (46)$$

as well.

It is easy to see that each of the $P_{n,t}$ matrices, for $t \in \{0, 1, \ldots, n-1\}$ is such that $\mu_n$ is reversible for $P_{n,t}$. For example, if $t$ is in $\{1, \ldots, n-1\}$, then

$$P_{n,t}((x_0, \ldots, x_n), (y_0, \ldots, y_n)) = \frac{Q(x_{t-1}, y_t)Q(y_t, x_{t+1})}{\sum_{z \in \mathcal{Y}} Q(x_{t-1}, z)Q(z, x_{t+1})} \cdot \nu(x_t)Q(x_0, x_1) \cdots Q(x_{n-1}, x_n), \quad (47)$$

if $y_s = x_s$ for all $s \in \{1, \ldots, n\} \setminus \{t\}$ (and it equals 0 otherwise). Isolating $x_t$ and $y_t$, and assuming $y_s = x_s$ for all $s \in \{1, \ldots, n\} \setminus \{t\}$ in order not to get 0, this is

$$P_{n,t}((x_0, \ldots, x_n), (y_0, \ldots, y_n)) \mu_n((x_0, x_1, \ldots, x_n)) = Q(x_{t-1}, y_t)Q(y_t, x_{t+1})Q(x_{t-1}, x_t)Q(x_t, x_{t+1}) \cdot F_{n,t}(x_0, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n), \quad (48)$$

for

$$F_{n,t}(x_0, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n) = \frac{\nu(x_0)Q(x_0, x_1) \cdots Q(x_{t-2}, x_{t-1}) \cdot Q(x_{t+1}, x_{t+2}) \cdots Q(x_{n-1}, x_n)}{\sum_{z \in \mathcal{Y}} Q(x_{t-1}, z)Q(z, x_{t+1})}. \quad (49)$$

Clearly (48) is symmetric in interchange of the two coordinates of $(x_t, y_t)$. Also, the conditions imposed by multiplying by $\prod_{s \in \{0, \ldots, n\} \setminus \{t\}} \delta(x_s, y_s)$ is also symmetric in interchange of every $(x_s, y_s)$ for $s \in \{0, \ldots, n\} \setminus \{t\}$. So $\mu$ is a reversible measure for Glauber dynamics. We refer to Chapter 3 of Levin and Peres, for example [6], for more details on Glauber dynamics.

**Proposition 3.1** For the Glauber dynamics we have been considering, there is a constant $\lambda_1$ satisfying

$$\lambda_1 > 0, \quad (50)$$

such that

$$\Lambda_{\mu_n}^{(1)}(P_n) \geq \frac{\lambda_1}{n+1}. \quad (51)$$

We will not prove this, here. But it is reportedly well-known. A reference is Lu and Yau’s paper on Glauber dynamics and Kawasaki dynamics [3].

We have a specific application in mind, which we call target frequency analysis, which is also hypothesis testing for the power spectrum (Fourier transform amplitudes-squared) integrated over certain intervals. But before moving to that example, let us just quickly state the general result.
Corollary 3.2 Suppose that we have a function \( f_n : \mathcal{X}_n \to \mathbb{R} \) satisfying
\[
\bar{\Phi}_{\text{Lip}}^{\text{asym}}(P_n; f_n) \leq \frac{C}{(n+1)^p},
\]for some power \( p \). Then we have the bound
\[
\ln \left( \frac{\mathcal{F}^+_\mu_n \left( f_n; a(n+1)^{-p+\frac{1}{2}} \right) }{\mu_n} \right) \leq \varphi - \vartheta \left( a \cdot \frac{2\sqrt{\lambda_1}}{C} \right),
\]using the notation from Corollary 1.3.

For us, the power we will obtain will be \( p = 1 \), so that the fluctuations will be shown to be bounded by \( O(n^{-1/2}) \) in this way, for a nonnegative observable whose mean is order-1.

3.1 Example: Target frequency analysis for Markov chains

As another basic application, we consider a statistic for time series which was considered by the authors and Jung in [3]. This is called “target frequency analysis.” The application is important in biostatistics. But it also supplies a pedagogically valuable example for the technique.

