Lecture 1

Lecture date: Aug 27, 2007 *Scribe: David Rosenberg*

1 Introduction

Stein's method was invented in early 70s by Charles Stein as a method for proving central limit theorems.

1.1 Convergence in distribution

The cumulative distribution function of a random variable X is defined as

$$F(t) := \mathbf{P}(X \leq t).$$

Suppose we have a sequence of r.v. $\{X_n\}$ with c.d.f. $\{F_n\}$ and a r.v. X with c.d.f. X. We say that X_n converges in distribution (or converges in law, or converges weakly) to X if

$$\lim_{n\to\infty} F_n(t) = F(t) \text{ for all continuity points } t \text{ of } F.$$

This is denoted by $X_n \Longrightarrow X$ or $F_n \Longrightarrow F$ or $X_n \Longrightarrow F$. The following theorem is standard.

Theorem 1 *The following are equivalent:*

1. $X_n \Longrightarrow X$
2. $\mathbf{E}\, f(X_n) \to \mathbf{E}\, f(X)$ *for all bounded continuous* f
3. $\mathbf{E}\, f(X_n) \to \mathbf{E}\, f(X)$ *for all bounded Lipschitz* f
4. $\mathbf{E}\left(e^{itX_n}\right) \to \mathbf{E}\, e^{itX}$ *for all* t

Recall that the function $\varphi(t) := \mathbf{E}(e^{itX})$ is known as the characteristic function of X.

1.2 Central limit theorems

Recall: The standard gaussian distribution $N(0,1)$ has density

$$\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

We will usually denote standard gaussian r.v. by Z.

Basic central limit theorem: If $X_1, X_2, \dots$ iid r.v. with mean 0 variance 1 then

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}} \Longrightarrow N(0,1).$$

One standard method of proof uses characteristic functions.

$$\begin{aligned}
\mathbf{E}\left[e^{it\sum_1^n X_i/\sqrt{n}}\right] &= \mathbf{E}\left[e^{it\sum_1^n X_i/\sqrt{n}}\right]^n \\
&= \left[1 + \frac{it}{\sqrt{n}}\,\mathbf{E}\,X + \frac{(it)^2}{2n}\,\mathbf{E}\,X^2 + \cdots\right]^n \\
&\approx \left(1 - \frac{t^2}{2n}\right)^n \to e^{-t^2/2} = E\left(e^{itZ}\right).
\end{aligned}$$

1.3 Some examples that we will cover

We will apply Stein's method to situations where it's hard to apply standard arguments. Some examples are as follows.

Hoeffding's combinatorial CLT

Suppose π is a random (uniform) permutation of $\{1, \dots, n\}$, and consider the following distance from identity:

$$W_n = \sum_{i=1}^n |i - \pi(i)|$$

This is known as *Spearman's footrule.*

Known result: As n becomes large,

$$\frac{W_n - \mathbf{E}\,W_n}{\sqrt{\mathrm{Var}(W_n)}} \Longrightarrow N(0,1).$$

More generally, we have *Hoeffding's combinatorial CLT*.

- Array of numbers $(a_{ij})_{i\leq i,j\leq n}$ satisfying certain conditions.
- π a random permutation
- $W_n = \sum_i a_{i\pi(i)}$ (for spearman, $a_{ij} = |i-j|$)

How close is $\frac{W_n - \mathbf{E}\, W_n}{\sqrt{\mathrm{Var}(W_n)}}$ to $N(0,1)$? This was Stein's original motivation.

Linear statistics of eigenvalues

Suppose $(X_{ij})_{1\leq i,j\leq n}$ are iid rv's with mean 0 and variance 1, $X_{ji} = X_{ij}$. Then

$$A_n = \frac{1}{\sqrt{n}}(X_{ij})$$

is known as a *Wigner matrix.* Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A_n. Then it is known that

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i} \Longrightarrow \text{semicircle law}$$

which has density

$$\frac{1}{2\pi}\sqrt{4-x^2}$$

on $[-2,2]$.

We may want to look at fluctations of random distribution about a fixed distribution.

Look at $W_n = \sum_{i=1}^{n} f(\lambda_i)$.

Then $W_n - \mathbf{E}\, W_n \Longrightarrow N(0, \sigma^2(f))$. Main restrictions needed:

1. $\mathbf{E}\left(X_{ij}^{2m}\right) \leq (Cm)^m$ for all m.
2. X_{ij}'s have symmetric distribution around zero; not needed to be iid.

Original proof is by method of moments.

Curie-Weiss Model

- N magnetic particles, each with spin $+1$ or -1.
- Spins denoted by $\sigma_1, \ldots, \sigma_N$.

- The particles try to align themselves together with the same spin.
- Simplest model [mean-field model]: $P(\sigma) = Z^{-1} \exp\left(\frac{\beta}{n}\sum_{1\le i<j\le n}\sigma_i\sigma_j\right)$ where $\beta = 1/kT$ with T being the temperature and k the Boltzmann constant.
- The magnetization of the system

$$m(\sigma) = \frac{1}{n}\sum \sigma_i$$

If $\beta = 0$, then we have iid, and magnetization is close to 0.

- Known that for $\beta \le 1$, $m(\sigma) \to 0$ in probability as $n \to \infty$.
- For $\beta > 1$, the equation $x = \tanh(\beta x)$ has two solutions, $m^*(\beta)$ and $-m^*(\beta)$, say $m^* > 0$ and

$$m(\sigma) \Longrightarrow \frac{1}{2}\left(\delta_{m^*(\beta)} + \delta_{-m^*(\beta)}\right).$$

- If $\beta < 1$, then

$$\sqrt{n}m(\sigma) \Longrightarrow N(0, ?)$$

- The model has a phase transition at $\beta = 1$. If $\beta = 1$, then

$$n^{1/4}m(\sigma) \Longrightarrow \text{ the distribution with density } \propto e^{-x^4/12}.$$

Sherrington-Kirkpatrick Model

Spin glass model for N spins.

$$P(\sigma) = Z^{-1}\exp\left(\frac{\beta}{\sqrt{N}}\sum_{1\le i<j\le N} g_{ij}\sigma_i\sigma_j + h\sum\sigma_i\right)$$

where g_{ij} is a fixed realization of iid $N(0,1)$. The idea is that some particles try to align in the same direction, and some repel each other.

We will prove various results about this model using Stein's method.

If $h = 0$, it is known that $\beta = 1$ is the critical temperature.

Overlap: generate two vectors, σ^1 and σ^2 independently from the Gibbs measure.

$$R_{1,2} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^1\sigma_i^2$$

It is known that $R_{1,2} = O(\frac{1}{\sqrt{N}})$ if $\beta < 1$.

Open question: What is the magnitude of $R_{1,2}$ at $\beta = 1$?

Lecture 2

Lecture date: Aug 29, 2007 *Scribe: David Rosenberg*

2 Distances between probability measures

Stein's method often gives bounds on how close distributions are to each other.

A typical distance between probability measures is of the type

$$d(\mu,\nu) = \sup\left\{\left|\int f d\mu - \int f d\nu\right| \,:\, f \in \mathcal{D}\right\},$$

where $\mathcal{D}$ is some class of functions.

2.1 Total variation distance

Let $\mathcal{B}$ denote the class of Borel sets. The total variation distance between two probability measures μ and ν on $\mathbb{R}$ is defined as

$$\mathrm{TV}(\mu,\nu) := \sup_{A\in\mathcal{B}} |\mu(A) - \nu(A)|\,.$$

Here

$$\mathcal{D} = \{1_A \,:\, A \in \mathcal{B}\}\,.$$

Note that this ranges in $[0,1]$. Clearly, the total variation distance is not restricted to the probability measures on the real line, and can be defined on arbitrary spaces.

2.2 Wasserstein distance

This is also known as the Kantorovich-Monge-Rubinstein metric.

Defined only when probability measures are on a metric space.

$$\mathrm{Wass}(\mu,\nu) := \sup\left\{\left|\int f\,d\mu - \int f\,d\nu\right| : f \text{ is 1-Lipschitz}\right\},$$

i.e. sup over all f s.t. $|f(x) - f(y)| \leq d(x,y)$, d being the underlying metric on the space. The Wasserstein distance can range in $[0,\infty]$.

