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

The Efron-Stein Inequality

1 Review

Last time, we proved the equivalence of the following two statements:

1. (Poincare Inequality Form) If $X = (X_1, ..., X_n)$ has independent coordinates $\mu = \otimes_{i=1}^n \mu_i$, then $\forall f : \mathbb{R}^n \to \mathbb{R}$,
$$\mathrm{Var}\left(f(X)\right) \leq \sum_{i=1}^n \mathbb{E}_{x\sim\mu}[\mathrm{Var}\left(f(X)|X_{\sim i}\right)]$$
2. (Spectral Form) Let $P = \frac{1}{n}P_i$ be the Glauber dynamics, with $P_i f := \mathbb{E}[f(X)|X_{\sim i}]$, then $\lambda_2(P) = 1 - \frac{1}{n}$.

The two (equivalent) statements above are statements of the Efron-Stein Inequality, and the goal of this lecture is to prove the inequality and put it to some practical use.

Theorem 1 (Efron-Stein Inequality)**.** *If $X = (X_1, ..., X_n)$ has independent coordinates $\mu = \otimes_{i=1}^n \mu_i$, then $\forall f : \mathbb{R}^n \to \mathbb{R}$,*
$$\mathrm{Var}\left(f(X)\right) \leq \sum_{i=1}^n \mathbb{E}_{x\sim\mu}[\mathrm{Var}\left(f(X)|X_{\sim i}\right)]$$
or equivalently, if $P = \frac{1}{n}P_i$ be the Glauber dynamics, with $P_i f := \mathbb{E}[f(X)|X_{\sim i}]$, then $\lambda_2(P) = 1 - \frac{1}{n}$.

In fact, the Spectral Form of the Efron-Stein inequality can be weakened to $\lambda_2(P) \leq 1 - \frac{1}{n}$ by considering functions depending on one entry, but we'll stick to the strong form in these notes.

2 The Spectral Decomposition

What are the eigenvectors of P_i? First, lets introduce some notation regarding the tensor product of functions. Suppose that $f_1 : \Sigma_1 \to \mathbb{R}, ..., f_n : \Sigma_n \to \mathbb{R}$, then we define
$$(f_1 \otimes ... \otimes f_n)(x) := f_1(x_1) \cdot ... \cdot f_n(x_n)$$
where $(f_1 \otimes ... \otimes f_n) : \Sigma_1 \times ... \times \Sigma_n \to \mathbb{R}$.

Lemma 1. *Let $1 = \psi_i^{(1)}, ..., \psi_i^{(|\Sigma_i|)}$ be an $L^2(\mu_i)$ orthonormal basis, i.e. $\mathbb{E}_{x_i\sim\mu_i}[\psi_i^{(a)} \cdot \psi_i^{(b)}] = \mathbb{I}_{a=b}$. Then*
$$P_i(f_1 \otimes ... \otimes f_{i=1} \otimes \psi_i^{(a)} \otimes f_{i+1} \otimes ... \otimes f_n) = \mathbb{I}_{a=1} f_1 \otimes ... \otimes f_{i=1} \otimes \psi_i^{(a)} \otimes f_{i+1} \otimes ... \otimes f_n$$

In words, these specific functions produce eigenfunctions with eigenvalues either 0 or 1, and as we shall see, these functions (namely their tensor products) form a basis of the eigenspace of P. Note that the above functions exist by the Gram-Schmidt process. Before proceeding with a proof, let's consider a concrete example of such a basis.

Example 1. Let $\Sigma_1 = ... = \Sigma_n$ with $\mu_i \sim \mathrm{unif}(\Sigma_i)$. Then, by definition $\psi_i^{(1)}(x_i) = 1$, and then $\psi_i^{(2)}(x_i) = 1$ if $x_i = 1$ or -1 if $x_i = -1$, etc.