Let \( \mathcal{Y} \) be \( \mathcal{Y}_{m,\epsilon} = \{ -m\epsilon, (-m+1)\epsilon, \ldots, m\epsilon \} \). For us, an important quantity is the radius of this chain \( R = m\epsilon \). Note that \( \mathcal{Y}_{m,\epsilon} \subset \mathbb{R} \), and the next step in the description of the function \( f_n \) on \( \mathcal{X}_n = (\mathcal{Y}_{m,\epsilon})^{n+1} \) only relies on that. Given a real sequence \( (x_0, \ldots, x_n) \) we define the Fourier transform \( \phi_{(x_0, \ldots, x_n)} : \mathbb{Z} \to \mathbb{C} \) defined as if \( (x_0, \ldots, x_n) \) were the \( n+1 \) components of a periodic signal
\[
\phi_{(x_0, \ldots, x_n)}(k) = \frac{1}{\sqrt{n+1}} \sum_{t=0}^{n} \exp \left( -\frac{2\pi i k t}{n+1} \right) x_t,
\]
where \( i = \sqrt{-1} \), as usual (despite the fact that in earlier sections the symbol \( i \) was used for an integer index). The choice of the prefactor \( 1/\sqrt{n+1} \) is such that Parseval’s identity is satisfied
\[
\sum_{k=0}^{n} |\phi_{(x_0, \ldots, x_n)}(k)|^2 = \frac{1}{n+1} \sum_{k=0}^{n} \sum_{s=0}^{n} \sum_{t=0}^{n} \exp \left( \frac{-2\pi i k (t-s)}{n+1} \right) x_t x_s
\]
\[
= \frac{1}{n+1} \sum_{s=0}^{n} \sum_{t=0}^{n} \exp \left( \frac{-2\pi i k (t-s)}{n+1} \right) x_t x_s
\]
\[
= \frac{1}{n+1} \sum_{s=0}^{n} \sum_{t=0}^{n} \sum_{k=0}^{n} \exp \left( \frac{-2\pi i k (t-s)}{n+1} \right) x_t x_s
\]
\[
= \sum_{s=0}^{n} x_s^2
\]
Now, given any choice of \( a, b \in \mathbb{Z} \) satisfying \( 0 < a < b < (n+1)/2 \), we consider the observable of interest to be
\[
f_n(x_0, \ldots, x_n) = \frac{1}{n+1} \sum_{k=a}^{b} |\phi_{(x_0, \ldots, x_n)}(k)|^2.
\]
In other words, using the language of signal processing, it is the power spectrum integrated from \( a \) to \( b \). Now it is rescaled by \( 1/(n+1) \) because for a signal of length \( n+1 \), we expect the total \( \ell^2 \)-norm (also called the total power) to be of order \( (n+1) \). So this rescales to give an order-1 quantity. Note that the Fourier transform is an isometry by (54), therefore \( f_n \) may be viewed as a contraction mapping times a constant \( 1/(n+1) \). For this reason, we obtain
\[
\Phi_{P_n}^{\text{Lip}}(f_n) \leq \frac{2R^2}{n+1}.
\]
Since \( \Phi_{\text{asym}}^{\text{Lip}}(P_n; f_n) \leq \Phi_{P_n}^{\text{Lip}}(f_n) \), this proves the following using Corollary 3.2.
Corollary 3.3 For the function $f_n$ written above, we have
\[
\ln \left( \mathcal{F}_{\mu_n}^+ \left( f_n; \frac{a}{\sqrt{n+1}} \right) \right) \leq \kappa - \theta \left( a \cdot \frac{\sqrt{\lambda_1}}{R^2} \right), \quad (57)
\]

We call the function $f_n$ by the name “target frequency analysis.” It has a special property: if we replace the present set-up by the case where $Y = \mathbb{R}$ and allow $X_0, \ldots, X_n$ to be IID standard, normal random variables (also called white noise, by some), then the Fourier transform has the property for frequencies $k$ satisfying $0 < k < (n + 1)/2$ the real and imaginary parts are all IID standard, normal random variables. (Here IID refers to the independence of the real and imaginary parts, as well as independence for different values of $k$.) By the Parseval identity, isometry property, it is elementary that the Fourier transform of IID complex-valued signals with real and imaginary parts being IID standard, normal random variables would have the same property. But the property stated for the Fourier transform of a real signal is slightly less trivial, although it may be easily checked using covariance matrices. One may also thinking of this fact as arising from the slight extra information included in the dihedral symmetry over the usual cyclic symmetry, for the dihedral group $D_n$ being the semi-direct product of the cyclic group $(\mathbb{Z}/n\mathbb{Z})$ with the involution group $(\mathbb{Z}/2\mathbb{Z})$. One can also see it by using properties of the complex conjugation, which amounts to the same. But it is probably the simplest example of a more general phenomenon where for special symmetrical models, random variables defined on a large space have unexpected projections into some components which also have simple, explicit distributions on smaller spaces.

The corollary above shows that more generally, for Markov chain models of a time series, the target frequency analysis will still be concentrating at least in the sense that the fluctuations are no larger than order $n^{1/2}$.

References

[1] Aida and Stroock. Moment estimates derived from Poincaré and logarithmic Sobolev inequalities. Math. Res. Lett. 1, 75–86 (1995).

[2] Jinho Baik, Percy Deift and Kurt Johansson. Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem. J. Amer. Math. Soc. 12, 1119–1178 (1999).

[3] A Target Frequency Analysis of functional MRI Data. Michael Fröhlich, Paul Jung and Shannon Starr. Int. J. Clin. Biostat-Biom (2015) 2015, no 1:2 (6 pages).

[4] John Michael Hammersley. A few seedlings of research. In Proc. Sixth Berkeley Symp. Math. Statist. Probab. v. 1, pp. 345–394. Univ. California Press, Berkeley, 1972.

[5] Michel Ledoux. The Concentration of Measure Phenomenon. AMS, Providence, RI, 2001.

[6] David A. Levin and Yuval Peres. Markov Chains and Mixing Times. Second edition. American Mathematical Society, Providence, RI, 2017.

[7] Benjamin F. Logan and Lawrence A. Shepp. A variational problem for random Young tableaux. Adv. Math. 26, 206–222 (1977).

[8] Sheng Lin Lu and Horng-Tzer Yau. Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. Comm. Math. Phys. 156, no. 2, 399–433 (1993).

[9] J. Michael Steele. Probability Theory and Combinatorial Optimization. CBMS-NSF Regional Conference Series in Applied Mathematics (Book 69). SIAM, Philadelphia, PA, 1987.

[10] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Publications Mathématiques de l’IHÉS 81, 73–205 (1995).
[11] Anatoly M. Vershik and Sergei V. Kerov. Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux. *Soviet Math. Dokl.* **18**, 527–531 (1977). (Translation of *Dokl. Acad. Nauk. SSR* **32**, 1024–1027.)

**Contact:** slstarr@uab.edu