2.3 Kolmogorov-Smirnov distance

Only for probability measures on $\mathbb{R}$.

$$\begin{aligned}\mathrm{Kolm}(\mu,\nu) &:= \sup_{x\in\mathbb{R}} |\mu((-\infty,x]) - \nu((-\infty,x])| \\ &\le \mathrm{TV}(\mu,\nu).\end{aligned}$$

2.4 Facts

- All three distances defined above are stronger than weak convergence (i.e. convergence in distribution, which is weak* convergence on the space of probaility measures, seen as a dual space). That is, if any of these metrics go to zero as $n \to \infty$, then we have weak convergence. But converse is not true. However, weak convergence is metrizable (e.g. by the Prokhorov metric).

- Important coupling interpretation of total variation distance:

$$TV(\mu,\nu) = \inf\{P(X \neq Y) : (X,Y) \text{ is a r.v. s.t. } X \sim \mu, Y \sim \nu\}$$

  (i.e. infimum over all joint distributions with given marginals.)

- Similarly, for μ, ν on the real line,

$$\mathrm{Wass}(\mu,\nu) = \inf\{\mathbf{E}\,|X-Y| : (X,Y) \text{ is a r.v. s.t. } X\sim\mu, Y\sim\nu\}$$

  (So it's often called the Wass_1, because of L_1 norm.)

- TV is a very strong notion, often too strong to be useful. Suppose $X_1, X_2, \ldots$ iid ± 1. $S_n = \sum_1^n X_i$. Then

$$\frac{S_n}{\sqrt{n}} \Longrightarrow N(0,1)$$

  But $TV(\frac{S_n}{\sqrt{n}}, Z) = 1$ for all n. Both Wasserstein and Kolmogorov distances go to 0 at rate $1/\sqrt{n}$.

Lemma 2 *Suppose W, Z are two r.v.'s and Z has a density w.r.t. Lebesgue measure bounded by a constant C. Then* $\mathrm{Kolm}(W,Z) \le 2\sqrt{C\mathrm{Wass}(W,Z)}$.

Proof: Consider a point t, and fix an ϵ. Define two functions g_1 and g_2 as follows. Let $g_1(x) = 1$ on $(-\infty,t)$, 0 on $[t+\epsilon,\infty)$ and linear interpolation in between. Let $g_2(x) = 1$ on $(-\infty, t-\epsilon]$, 0 on $[t,\infty)$, and linear interpolation in between. Then g_1 and g_2 form upper and lower 'envelopes' for $1_{(-\infty,t]}$. So

$$P(W \le t) - P(Z \le t) \le \mathbf{E}\,g_1(W) - \mathbf{E}\,g_1(Z) + \mathbf{E}\,g_1(Z) - P(Z \le T).$$

Now $\mathbf{E}\, g_1(W) - \mathbf{E}\, g_1(Z) \le \frac{1}{\epsilon}\text{Wass}(W, Z)$ since g_1 is $(1/\epsilon)$-Lipschitz, and $\mathbf{E}\, g_1(Z) - P(Z \le t) \le C\epsilon$ since Z has density bdd by C.

Now using g_2, same bound holds for the other side: $P(Z \le t) - P(W \le t)$. Optimize over ϵ to get the required bound. □

2.5 A stronger notion of distance

Exercise 1: S_n a simple random walk (SRW). $S_n = \sum_1^n X_i$, with X_i iid ± 1. Then

$$\frac{S_n}{\sqrt{n}} \Longrightarrow Z \sim N(0, 1).$$

The Berry-Esseen bound: Suppose $X_1, X_2, \ldots$ iid $\mathbf{E}(X_1) = 0, \mathbf{E}(X_1^2) = 1, \mathbf{E}\,|X|^3 < \infty$. Then

$$\text{Kolm}\left(\frac{S_n}{\sqrt{n}}, Z\right) \le \frac{3\,\mathbf{E}\,|X_1|^3}{\sqrt{n}}$$

Can also show that for SRW,

$$\text{Wass}\left(\frac{S_n}{\sqrt{n}}, Z\right) \le \frac{Const}{\sqrt{n}}$$

This means that it is possible to construct $\frac{S_n}{\sqrt{n}}$ and Z on the same space such that

$$\mathbf{E}\left|\frac{S_n}{\sqrt{n}} - Z\right| \le \frac{C}{\sqrt{n}}$$

Can we do it in the strong sense? That is:

$$P\left(\left|\frac{S_n}{\sqrt{n}} - Z\right| > \frac{t}{\sqrt{n}}\right) \le Ce^{-ct}.$$

This is known as Tusnády's Lemma. Will come back to this later.

3 Integration by parts for the gaussian measure

The following result is sometimes called 'Stein's Lemma'.

Lemma 3 *If $Z \sim N(0, 1)$, and $f : \mathbb{R} \to \mathbb{R}$ is an absolutely continuous function such that $\mathbf{E}\,|f'(Z)| < \infty$, then $\mathbf{E}\, Zf(Z) = \mathbf{E}\, f'(Z)$.*

Proof: First assume f has compact support contained in (a, b). Then the result follows from integration by parts:

$$\int_a^b xf(x)e^{-x^2/2}dx = \left[f(x)e^{-x^2/2}\right]_a^b + \int_a^b f'(x)e^{-x^2/2}\,dx.$$

Now take any f s.t. $\mathbf{E}\,|Zf(Z)| < \infty, \mathbf{E}\,|f'(Z)| < \infty, \mathbf{E}\,|f(Z)| < \infty$.

Take a piecewise linear function g that takes value 1 in $[-1, 1]$, 0 outside $[-2, 2]$, and between 0 and 1 elsewhere. Let

$$f_n(x) := f(x)g(x/n).$$

Then clearly,

$$|f_n(x)| \le |f(x)| \text{ for all } x \text{ and } f_n(x) \to f(x) \text{ pointwise.}$$

Similarly, $f_n' \to f'$ pointwise. Rest follows by DCT. The last step is to show that the finiteness of $\mathbf{E}\,|f'(Z)|$ implies the finiteness of the other two expectations.

Suppose $\mathbf{E}\,|f'(Z)| < \infty$. Then

$$\begin{aligned}\int_0^\infty |xf(x)|\, e^{-x^2/2}\,dx &\le \int_0^\infty x \int_0^x \left|f'(y)\right|\,dy\, e^{-x^2/2}\,dx \\ &= \int_0^\infty \left|f'(y)\right| \underbrace{\int_y^\infty xe^{-x^2/2}dx}_{e^{-y^2/2}}\,dy.\end{aligned}$$

Finiteness of $\mathbf{E}\,|f(Z)|$ follows from the inequality $|f(x)| \le \sup_{|t|\le 1} |f(t)| + |xf(x)|$. □

Exercise 2: Find f s.t. $\mathbf{E}\,|Zf(Z)| < \infty$ but $\mathbf{E}\,|f'(Z)| = \infty$.

Next time, Stein's method. Sketch:

Suppose you have a r.v. W and $Z \sim N(0, 1)$ and you want to bound

$$\sup_{g\in\mathcal{D}} |\mathbf{E}\, g(W) - \mathbf{E}\, g(Z)| \le \sup_{f\in\mathcal{D}'} \left|\mathbf{E}\left(f'(W) - Wf(W)\right)\right|$$

Main difference between stein's method and characteristic functions is that Stein's method is a *local* technique. We transfer a *global* problem to a local problem. It's a theme that is present in many branches of mathematics.

Lecture 3

Lecture date: Aug 31, 2007 *Scribe: David Rosenberg*

4 First step in Stein's method

Suppose you have a r.v. W and a standard gaussian r.v. Z and you want to bound

$$\sup_{g\in\mathcal{D}} |\mathbf{E}\, g(W) - \mathbf{E}\, g(Z)|,$$

where $\mathcal{D}$ is some given class of functions. The first step is to find another class of functions $\mathcal{D}'$ such that

$$\sup_{g\in\mathcal{D}} |\mathbf{E}\, g(W) - \mathbf{E}\, g(Z)| \le \sup_{f\in\mathcal{D}'} \left|\mathbf{E}\left(f'(W) - Wf(W)\right)\right|. \tag{1}$$

Stein's idea: If $\mathcal{D}'$ is a class of functions such that for every $g \in \mathcal{D}$, $\exists f \in \mathcal{D}'$ s.t.

$$f'(x) - xf(x) = g(x) - \mathbf{E}\, g(Z) \tag{2}$$

for $Z \sim N(0,1)$, then (1) holds. (This o.d.e. is sometimes called the 'Stein equation'.) Indeed, take any $g \in \mathcal{D}$, and find $f \in \mathcal{D}'$ that solves the above equation. Then

$$\begin{aligned}\mathbf{E}\, g(W) - \mathbf{E}\, g(Z) &= \mathbf{E}\left[g(W) - \mathbf{E}\, g(Z)\right] \\ &= \mathbf{E}\left(f'(W) - Wf(W)\right).\end{aligned}$$

Clearly, given $\mathcal{D}$ it is in our interest to have $\mathcal{D}'$ as small as possible.

