These notes have not received the scrutiny of publication. They could be missing important references, etc.

Glauber dynamics

1 A Poincaré inequality

We first review some relevant definitions of Markov chains.

Definition 1. Let P be the transition matrix of a (finite) Markov chain with state space Ω. Then $L = I - P$ is called the *Laplacian* of the Markov chain. For functions f and g from Ω to $\mathbb{R}$, we define the *Dirichlet form* by

$$\mathcal{E}_P(f, g) = \langle f, Lg \rangle_\pi = \sum_{x \in \Omega} f(x)(Lg)(x)\pi(x)\,.$$

For reversible Markov chains, we have a more insightful formula for the Dirichlet form. Rewriting the definition, we have

$$\begin{aligned}
\mathcal{E}_P(f, g) &= \langle f, (I - P)g \rangle_\pi \\
&= \sum_{x \in \Omega} \pi(x) f(x)((I - P)g)(x) \\
&= \sum_{x \in \Omega} \pi(x) f(x) \left(g(x) - \sum_{y \in \Omega} P(x, y) g(y) \right) \\
&= \sum_{x, y \in \Omega} \pi(x) f(x) P(x, y)(g(x) - g(y))\,.
\end{aligned}$$

Reversibility gives $\pi(x)P(x, y) = \pi(y)P(y, x)$, so this equals

$$\sum_{x, y \in \Omega} \pi(x) P(x, y) f(y)(g(y) - g(x))\,.$$

Averaging these two expressions gives a new formula for the Dirichlet form

$$\mathcal{E}_P(f, g) = \frac{1}{2} \sum_{x, y \in \Omega} \pi(x) P(x, y)(f(x) - f(y))(g(x) - g(y))\,.$$

In particular, we get

$$\mathcal{E}_P(f, f) = \frac{1}{2} \sum_{x, y \in \Omega} \pi(x) P(x, y)(f(x) - f(y))^2\,. \tag{1}$$

Note that by reversibility the quantity $\pi(x)P(x, y)$ is a property of the edge $e = \{x, y\}$ and does not depend on the order of the endpoints. This is sometimes called the *conductance* of e, a terminology derived from electrical network. The quantities $f(x) - f(y)$ and $g(x) - g(y)$ can be understood as the gradients of f and g along the edge e. Thus, the expression (1) represents the inner product (induced by edge conductance) of the gradients of two functions, similar to the Dirichlet form in functional analysis.

Recall that if P is ergodic and reversible with stationary distribution π, any function $f : \Omega \to \mathbb{R}$ satisfies a *Poincaré inequality*

$$\mathrm{Var}_{X \sim \pi}(f(X)) \leq \frac{1}{1 - \lambda_2(P)} \mathcal{E}_P(f, f)\,. \tag{2}$$

The inequality (2) is also called the *spectral gap inequality* since 1 is the unique largest eigenvalue of P if it is ergodic and reversible.

Intuitively, the Dirichlet form (1) quantifies the "local variance" of f along the transitions. In this regard, the spectral gap inequality (2) implies that the greater the spectral gap, the greater the local variance compared to the global variance, which means that the Markov chain mixes more rapidly. This matches our intuition from spectral graph theory where graphs with large spectral gap possess "good expansion properties."

Example 1. We consider a random walk on the complete graph on n vertices. Its adjacency matrix A is given by $\mathbf{1}\mathbf{1}^\top - I$, so we have the transition matrix

$$P = \frac{1}{n-1}A = \begin{bmatrix} 0 & \frac{1}{n-1} & \cdots & \frac{1}{n-1} \\ \frac{1}{n-1} & 0 & \cdots & \frac{1}{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n-1} & \frac{1}{n-1} & \cdots & 0 \end{bmatrix}.$$

It is not difficult to see that the spectrum of P is given by

$$\mathrm{spec}(P) = \left(1, -\frac{1}{n-1}, \cdots, -\frac{1}{n-1}\right).$$

In particular, we have $\lambda_2(P) = -\frac{1}{n-1} < 0$, so complete graphs have excellent spectral gap.

In general, for a random walk on d-regular graphs, the transition matrix is given by $P = \frac{1}{d}A$. Some d-regular graphs (e.g. Ramanujan graphs) satisfy $\lambda_2(P) = \Theta(d^{-1/2})$, which make them good expanders and also rapidly mixing Markov chains. It is known (e.g. the Alon–Boppana bound) that such graphs exhibit asymptotically largest possible spectral gap.

2 Glauber dynamics

Glauber dynamics or *Gibbs sampler* is an algorithm for sampling from a probability distribution π on an n-dimensional state space $\Omega = \Sigma_1 \times \cdots \times \Sigma_n$ (for the Ising model, we set $\Sigma_1 = \cdots \Sigma_n = \{\pm 1\}$). It is useful when the conditional distribution of one component given all the other components is easy to compute. Glauber dynamics defines a probability transition kernel

$$P = \frac{1}{n}\sum_{i=1}^{n} P_i$$

where each P_i is defined by

$$(P_i f)(y) = \mathbb{E}_{X\sim\pi}[f(X) \mid X_{\sim i} = y_{\sim i}]$$

or equivalently,

$$e_y^\top P_i = \pi(\cdot \mid X_{\sim i} = y_{\sim i}).$$

Here, $x_{\sim i}$ means all the components of x except for i. One step of the transition of Glauber dynamics can be described as the following.