Proof of Lemma 1. We have that

$$P_i(f_1 \otimes ... \otimes f_{i=1} \otimes \psi_i^{(a)} \otimes f_{i+1} \otimes ... \otimes f_n)(x) = f_1(x_1) \cdot ... \cdot f_{i-1}(x_{i-1}) \cdot \mathbb{E}_\mu[\psi_i^{(a)} | x_{\sim i}] \cdot f_{i+1}(x_{i+1}) \cdot ... \cdot f_n(x_n)$$

Note that by definition, the expectation wraps around the entire term, but by linearity of conditional expectation, we end up with an expression containing an expectation over only the target term.

Plucking out the expectation portion, $\mathbb{E}_\mu[\psi_i^{(a)} | x_{\sim i}] = \mathbb{E}_{x_i \sim \mu_i}[\psi_i^{(a)}] = \mathbb{E}_{x_i \sim \mu_i}[\psi_i^{(1)} \cdot \psi_i^{(a)}(x_i)] = \mathbb{I}_{a=1}$, where the first equality follows from definition of the respective expectations, the second equality follows from the definition $\psi_i^{(1)} = 1$, and the third equality follows from orthonormality of the basis. In other words, the proof follows by unwinding the carefully laid definitions above. $\square$

With this technical lemma established, we're now in a position to finish up the proof of Efron-Stein.

Lemma 2. $\forall a_1 \in [|\Sigma_1|], ..., a_n \in [|\Sigma_n|]$, *we have that*

$$P(\psi_1^{(a_1)} \otimes ... \otimes \psi_n^{(a_n)}) = \frac{\#\{i : a_i = 1\}}{n} (\psi_1^{(a_1)} \otimes ... \otimes \psi_n^{(a_n)})$$

and this is a complete eigenbasis for P, where P represents the Glauber dynamics.

Proof.

$$P(\otimes_{i=1}^m \psi_i^{(a_i)}) = \frac{1}{n} \sum_{j=1}^n P_j(\otimes_{i=1}^m \psi_i^{(a_i)}) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}_{a_j=1} \psi_i^{(a_i)}$$

where again, our proof follows largely from the definitions in the setup. It follows that for the eigenvector $1 \otimes ... \otimes 1$ we have eigenvalue 1, and for $1 \otimes ... \otimes \psi_i^{(a)} \otimes ... \otimes 1$ we have eigenvalue $\frac{n-1}{n} \implies \lambda_2(P) = 1 - \frac{1}{n}$. $\square$

As the last line in the above proof indicates, Lemmas 1 and 2 establish the spectral form of the Efron-Stein Inequality. Having established our core result, we may now move on to applications involving control of variance of quantities of interest.

3 Applications

Let $x \in \{\pm 1\}^n$ be independent coordinates, then by Efron-Stein, we have that

Example 2 (Operator Norm)**.**

$$\mathrm{Var}\left(\langle w, x \rangle\right) \le \sum_{i=1}^n \mathbb{E}\mathrm{Var}\left(\langle w, x \rangle | x_{\sim i}\right) \le \sum_{i=1}^n w_i^2 = ||w||_2^2$$

and so if $\Sigma := \mathbb{E}[x \cdot x^T] - \mathbb{E}[x] \cdot \mathbb{E}[x^T]$, then

$$\implies |\Sigma|_{op} = \sup_{||w||_2 = 1} \langle w, \Sigma w \rangle \le 1$$

We state a basic technical fact, which we use in tandem with Efron-Stein to control the Rademacher complexity of a rectangularly-bounded set.

Fact 1. *X a r.v. such that $X \in [a, a + 2M]$, then* $\mathrm{Var}\left(X\right) \le M^2$.