Lemma 4 (Stein) *Given a function $g : \mathbb{R} \to \mathbb{R}$ that is bounded, $\exists$ absolutely continuous f solving $f'(x) - xf(x) = g(x) - \mathbf{E}\, g(Z)$ for all x, satisfying*

$$|f|_\infty \le \sqrt{\frac{\pi}{2}}\, |g - Ng|_\infty \quad \textit{and} \quad |f'|_\infty \le 2\,|g - Ng|_\infty$$

(where $|f|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$, $Ng := \mathbf{E}\, g(Z), Z \sim N(0,1)$).

If g is Lipschitz, but not necessarily bounded, then

$$|f|_\infty \le |g'|_\infty, \quad |f'|_\infty \le \sqrt{\frac{2}{\pi}}\, |g'|_\infty, \quad \textit{and} \quad |f''|_\infty \le 2\,|g'|_\infty.$$

Actually, the third and fourth inequalities are not due to Stein, but were obtained later.

Exercise 1: Show that all five constants are the best possible.

So we can now take $\mathcal{D}'$ to be these f's (which in particular have the given bounds).

The above lemma tells us, for instance, that

$$\mathrm{Wass}(W,Z) \le \sup\left\{\left|\mathbf{E}\left(f'(W) - Wf(W)\right)\right| : |f|_\infty \le 1, |f'|_\infty \le \sqrt{2/\pi}, |f''|_\infty \le 2\right\}.$$

5 Example: Ordinary CLT in the Wasserstein metric

Suppose $X_1, X_2, \ldots, X_n$ are independent, mean 0, variance 1, $\mathbf{E}\,|X_i|^3 < \infty$. Let $S_n = \sum_1^n X_i$. Take any $f \in C^1$ with f' absolutely continuous, and satisfying $|f| \le 1$,$|f'| \le \sqrt{2/\pi}$, and $|f''| \le 2$. First, note that

$$\mathbf{E}\,Wf(W) = \frac{1}{\sqrt{n}} \sum \mathbf{E}\left(X_i f(W)\right). \tag{3}$$

Now let

$$W_i = W - \frac{X_i}{\sqrt{n}} = \frac{\sum_{j\neq i} X_j}{\sqrt{n}}$$

Then X_i, W_i are independent. Thus

$$\mathbf{E}\,X_i f(W_i) = \underbrace{\mathbf{E}(X_i)}_{=0}\,\mathbf{E}\,f(W_i) = 0$$

and so

$$\begin{aligned}\mathbf{E}\left(X_i f(W)\right) &= \mathbf{E}\left(X_i\left(f(W) - f(W_i)\right)\right)\\ &= \mathbf{E}\left(X_i\left(f(W) - f(W_i) - (W - W_i)f'(W_i)\right)\right)\\ &\quad + \mathbf{E}\left[X_i(W - W_i)f'(W_i)\right].\end{aligned}$$

Note that

$$\left|f(b) - f(a) - (b-a)f'(a)\right| \le \frac{1}{2}(b-a)^2\left|f''\right|_\infty$$

and that $W - W_i = X_i/\sqrt{n}$. Thus

$$\begin{aligned}&\left|\mathbf{E}\left[X_i\left(f(W) - f(W_i) - \frac{X_i}{\sqrt{n}}f'(W_i)\right)\right]\right|\\ &\le \frac{1}{2}\,\mathbf{E}\left|X_i\frac{X_i^2}{n}\right|\cdot\left|f''\right|_\infty \le \frac{1}{n}\,\mathbf{E}\,|X_i|^3.\end{aligned}$$

Again,

$$\begin{aligned}\mathbf{E}\left[X_i\,(W-W_i)\,f'(W_i)\right] &= \frac{1}{\sqrt{n}}\,\mathbf{E}\,X_i^2 f'(W_i)\\ &= \frac{1}{\sqrt{n}}\,\mathbf{E}\,f'(W_i)\end{aligned}$$

since $\mathbf{E}\,X_i^2 = 1$ and X_i is independent of W_i.

From (3) and the above calculation we see that

$$\left|\mathbf{E}\,W f(W) - \frac{1}{n}\sum \mathbf{E}\,f'(W_i)\right| \le \frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbf{E}\,|X_i|^3 .$$

Finally, note that

$$\begin{aligned}\left|\frac{1}{n}\sum \mathbf{E}\,f'(W_i) - \mathbf{E}\,f'(W)\right| &\le \frac{|f''|_\infty}{n}\sum \mathbf{E}\,|W-W_i|\\ &= \frac{|f''|_\infty}{n^{3/2}}\sum \mathbf{E}\,|X_i| \le \frac{2}{n^{3/2}}\sum \mathbf{E}\,|X_i| .\end{aligned}$$

Combining, we have

$$\begin{aligned}&\left|\mathbf{E}\,f(W)W - \mathbf{E}\,f'(W)\right|\\ &\le \frac{1}{n^{3/2}}\sum \mathbf{E}\,|X_i|^3 + \frac{2}{n^{3/2}}\sum \mathbf{E}\,|X_i| .\end{aligned}$$

Since $\mathbf{E}\,X_i^2 = 1$ we can conclude that $\mathbf{E}\,|X_i|^3 \ge 1$ and hence $\mathbf{E}\,|X_i| \le \left(\mathbf{E}\,|X_i|^3\right)^{1/3} \le \mathbf{E}\,|X_i|^3$. We have now arrived at a 'Berry-Esséen bound' for the Wasserstein metric:

Theorem 5 *Suppose $X_1, \ldots, X_n$ are independent with mean 0, variance 1, and finite third moments. Then*

$$\mathrm{Wass}\left(\frac{\sum_1^n X_i}{\sqrt{n}},\, Z\right) \le \frac{3}{n^{3/2}}\sum_1^n \mathbf{E}\,|X_i|^3 ,$$

where $Z \sim N(0,1)$.

Unfortunately, this isn't a real Berry-Esséen bound, since it's a bound on the Wasserstein metric and not the Kolmogorov metric. From a lemma proved in Lecture 2, we can get

$$\mathrm{Kolm}(W,Z) \le 2\sqrt{\frac{1}{\sqrt{2\pi}}\mathrm{Wass}(W,Z)} = \frac{2}{(2\pi)^{1/4}}\sqrt{\mathrm{Wass}(W,Z)}.$$

But this is of order $n^{-1/4}$, which is suboptimal.

Exercise 2: Get the true Berry-Esséen bound using Stein's method. This involves analyzing the solution of the Stein equation (2) for $g(x) = 1_{\{x \leq t\}}$ for arbitrary $t \in \mathbb{R}$.

Exercise 3: Consider Erdős-Rényi graph $G(n, p)$. Has n vertices and $\binom{n}{2}$ possible edges, each edge being open or closed with prob p and $1 - p$, independently of each other. Let T_n = number of triangles in this graph. Find a way to use Stein's method to prove the CLT for T_n when (a) p is fixed, and (b) p is allowed to go to zero with n.

Lecture 4

Lecture date: Sep 5, 2007 *Scribe: Arnab Sen*

In this lecture we are going to study the solution of the differential equation

$$f'(x) - xf(x) = g(x) - \mathbf{E}g(Z), \quad Z \sim N(0,1). \tag{4}$$

Lemma 6 *Given function $g : \mathbb{R} \to \mathbb{R}$ such that $\mathbf{E}|g(Z)| < \infty$ where $Z \sim N(0,1)$,*

$$f(x) = e^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2}(g(y) - \mathbf{E}g(Z))dy \tag{5}$$

is an absolutely continuous solution of (4).

Moreover, any a.c. solution $\tilde{f}$ of (4) *is of the form*

$$\tilde{f}(x) = f(x) + ce^{x^2/2}, \quad c \in \mathbb{R}.$$

Finally, f is the only solution that satisfies $\lim_{|x|\to\infty} f(x)e^{-x^2/2} = 0$.