1. Select i from $\{1, \cdots, n\}$ uniformly at random.

2. Resample the ith component X_i from the conditional distribution $X_i \mid X_{\sim i}$.

Our goal now is to analyze the behavior of Glauber dynamics. For notational convenience, we write $\Omega_{\sim i} = \Sigma_1 \times \cdots \times \Sigma_{i-1} \times \Sigma_{i+1} \times \cdots \times \Sigma_n$ and let $\pi_{\sim i}$ denote the marginal distribution on $\Omega_{\sim i}$.

Fact 1. *Suppose $\pi(x) > 0$ for all $x \in \Omega$. Then P is ergodic.*

Proof. We show that P^n has positive entries. Indeed, for any $(x_1, \cdots, x_n), (y_1, \cdots, y_n) \in \Omega$, each step of the transition

$$(x_1, \cdots, x_n) \to (y_1, x_2, \cdots, x_n) \to (y_1, y_2, x_3, \cdots, x_n) \to \cdots \to (y_1, \cdots, y_n)$$

has positive probability. □

Fact 2. *P is reversible with stationary distribution π.*

Proof. We verify that P and π satisfy the detailed balance equation:

$$\begin{aligned}
\pi(x)P(x,y) &= \sum_{i=1}^{n} \mathbf{1}\{x_{\sim i} = y_{\sim i}\}\pi(x) \Pr_{X\sim\pi}[X = y \mid X_{\sim i} = x_{\sim i}] \\
&= \sum_{i=1}^{n} \mathbf{1}\{x_{\sim i} = y_{\sim i}\} \frac{\pi(x)\pi(y)}{\pi_{\sim i}(x_{\sim i})} \\
&= \sum_{i=1}^{n} \mathbf{1}\{x_{\sim i} = y_{\sim i}\} \frac{\pi(x)\pi(y)}{\pi_{\sim i}(y_{\sim i})} \\
&= \pi(y)P(y,x) .
\end{aligned}$$

□

Given $\pi(x) > 0$ for all $x \in \Omega$, we have proved that P is ergodic and reversible, thus satisfying the spectral gap inequality (2). Now we compute the Dirichlet form $\mathcal{E}_P(f,f)$ using the formula (1):

$$\begin{aligned}
\mathcal{E}_P(f,f) &= \frac{1}{2}\sum_{x,y\in\Omega} \pi(x)P(x,y)(f(x)-f(y))^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{x,y\in\Omega} \pi(x)P_i(x,y)(f(x)-f(y))^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{x_{\sim i}=y_{\sim i}} \frac{\pi(x)\pi(y)}{\pi_{\sim i}(x_{\sim i})}(f(x)-f(y))^2 \\
&= \frac{1}{2n}\sum_{i=1}^{n}\sum_{z\in\Omega_{\sim i}}\sum_{x_{\sim i}=y_{\sim i}=z} \pi(x \mid x_{\sim i} = z)\pi(y \mid y_{\sim i} = z)\pi_{\sim i}(z)(f(x)-f(y))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{z\in\Omega_{\sim i}} \pi_{\sim i}(z)\mathrm{Var}_{X\sim\pi}(f(X) \mid X_{\sim i} = z) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{X\sim\pi}\mathrm{Var}(f(X) \mid X_{\sim i}) \\
&= \frac{1}{n}\mathbb{E}_{X\sim\pi}\sum_{i=1}^{n} \mathrm{Var}(f(X) \mid X_{\sim i}) .
\end{aligned}$$

Hence, the spectral gap inequality implies

$$\mathrm{Var}_{X\sim\pi}(f(X)) \le \frac{1}{1-\lambda_2(P)} \cdot \frac{1}{n}\mathbb{E}_{X\sim\pi}\sum_{i=1}^{n} \mathrm{Var}(f(X) \mid X_{\sim i}) . \tag{3}$$

Remark 1. If π is the product measure, i.e. if $X_1, \cdots, X_n$ are independent, the Efron–Stein inequality states that

$$\mathrm{Var}_{X\sim\pi}(f(X)) \le \mathbb{E}_{X\sim\pi} \sum_{i=1}^{n} \mathrm{Var}(f(X) \mid X_{\sim i})$$

which looks very similar to the inequality (3), the only difference being the multiplicative factor. Thus, the Efron–Stein inequality can be viewed as a Poincaré inequality for product measure. This is the topic of the next lecture.

Example 2. Suppose $X_1, \cdots, X_n \in \{\pm 1\}$ are random variables whose joint distribution function is positive everywhere. We may obtain a bound of the variance of $\sum_{i=1}^{n} X_i$ using the Glauber dynamics P of the joint distribution. Applying the inequality (3) gives

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) &\le \frac{1}{1-\lambda_2(P)} \cdot \frac{1}{n} \mathbb{E} \sum_{i=1}^{n} \mathrm{Var}\left(\sum_{i=1}^{n} X_i \,\middle|\, X_{\sim i}\right) \\
&= \frac{1}{1-\lambda_2(P)} \cdot \frac{1}{n} \mathbb{E} \sum_{i=1}^{n} \mathrm{Var}(X_i \mid X_{\sim i}) \\
&\le \frac{1}{1-\lambda_2(P)}
\end{aligned}$$

where the last inequality comes from the fact that the variance of a random variable on $\{\pm 1\}$ is at most 1.