Proof. This follows readily from observing that a random variable $X \in [-1, 1]$ has variance $\mathrm{Var}(X) \le 1$, and the general result follows by shifting and scaling. $\square$

Example 3 (Rademacher Complexity). $F \subseteq [-M, M]^n$, and we define

$$\mathrm{Rad}\,(F) := \mathbb{E}_{\sigma \sim \mathrm{unif}(\{\pm 1\}^n)} \sup_{f \in F} \langle \sigma, f \rangle$$

We now compute by Efron-Stein

$$\mathrm{Var}\,(\sup_{f \in F} \langle \sigma, f \rangle) \leq \sum_{i=1}^{n} \mathbb{E}_\sigma \mathrm{Var}\,(\sup_{f \in F} \langle \sigma, f \rangle | \sigma_{\sim i}) \leq O(nM^2)$$

Where the second inequality follows by noticing that by flipping σ_i, then $\langle \sigma, f \rangle$ changes by $\pm 2 f_i$ (which is at most $2M$).

Example 4 (A simple statistical application). $f^* \in F \subseteq \mathbb{R}^n$, and $y = f^* + \epsilon$ where ϵ is noise with independent coordinates ($\epsilon \in \{-\sigma, \sigma\}^n$), and we consider the Ordinary Least Squares Estimator (LSE), defined as

$$\hat{f} = \mathrm{argmin}_{f \in F} |y - f|_2^2$$

We have that

$$|f^* - \hat{f}|_2^2 \leq 2\langle \hat{f} - f^*, \epsilon \rangle \leq 2 \sup_{f \in F} \langle f - f^*, \epsilon \rangle$$

and so with high probability,

$$\leq 2\sigma \mathbb{E} \sup_{g \in F - f^*} \langle g, \epsilon \rangle + O((\mathrm{Var}\,(\sup_{g \in F - f^*})\langle g, \epsilon \rangle)^{1/2}) \leq 2\sigma \mathrm{Rad}\,(F - f^*) + O(\sigma M \sqrt{n})$$

4 Lipschitz Concentration

We now establish a more abstract definition, which we can use to address concentration of random variables with product distributions under Lipschitz functions. Note that in what follows, we aren't necessarily dealing with product distributions, as we are really appealing to the characterization of the spectral gap of the Glauber dynamics in terms of the corresponding Poincare inequality exhibited in the prior lecture. In the special case of product measures, we have $\lambda_2(P) = 1 - 1/n$, which is the content of the Efron-Stein Inequality.

Definition 1 (Hamming Distance and L-Lipschitz Functions). Letting the Hamming Distance be defined as $d_H(x, y) := \#\{i : x_i \neq y_i\}$, $f : \otimes_{i=1}^n \Sigma_i \to \mathbb{R}$ is L-Lipschitz with respect to d_H if

$$|f(x) - f(x')| \leq L d_H(x, y)$$

Example 5 (L-Lipschitz Functions). If f L-Lipschitz, then

$$\mathrm{Var}\,(f(X)) \leq \frac{1}{1 - \lambda_2(P)} \cdot \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\mathrm{Var}\,(f(X)|X_{\sim i}) \leq \frac{1}{1 - \lambda_2} \cdot \frac{L^2}{2}$$

where the first inequality follows from the corollary to the spectral theorem, and the second from the definition of Lipschitzness.

Lastly, we state but do not prove the following concentration of measure result for L-Lipschitz functions, which follows by recursively applying the Poincare inequality to the moment generating function.

Theorem 2.

$$\mathbb{P}[|f(X) - \mathbb{E}[f(X)]| \geq t] \leq 6 \exp\left(\frac{-ct\sqrt{1 - \lambda_2}}{L} \right)$$

for some constant c. Moreover, letting $f = \sum_{i=1}^n x_i$, suppose that $1 - \lambda_2 \asymp \frac{c}{n}$, then $f(X) - \mathbb{E}[f(X)] \in [-c\log(2/\delta)\sqrt{n}, c\log(2/\delta)\sqrt{n}]$ with probability $1 - \delta$.

A detailed proof of the above result can be found in [1], Section 4.4.

References

[1] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. 2014.