Proof: By the method of integrating factors, we have that if f is a solution to (4), then

$$\frac{d}{dx}(e^{-x^2/2}f(x)) = e^{-x^2/2}(f'(x) - xf(x)) = e^{-x^2/2}(g(x) - \mathbf{E}g(Z)).$$

So, (5) is a reasonable candidate as a solution of (4). And it is easy to verify directly that (5) indeed satisfies (4).

If $\tilde{f}$ is any other solution of (4), then

$$\frac{d}{dx}\left(e^{-x^2/2}(f(x) - \tilde{f}(x))\right) = 0.$$

Hence, $\tilde{f}(x) = f(x) + ce^{x^2/2}$ for some $c \in \mathbb{R}$.

Clearly, from definition

$$\lim_{x\to-\infty} f(x)e^{-x^2/2} = 0 \quad \text{(by DCT)}.$$

Note that since $Z \sim N(0,1)$, we have

$$\int_{-\infty}^{\infty} e^{-y^2/2}(g(y) - \mathbf{E}g(Z))dy = 0.$$

So, f can also be written as follows

$$f(x) = -e^{x^2/2} \int_x^\infty e^{-y^2/2}(g(y) - \mathbf{E}g(Z))dy. \tag{6}$$

Therefore, by DCT, $\lim_{x\to+\infty} f(x)e^{-x^2/2} = 0$.

□

Remark 7 *If, instead of standard gaussian, Z follows any other distribution then all of the statements of the above lemma still hold except* $\lim_{x\to+\infty} f(x)e^{-x^2/2} = 0$.

5.1 Another form of the solution

Lemma 8 *Assume g is Lipschitz. Then*

$$f(x) = -\int_0^1 \frac{1}{2\sqrt{t(1-t)}} \mathbf{E}\left[Zg(\sqrt{t}x + \sqrt{1-t}Z)\right] dt, \quad Z \sim N(0,1) \tag{7}$$

is a solution of (4)*. In fact, it must be the same as* (5)*, because* $\lim_{|x|\to\infty} f(x)e^{-x^2/2} = 0$.

Proof: Let g is C-Lipschitz. Then[1] $|g'|_\infty \le C$.

On differentiating f and carrying the derivative inside the integral and expectation which can be justified using DCT, we have

$$f'(x) = -\int_0^1 \frac{1}{2\sqrt{1-t}} \mathbf{E}\left[Zg'(\sqrt{t}x + \sqrt{1-t}Z)\right] dt. \tag{8}$$

On the other hand, the Stein identity gives us

$$\mathbf{E}\left[Zg(\sqrt{t}x + \sqrt{1-t}Z)\right] = \sqrt{1-t}\mathbf{E}\left[g'(\sqrt{t}x + \sqrt{1-t}Z)\right].$$

Thus,

$$\begin{aligned} f'(x) - xf(x) &= \int_0^1 \mathbf{E}\left[\left(-\frac{Z}{2\sqrt{1-t}} + \frac{x}{2\sqrt{t}}\right) g'(\sqrt{t}x + \sqrt{1-t}Z)\right] dt \\ &= \int_0^1 \mathbf{E}\left[\frac{d}{dt} g'(\sqrt{t}x + \sqrt{1-t}Z)\right] dt \\ &= \mathbf{E}\left[\int_0^1 \frac{d}{dt} g'(\sqrt{t}x + \sqrt{1-t}Z) dt\right] = g(x) - \mathbf{E}g(Z). \end{aligned}$$

[1] Any Lipschitz function g is absolutely continuous. Hence, it is (Lebesgue) almost surely differentiable. Define g' to be derivative of g at the points where it exists and 0 elsewhere.

□

Recall the notation $Ng := \mathbf{E}g(Z)$. Now we will prove that if $g : \mathbb{R} \to \mathbb{R}$ is bounded,

$$I.\ |f|_\infty \le \sqrt{\frac{\pi}{2}}\,|g - Ng|_\infty \quad \text{and} \quad II.\ \left|f'\right|_\infty \le 2\,|g - Ng|_\infty$$

and if g is Lipschitz, but not necessarily bounded, then

$$III.\ |f|_\infty \le \left|g'\right|_\infty,\quad IV.\ \left|f'\right|_\infty \le \sqrt{\frac{2}{\pi}}\,\left|g'\right|_\infty,\quad \text{and} \quad V.\ \left|f''\right|_\infty \le 2\left|g'\right|_\infty.$$

This will prove the Lemma 4 of Lecture 3. The bounds (I), (II) and (V) were obtained by Stein.

Proof of bound (III) : Applying Stein's identity on (7), we have

$$f(x) = -\int_0^1 \frac{1}{2\sqrt{t}}\mathbf{E}\left[g'(\sqrt{t}x + \sqrt{1-t}Z)\right] dt.$$

Hence,

$$|f|_\infty \le |g'|_\infty \int_0^1 \frac{1}{2\sqrt{t}} = |g'|_\infty.$$

□

Proof of bound (IV) : From (8), it follows that

$$|f|_\infty \le (\mathbf{E}|Z|)\,|g'|_\infty \int_0^1 \frac{1}{2\sqrt{1-t}} = \sqrt{\frac{2}{\pi}}|g'|_\infty.$$

□

Exercise 9 *Get the bound (V) from the representation* (7).

Proof of bound (I) : Take f as in (5). Suppose $x > 0$. Using the representation in (6), we have

$$|f(x)| \le |g - Ng|_\infty \left(e^{x^2/2}\int_x^\infty e^{-y^2/2}dy\right).$$

Now, $\frac{d}{dx}e^{x^2/2}\int_x^\infty e^{-y^2/2}dy = -1 + xe^{x^2/2}\int_x^\infty e^{-y^2/2}dy \le 0 \quad \forall x > 0$. The last step follows from Mill's ratio inequality which says that $\int_x^\infty e^{-y^2/2}dy \le \frac{1}{x}e^{-x^2/2}$ (for a quick proof, note that LHS $\le \int_x^\infty \frac{y}{x}e^{-y^2/2}dy =$ RHS).

So, $e^{x^2/2}\int_x^\infty e^{-y^2/2}dy$ is maximized at $x = 0$ on $[0, \infty)$ where its value is $\sqrt{\frac{\pi}{2}}$. Hence,

$$|f(x)| \le \sqrt{\frac{\pi}{2}}|g - Ng|_\infty \quad \forall x > 0.$$

For $x < 0$, use the form (5) and proceed in the similar manner. $\square$

Proof of bound (II) : Again, we will only consider $x > 0$ case. The other case will be similar.

Note that

$$f'(x) = g(x) - Ng + xf(x) = g(x) - Ng - xe^{x^2/2} \int_x^\infty e^{-y^2/2}(g(y) - Ng)dy.$$

Therefore,

$$\begin{aligned} |f'(x)| &\le |g - Ng|_\infty \left(1 + xe^{x^2/2} \int_x^\infty e^{-y^2/2} dy\right) \\ &\le 2|g - Ng|_\infty \quad \text{(By Mill's ratio inequality)}. \end{aligned}$$

$\square$

Proof of bound (V) : On differentiating (4) and rearranging

$$\begin{aligned} f''(x) &= g'(x) + f(x) + xf'(x) \\ &= g'(x) + f(x) + x(g(x) - Ng + xf(x)) \\ &= g'(x) + x(g(x) - Ng) + (1 + x^2)f(x). \end{aligned} \tag{9}$$

We can write $g(x) - Ng$ in terms of g' as follows,

$$\begin{aligned} g(x) - Ng &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-y^2/2}(g(x) - g(y))dy \\ &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^x \int_y^x g'(z)e^{-y^2/2}dzdy - \int_x^\infty \int_x^y g'(z)e^{-y^2/2}dzdy\right] \\ &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^x g'(z) \int_{-\infty}^z e^{-y^2/2}dydz - \int_x^\infty g'(z) \int_z^\infty e^{-y^2/2}dydz\right] \\ &= \int_{-\infty}^x g'(z)\Phi(z)dz - \int_x^\infty g'(z)\overline{\Phi}(z)dz \end{aligned}$$

where $\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy$ is the distribution function for standard normal and $\overline{\Phi}(z) = 1 - \Phi(z)$.

Similarly,

$$\begin{aligned}
f(x) &= e^{x^2/2}\int_{-\infty}^{x} e^{-y^2/2}(g(y)-\mathbf{E}g(Z))dy \\
&= e^{x^2/2}\int_{-\infty}^{x} e^{-y^2/2}\left(\int_{-\infty}^{y} g'(z)\Phi(z)dz - \int_{y}^{\infty} g'(z)\overline{\Phi}(z)\right)dzdy \\
&= e^{x^2/2}\left(\int_{-\infty}^{x} g'(z)\Phi(z)\int_{z}^{x} e^{-y^2/2}dydz - \int_{-\infty}^{\infty} g'(z)\overline{\Phi}(z)\int_{-\infty}^{z\wedge x} e^{-y^2/2}dydz\right) \\
&= \sqrt{2\pi}e^{x^2/2}\left(\int_{-\infty}^{x} g'(z)\Phi(z)(\overline{\Phi}(z)-\overline{\Phi}(x))dz\right. \\
&\qquad \left. - \int_{-\infty}^{x} g'(z)\overline{\Phi}(z)\Phi(z)dz - \int_{x}^{\infty} g'(z)\overline{\Phi}(z)\Phi(x)dz\right) \\
&= -\sqrt{2\pi}e^{x^2/2}\left[\overline{\Phi}(x)\int_{-\infty}^{x} g'(z)\Phi(z)dz + \Phi(x)\int_{x}^{\infty} g'(z)\overline{\Phi}(z)dz\right]
\end{aligned}$$

Substituting the above expressions for $g - Ng$ and f in (9), we get

$$\begin{aligned}
f''(x) = g'(x) &+ \left(x - \sqrt{2\pi}(1+x^2)e^{x^2/2}\overline{\Phi}(x)\right)\int_{-\infty}^{x} g'(z)\Phi(z)dz \\
&+ \left(-x - \sqrt{2\pi}(1+x^2)e^{x^2/2}\Phi(x)\right)\int_{x}^{\infty} g'(z)\overline{\Phi}(z)dz.
\end{aligned}$$

To be continued in the next lecture.

Lecture 5

Lecture date: Sept 7, 2007 *Scribe: Guy Bresler*

6 Continuation of Stein Bound

In the previous lecture we proved bounds on f and its derivatives, f satisfying

$$f'(x) - xf(x) = g(x) - Ng,$$

where $Ng = \mathbf{E}\, g(Z)$ and $Z \sim N(0,1)$.

We are in the process of bounding $|f''|_\infty$ under the assumption that g is Lipschitz. Continuing from the previous lecture, we have

$$\begin{aligned} f''(x) = g'(x) &+ \left(x - \sqrt{2\pi}(1+x^2)e^{x^2/2}(1-\Phi(x))\right)\int_{-\infty}^{x} g'(z)\Phi(z)dz \\ &+ \left(-x - \sqrt{2\pi}(1+x^2)e^{x^2/2}\Phi(x)\right)\int_{x}^{\infty} g'(z)(1-\Phi(z))dz. \end{aligned}$$

This gives

$$\begin{aligned} |f''(x)|_\infty \le |g'|_\infty \Bigg[1 &+ \left|x - \sqrt{2\pi}(1+x^2)e^{x^2/2}(1-\Phi(x))\right| \int_{-\infty}^{x} \Phi(z)dz \\ &+ \left|-x - \sqrt{2\pi}(1+x^2)e^{x^2/2}\Phi(x)\right| \int_{x}^{\infty} (1-\Phi(z))dz\Bigg]. \end{aligned} \tag{10}$$

Recall the Mill's ratio inequality on $\Phi(x)$ for $x > 0$:

$$\frac{xe^{x^2/2}}{\sqrt{2\pi}(1+x^2)} \le 1 - \Phi(x) \le \frac{e^{x^2/2}}{x\sqrt{2\pi}}. \tag{11}$$

Exercise 10 *Prove the left inequality in* (11).

There is a similar bound for $x \le 0$. To proceed, we wish to remove the absolute values in equation (10), by determining the sign of the expressions within the absolute value. From the Mill's ratio (11) we have

$$x + \sqrt{2\pi}(1+x^2)e^{x^2/2}\Phi(x) > 0 \tag{12a}$$

and

$$-x + \sqrt{2\pi}(1 + x^2)e^{x^2/2}(1 - \Phi(x)) > 0. \tag{12b}$$

You can check (12b) by noting that for $x < 0$ the inequality is obvious, and for $x > 0$ use the lower Mill's ratio inequality; (12a) follows similarly. Hence both expressions within the absolute values in equation (10) are negative.

To finish the simplification, observe that integration by parts gives

$$\int_{-\infty}^{x} \Phi(z)dz = x\Phi(x) + \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

and

$$\int_{x}^{\infty} (1 - \Phi(z))dz = -x(1 - \Phi(x)) + \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

Combining, we get

$$\begin{aligned} |f''(x)| \le |g'|_\infty \Bigg[1 &+ \left(-x + \sqrt{2\pi}(1 + x^2)e^{x^2/2}(1 - \Phi(x))\right)\left(x\Phi(x) + \frac{e^{-x^2/2}}{\sqrt{2\pi}}\right) \\ &+ \left(x + \sqrt{2\pi}(1 + x^2)e^{x^2/2}\Phi(x)\right)\left(-x(1 - \Phi(x)) + \frac{e^{-x^2/2}}{\sqrt{2\pi}}\right)\Bigg] \\ = 2|g'|_\infty. \end{aligned} \tag{13}$$

This proves the desired bound.

The factor 2 in the bound above turns out to be sharp. In contrast to the calculation above, it is easy to attain a factor of 4, as follows. First we take the derivative of the equation

$$f'(x) - xf(x) = g(x) - Ng$$

to get

$$f''(x) - xf'(x) - f(x) = g'(x),$$

or

$$f''(x) - xf'(x) = g'(x) + f(x) := h(x).$$

Since $Nh = Eh(Z) = Ef''(Z) - EZf'(Z) = 0$ (from Stein's Lemma in lecture 2), we see that f' is a solution of the Stein equation with h. The triangle inequality and one of the earlier Stein bounds give

$$|g' + f| \le |g'| + |f| \le 2|g'|_\infty,$$

hence

$$|f''|_\infty \le 2|g' + f|_\infty \le 4|g'|_\infty.$$

This completes the discussion of the five Stein bounds.

7 Dependency Graphs

Let $\{X_i, i \in V\}$ be a collection of random variables, and $G = (V, E)$ be a graph with vertex set V.

Definition 11 *G is called a* dependency graph *for $\{X_i, i \in V\}$ if the following holds: for any two subsets of vertices $S, T \subseteq V$ such that there is no edge from any vertex in S to any vertex in T, the collections $\{X_i, i \in S\}$ and $\{X_i, i \in T\}$ are independent.*

The idea behind the usefulness of dependency graphs is that if the max degree is not too large, we get a CLT. Note that there is not a unique dependency graph (for example, the complete graph works for any set of r.v.).

Example 12 *Suppose $Y_1, Y_2, \ldots, Y_{n+1}$ are independent random variables, and let $X_i = Y_i Y_{i+1}$. We will want to study the behavior of*

$$\sum_i X_i = \sum_i Y_i Y_{i+1}.$$

A dependency graph for $\{X_i, i \in V\}$ with $V = \{1, \ldots, n\}$ is given by the graph with edge set $\{(i, i+1); 1 \leq i \leq n-1\}$.

Given a graph G, let $D = 1 +$ maximum degree of G. We have the following lemma.

Lemma 13 *Let $S = \sum_{i \in V} X_i$. Then $\mathrm{Var}(S) \leq D \sum_{i \in V} \mathrm{Var}(Xi)$.*

Proof: Assume without loss of generality that $\mathbf{E}(X_i) = 0$ for all $i \in V$. We write $j \sim i$ if j is a neighbor of i or $j = i$. Then

$$\mathrm{Var}(S) = \sum_{i,j} \mathbf{E}(X_i X_j) \overset{(a)}{=} \sum_{i,j \sim i} \mathbf{E}(X_i X_j) \overset{(b)}{\leq} \sum_{i,j \sim i} \frac{\mathbf{E}\, X_i^2 + \mathbf{E}\, X_j^2}{2} \leq D \sum_{i \in V} \mathrm{Var}(X_i),$$

where (a) follows by the zero-mean assumption and independence, and (b) from the AM-GM inequality ($ab \leq (a^2 + b^2)/2$). □

In the next lecture we will use the lemma to prove the following theorem. Let $\sigma^2 = \mathrm{Var}(\sum X_i)$ and $W = \frac{\sum X_i}{\sigma}$, where it is assumed that $\mathbf{E}(X_i) = 0$.

Theorem 14 *It holds that*

$$\mathrm{Wass}(W, Z) \leq \frac{4}{\sqrt{\pi}\sigma^2} \sqrt{D^3 \sum \mathbf{E}\,|X_i|^4} + \frac{D^2}{\sigma^3} \sum \mathbf{E}\,|X_i|^3,$$

where $Z \sim N(0, 1)$.

Remark 15 *The bound in the theorem is often tight. We can get a bound on the Kolmogorov metric from the bound*

$$\mathrm{Kolm}(W, Z) \leq \frac{2}{(2\pi)^{1/4}} \sqrt{\mathrm{Wass}(W, Z)},$$

but this is not a good bound.

Lecture 6

Lecture date: Sep 10, 2007 *Scribe: Allan Sly*

8 Method of Dependency Graphs

We recall the definition of a dependency graph from the previous lecture. For a collection of random variables $\{X_i, i \in V\}$ indexed by the vertices V of a graph $G = (V, E)$ we say that G is a dependency graph if for any disjoint subsets $S, T \subseteq V$ with no edges between S and T we have $\{X_i, i \in S\}$ and $\{X_i, i \in T\}$ independent. Let $D = 1 +$ max degree G.

Lemma 16 *Suppose that* $E(X_i) = 0, \sigma^2 = \mathrm{Var}(\sum X_i), W = \frac{\sum X_i}{\sigma}$ *and* $Z \sim N(0,1)$. *Then*

$$\mathrm{Wass}(W, Z) \leq \frac{4}{\sqrt{\pi}\sigma^2}\sqrt{D^3 \sum E|X_i|^4} + \frac{D^2}{\sigma^3}\sum E|X_i|^3.$$

Proof: Let $W_i = \frac{1}{\sigma}\sum_{j \in N_i} X_j$ where $N_i = \{i\} \cup \{\text{neighbours of i}\}$. As in the case of iid random variables we have X_i and W_i independent but we do not in general have that W_i and $W - W_i$ are independent.

Take any f such that

$$|f| \leq 1, \quad |f'| \leq \sqrt{\frac{2}{\pi}}, |f''| \leq 2.$$

Then

$$EWf(W) = \frac{1}{\sigma}\sum E(X_i f(W)) = \frac{1}{\sigma}\sum E(X_i(f(W) - f(W_i))) = (I) + (II)$$

where

$$(I) = \frac{1}{\sigma}\sum E[X_i(f(W) - f(W_i) - (W - W_i)f'(W))]$$

and

$$(II) = \frac{1}{\sigma}\sum E[X_i(W - W_i)f'(W)].$$

Now

$$(I) \leq \frac{1}{\sigma}\sum \frac{1}{2}E|X_i(W - W_i)^2||f''|_\infty \leq \frac{1}{\sigma^3}\sum E|X_i(\sum_{j \in N_i} X_j)^2|$$

since $W - W_i = \frac{1}{\sigma}\sum_{j\in N_i} X_j$. Also

$$(II) = \frac{1}{\sigma}\sum EX_i(\sum_{j\in N_i} X_j f(W)) = E\Bigg(f'(W)\underbrace{\Bigg[\frac{1}{\sigma^2}\sum X_i(\sum_{j\in N_i} X_j)\Bigg]}_{T}\Bigg).$$

We will proceed by showing that T is concentrated. Note that since $EX_iW_i = 0$,

$$\frac{1}{\sigma}E\sum X_i(W - W_i) = \frac{1}{\sigma}E\sum X_iW = EW^2 = 1$$

and so

$$|(II) - f'(W)| = |E(f'(W)(T-1)| \le |f'|_\infty E|T-1| \le \sqrt{\frac{2}{\pi}}\sqrt{E(T-1)^2} = \sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}(T)}$$

Combining these results we have

$$|EWf(W) - Ef'(W)| \le \sqrt{\frac{2}{\pi}}\sqrt{\underbrace{\mathrm{Var}(\frac{1}{\sigma^2}\sum X_i(\sum_{j\in N_i} X_j))}_{III} + \underbrace{\frac{1}{\sigma^3}\sum E|X_i(\sum_{j\in N_i} X_j)^2|}_{IV}}.$$

Now

$$\begin{aligned} IV &\le \frac{1}{\sigma^3}\sum_i \sum_{j,k\in N_i} E|X_iX_jX_k| \\ &\le \frac{1}{\sigma^3}\sum_i \sum_{j,k\in N_i} \frac{1}{3}(E|X_i|^3 + E|X_i|^3 + E|X_i|^3) \\ &\le \frac{D^2}{\sigma^3}\sum E|X_i|^3 \end{aligned}$$

where the second inequality follows from the AM-GM inequality.

Next we need to estimate

$$\mathrm{Var}(\sum_{i,j\in N_i} X_iX_j).$$

The collection $\{X_iX_j, i \in V, j \in N_i\}$ is a collection with a dependency graph of maximum degree $2D^2$. This can be seen as follows: X_iX_j is independent of X_kX_l if neither k nor l belongs to $N_i \cup N_j$. Now $|N_i \cup N_j| \le 2D$ and each vertex in this set has at most D neighbors so the maximum degree of the new dependency graph is $2D^2$. Using the variance bound on sums of dependency graph variables derived in the previous lecture we have

$$\mathrm{Var}(\sum_{i,j\in N_i} X_iX_j) \le 2D^2\sum_{i\sim j}\mathrm{Var}(X_iX_j) \le 2D^3\sum EX_i^4$$

using the fact that

$$\mathrm{Var}(X_iX_j) \le E(X_i^2X_j^2) \le \frac{1}{2}EX_i^4 + \frac{1}{2}EX_j^4.$$

The proof is completed by substituting this estimate. □

Example 17 *Let $Y_1, \ldots, Y_{n+1}$ be iid mean 0 variance 1 random variables and let $X_i = Y_iY_{i+1}$ A dependency graph for the X_i has edge set $\{(i, i+1) : 1 \le i \le n\}$ and $D = 3$. Then $\mathrm{Var}(\sum X_i) = \sigma^2 = Cn$ so*

$$\mathrm{Wass}\left(\frac{1}{\sigma}\sum X_i,\, Z\right) \le C\frac{1}{\sigma^2}\sqrt{D^3\sum EX_i^4} + \frac{D^2}{\sigma^3}\sum E|X_i|^3 \le \frac{c}{\sqrt{n}}$$

Exercise 18 *In an Erdos-Renyi random graph $G(n,p)$ let $T_{n,p}$ be the number of triangles. Using the method of dependency graphs show that for some absolute constant C*

$$\mathrm{Wass}\left(\frac{T_{n,p} - ET_{n,p}}{\sqrt{\mathrm{Var}(T_{n,p})}},\, Z\right) \le \frac{C}{np^{9/2}}$$

Exercise 19

- *Find out the best known result for the above problem.*
- *Show that the CLT can not hold if $np \not\to \infty$.*
- *Refine the method of dependency graphs to show a CLT when $np \to \infty$.*

STAT C206A / MATH C223A : Stein's method and applications	Fall 2007
Lecture 7	
Lecture date: Sept. 12, 2007	*Scribe: Partha Dey*

9 Method of Exchangeable Pair

Suppose (W, W') is an exchangeable pair of random variables, *i.e.* $(W, W') \stackrel{d}{=} (W', W)$, and there is a constant $\lambda \in (0, 1)$ such that

$$\mathbf{E}(W' - W|W) = -\lambda W \text{ a.e.} \tag{14}$$

Also suppose $\mathbf{E}W^2 = 1$, then we have

$$\mathbf{Wass}(W, Z) \leq \sqrt{\frac{2}{\pi}\text{Var}\left(\mathbf{E}(\frac{1}{2\lambda}(W' - W)^2|W)\right)} + \frac{1}{3\lambda}\mathbf{E}|W' - W|^3 \tag{15}$$

where $Z \sim \text{N}(0, 1)$.

In general λ is a small positive number, usually of the oder $1/n$ and W' is obtained by applying a small perturbation to W.

Note that from the given conditions we have $\mathbf{E}W = \mathbf{E}W'$ and $\mathbf{E}W^2 = \mathbf{E}W'^2 = 1$. Using this information along with (14) we have

1. $\mathbf{E}W = 0$. Since $\mathbf{E}[-\lambda W] = \mathbf{E}[\mathbf{E}(W' - W|W)] = \mathbf{E}[W' - W] = 0$ and $\lambda \neq 0$.
2. $\mathbf{E}(W' - W)^2 = 2\lambda$. Since

$$\begin{aligned}\mathbf{E}(W' - W)^2 &= \mathbf{E}[W'^2 + W^2 - 2W'W] \\ &= \mathbf{E}[2W^2 - 2W'W] \\ &= \mathbf{E}[2W(W - W')] = \mathbf{E}[2W\mathbf{E}(W - W'|W)] = \mathbf{E}[2\lambda W^2] = 2\lambda.\end{aligned}$$

Now take any twice differentiable function f with $\|f\|_\infty \leq 1, \|f'\|_\infty \leq \sqrt{\frac{2}{\pi}}$ and $\|f''\|_\infty \leq 2$. Let

$$F(x) = \int_0^x f(y)dy.$$

Clearly F is a well defined thrice differentiable function. So using Taylor series expansion for F we have

$$\begin{aligned}0 &= \mathbf{E}[F(W') - F(W)] \\ &= \mathbf{E}\left[(W' - W)f(W) + \frac{1}{2}(W' - W)^2 f'(W) + \text{Remainder}\right]\end{aligned} \tag{16}$$

where $|\text{Remainder}| \le \frac{1}{6}|W-W'|^3\|f''\|_\infty \le \frac{1}{3}|W-W'|^3$. Now

$$\begin{aligned}
-\lambda \mathbf{E}[Wf(W)] &= \mathbf{E}[(W'-W)f(W)] \\
&= -\mathbf{E}\left[\frac{1}{2}(W'-W)^2 f'(W) + \text{Remainder}\right] \\
&= -\mathbf{E}\left[\frac{1}{2}\mathbf{E}[(W'-W)^2|W]\ f'(W)\right] + \mathbf{E}[\text{Remainder}].
\end{aligned}$$

Dividing both sides by λ we get

$$|\mathbf{E}f'(W) - \mathbf{E}Wf(W)| \le \left|\mathbf{E}\left[f'(W)\cdot\left(\mathbf{E}(\frac{1}{2\lambda}(W-W')^2|W) - 1\right)\right]\right| + \frac{1}{3\lambda}\mathbf{E}|W-W'|^3.$$

Since $\|f'\|_\infty \le \sqrt{\frac{2}{\pi}}$ and $\mathbf{E}(W-W')^2 = 2\lambda$ we have

$$\begin{aligned}
|\mathbf{E}f'(W) - \mathbf{E}Wf(W)| &\le \sqrt{\frac{2}{\pi}}\cdot\mathbf{E}\left|\mathbf{E}(\frac{1}{2\lambda}(W-W')^2|W) - 1\right| + \frac{1}{3\lambda}\mathbf{E}|W-W'|^3 \\
&\le \sqrt{\frac{2}{\pi}\text{Var}\left(\mathbf{E}(\frac{1}{2\lambda}(W'-W)^2|W)\right)} + \frac{1}{3\lambda}\mathbf{E}|W-W'|^3.
\end{aligned}$$

Remark 20 *If W, W' just have the same distribution (need not be exchangeable) then also the above result holds.*

Now let us apply this method to the simplest case of sums of independent random variables.

Let $X_1, X_2, \ldots, X_n$ be independent random variables with mean 0 and variance 1. Define

$$W = \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i.$$

Let $X_1', X_2', \ldots, X_n'$ be an independent copy of $X_1, X_2, \ldots, X_n$. Choose an index I uniformly at random from $\{1, 2, \ldots, n\}$. Replace X_I by X_I'. Let

$$W' = \frac{1}{\sqrt{n}}\sum_{j\ne I} X_j + \frac{X_I'}{\sqrt{n}} = W + \frac{X_I' - X_I}{\sqrt{n}}.$$

Lemma 21 *(W, W') is an exchangeable pair.*

Proof: Exercise. □

Note that $W' - W = \frac{X_I' - X_I}{\sqrt{n}}$. Hence we have

$$\mathbf{E}[W' - W|W] = \frac{1}{\sqrt{n}}\mathbf{E}[X_I' - X_I|W]$$
$$= \frac{1}{\sqrt{n}} \cdot \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}[X_i' - X_i|W] = -\frac{1}{n}\mathbf{E}[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i|W] = -\frac{1}{n}W.$$

Here condition (15) is satisfied with $\lambda = n^{-1}$. Now,

$$\frac{1}{3\lambda}\mathbf{E}|W' - W|^3 = \frac{n}{3n^{3/2}}\mathbf{E}|X_I' - X_I|^3 = \frac{1}{3n^{3/2}}\sum_{i=1}^{n}\mathbf{E}|X_i' - X_i|^3 \leq \frac{8}{3n^{3/2}}\sum_{i=1}^{n}\mathbf{E}|X_i|^3$$

and

$$\mathbf{E}\left[\frac{1}{2\lambda}(W' - W)^2|W\right] = \frac{n}{2n}\mathbf{E}((X_I' - X_I)^2|W) = \frac{1}{2n}\sum_{i=1}^{n}\mathbf{E}((X_i' - X_i)^2|W).$$

Note that,

$$\mathbf{E}((X_i' - X_i)^2|W) = \mathbf{E}(X_i'^2 - 2X_i'X_i + X_i^2|W) = 1 + \mathbf{E}(X_i^2|W).$$

Hence

$$\mathrm{Var}\left(\mathbf{E}(\frac{1}{2\lambda}(W' - W)^2|W)\right) = \mathrm{Var}\left(\mathbf{E}(\frac{1}{2n}\sum_{i=1}^{n} X_i^2|W)\right)$$
$$\leq \mathrm{Var}\left(\frac{1}{2n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{4n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i^2) \leq \frac{1}{4n^2}\sum_{i=1}^{n}\mathbf{E}X_i^4$$

and we have,

$$\mathbf{Wass}(W, Z) \leq \sqrt{\frac{1}{2\pi n^2}\sum_{i=1}^{n}\mathbf{E}X_i^4} + \frac{8}{3n^{3/2}}\sum_{i=1}^{n}\mathbf{E}|X_i|^3.$$

Lecture 8

Lecture date: Sept. 14, 2007 *Scribe: Laura Derksen*

10 Exchangeable pairs

Recall that a pair of random variables is called exchangeable if (W, W') and (W', W) are equal in distribution.

During the last lecture we obtained an upper bound on the Wasserstein distance between such a W and a Gaussian random variable Z:

Theorem 22 *Let Z be a standard Gaussian random variable. If (W, W') is an exchangeable pair of r.v.'s, $\mathbf{E}(W' - W|W) = -\lambda W$ for some $0 < \lambda < 1$, and $\mathbf{E}(W^2) = 1$ (or $\mathbf{E}((W' - W)^2) = 2\lambda$), then*

$$Wass(W, Z) \leq \sqrt{\frac{2}{\pi}\mathrm{Var}\left(\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|W\right)\right)} + \frac{1}{3\lambda}\mathbf{E}(|W' - W|^3). \tag{17}$$

Intuitively, if $\mathbf{E}(W' - W|W) = -\lambda W$, $\mathbf{E}((W' - W)^2) = 2\lambda + o(\lambda)$, and $\mathbf{E}(|W' - W|^3) = o(\lambda)$ then $\mathrm{Wass}(W, Z) = o(1)$.

Usually the quantity $\frac{1}{2\lambda}(W' - W)^2$ is not concentrated. However, we will often have a σ-algebra $\mathcal{F}$ such that W is measurable with respect to $\mathcal{F}$ and

$$\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|\mathcal{F}\right)$$

is concentrated. By Jensen's inequality,

$$\mathrm{Var}\left(\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|W\right)\right) \leq \mathrm{Var}\left(\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|\mathcal{F}\right)\right).$$

11 Example: CLT for the scaled sum of i.i.d. random variables

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with mean 0 and variance 1. Let

$$W = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i.$$

As seen last lecture, we define W' as follows:

$$W' = \frac{1}{\sqrt{n}} \sum_{j \neq I} X_j + \frac{X_I}{\sqrt{n}},$$

where the index I is chosen uniformly at random from $\{1, 2, ..., n\}$ and X_I is independent from, and equal in distribution to the other X_i's.

Then $\mathbf{E}(W' - W|W) = -\frac{1}{n}W$ so $\lambda = \frac{1}{n}$. $\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|W\right)$ is hard to compute. However, we can write $\frac{1}{2\lambda}(W' - W)^2 = \frac{1}{2}(X_I' - X_I)^2$, and if $\mathcal{F}$ is $\sigma(X_1, X_2, ..., X_n)$ then

$$\mathbf{E}\left(\frac{1}{2\lambda}(W' - W)^2|\mathcal{F}\right) = \frac{1}{2} + \frac{1}{2n}\sum_{i=1}^{n} X_i^2,$$

which is concentrated.

12 Hoeffding combinatorial central limit theorem

Suppose $(a_{ij})_{i,j=1}^n$ is an array of numbers. Let π be a uniform random permutation of $\{1, 2, ..., n\}$. Let $W = \sum_{i=1}^n a_{i\pi(i)}$.

We would like to say something about how close $\frac{W - \mathbf{E}(W)}{\sqrt{\mathrm{Var}(W)}}$ is to the standard Gaussian distribution $N(0, 1)$.

Hoeffding's original proof involved a sequence of matrices $(a_{ij}^{(n)})_{i,j=1}^n$ and gave conditions for convergence to normality. The method of moments was used for the proof. The idea is to show that

$$\mathbf{E}\left(\left(\frac{W_n - \mathbf{E}(W_n)}{\sqrt{\mathrm{Var}(W_n)}}\right)^k\right)$$

converges to 0 for k odd, and to $\frac{(2k)!}{2^k k!}$ for k even.

Bolthausen ('83 or '84) proved a Berry-Esseen bound for finite n using Stein's method.

We assume the following, without loss of generality:

$$\sum_{j=1}^{n} a_{ij} = 0, \ \sum_{i=1}^{n} a_{ij} = 0 \text{ and } \frac{1}{n-1}\sum_{i,j=1}^{n} a_{ij}^2 = 1. \tag{18}$$

To see why this does not compromise generality, for an arbitrary $(a_i j)_{i,j=1}^n$ we define

$$a_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n} a_{ij},$$

$$a_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij},$$

$$a_{\cdot\cdot} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{ij},$$

and

$$\tilde{a}_{ij} = a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot}.$$

Now,

$$\begin{aligned}
\sum_{i=1}^{n} \tilde{a}_{ij} &= \sum_{i=1}^{n} a_{ij} - \sum_{i=1}^{n} a_{i\cdot} - \sum_{i=1}^{n} a_{\cdot j} + \sum_{i=1}^{n} a_{\cdot\cdot} \\
&= \sum_{i=1}^{n} a_{ij} - \frac{1}{n}\sum_{i,j=1}^{n} a_{ij} - \sum_{i=1}^{n} a_{ij} + \frac{1}{n}\sum_{i,j=1}^{n} a_{ij} \\
&= 0.
\end{aligned}$$

Similarly, we can check that the other assumptions in (18) are satisfied by $(\tilde{a}_{ij})$.

We define

$$\tilde{W} = \sum_{i=1}^{n} \tilde{a}_{i\pi(i)} = \sum_{i=1}^{n} a_{i\pi(i)} - \sum_{i=1}^{n} a_{i\cdot} - \sum_{i=1}^{n} a_{\cdot\pi(i)} + na_{\cdot\cdot} = \sum_{i=1}^{n} a_{i\pi(i)} - na_{\cdot\cdot}.$$

It can easily be checked that

$$\frac{\tilde{W} - \mathbf{E}(\tilde{W})}{\sqrt{\mathrm{Var}(\tilde{W})}} = \frac{W - \mathbf{E}(W)}{\sqrt{\mathrm{Var}(W)}},$$

justifying (18).

We now return to our original problem, and assume (18). Then we have

$$\mathbf{E}(a_{i\pi(i)}) = \frac{1}{n}\sum_{j=1}^{n} a_{ij} = 0,$$

so $\mathbf{E}(W) = 0$. For the variance, we can write

$$\mathrm{Var}(W) = \sum_{i=1}^{n} \mathrm{Var}(a_{i\pi(i)}) + \sum_{i \neq j} \mathrm{Cov}(a_{i\pi(i)}, a_{j\pi(j)}).$$

First,

$$\mathrm{Var}(a_{i\pi(i)}) = \mathbf{E}(a_{i\pi(i)}^2) = \frac{1}{n}\sum_{j=1}^{n} a_{ij}^2,$$

so

$$\sum_{i=1}^{n} \mathrm{Var}(a_{i\pi(i)}) = \frac{1}{n} \sum_{i,j=1}^{n} a_{ij}^2.$$

Now we will calculate the covariance.

$$\begin{aligned}
\mathrm{Cov}(a_{i\pi(i)}, a_{j\pi(j)}) &= \mathbf{E}(a_{i\pi(i)} a_{j\pi(j)}) \\
&= \frac{1}{n-1} \sum_{k,l\neq k} a_{ik} a_{jl} \\
&= \frac{-1}{n(n-1)} \sum_{k} a_{ik} a_{jk}
\end{aligned}$$

where the last equality comes from the fact that $\sum_{l\neq k} a_{jl} = -a_{jk}$.

We now obtain

$$\begin{aligned}
\sum_{i\neq j} \mathrm{Cov}(a_{i\pi(i)}, a_{j\pi(j)}) &= \frac{-1}{n(n-1)} \sum_{i\neq j} \sum_{k} a_{ik} a_{jk} \\
&= \frac{1}{n(n-1)} \sum_{i,k} a_{ik}^2.
\end{aligned}$$

Combining the variance and covariance calculations above, and keeping (18) in mind, we obtain

$$\mathrm{Var}(W) = \frac{1}{n-1} \sum_{i,j=1}^{n} a_{ij}^2 = 1.$$

Next, we will create an exchangeable pair (π, π') by defining $\pi' = \pi \circ (I, J)$ and $W' = \sum_{i=1}^{n} a_{i\pi'(i)}$ where (I, J) is a uniformly random transposition.

To be continued in the next lecture.

Lecture 9

Lecture date: Sep 17, 2007 *Scribe: Tanya Gordeeva*

13 Proof of the Hoeffding combinatorial CLT

Recall the Hoeffding CLT from the previous lecture:

Theorem 23 *Suppose $(a_{ij})_{1\le i,j\le n}$ is an array of numbers. Let π be a uniform random permutation of $\{1,\dots,n\}$. Let $W=\sum_{i=1}^n a_{i\pi(i)}$. Under suitable conditions on $(a_{ij})_{1\le i,j\le n}$, W converges to the normal distribution (after centering and scaling).*

We assumed, without loss of generality, that

$$\forall i \sum_{j=1}^n a_{ij}=0, \quad \forall j \sum_{i=1}^n a_{ij}=0, \quad \frac{1}{n-1}\sum_{i,j} a_{ij}^2=1$$

so that

$$\mathbf{E}\, W=0, \quad \mathbf{E}\, W^2=1$$

We will create an exchangeable pair (π,π') by defining $\pi'=\pi\circ(I,J)$, where (I,J) is selected uniformly at random over the set of transpositions on $\{1,\dots,n\}$. That is, $\pi'(I)=\pi(J)$, $\pi'(J)=\pi(I)$, and $\pi'(k)=\pi(k)$ for $k\ne I,J$.

Exercise 1: Show that (π,π') is an exchangeable pair.

Let $W'=\sum_{i=1}^n a_{i\pi'(i)}$. So (W,W') is an exchangeable pair. Note that

$$\begin{aligned} W'-W &= a_{I\pi'(I)}+a_{J\pi'(J)}-a_{I\pi(I)}-a_{J\pi(J)} \\ &= a_{I\pi(J)}+a_{J\pi(I)}-a_{I\pi(I)}-a_{J\pi(J)} \end{aligned}$$

So, by summing over the choices for (I,J),

$$\mathbf{E}(W'-W|\pi)=\frac{1}{n(n-1)}\sum_{1\le i\ne j\le n}(a_{i\pi(j)}+a_{j\pi(i)}-a_{i\pi(i)}-a_{j\pi(j)})$$

Note that

$$\frac{1}{n(n-1)}\sum_{i\neq j} a_{i\pi(i)} = \frac{1}{n}W.$$

Again, by assumption, $\sum_{j\neq i} a_{i\pi(j)} = -a_{i\pi(i)}$ for fixed i, so

$$\frac{1}{n(n-1)}\sum_{i\neq j} a_{i\pi(j)} = \frac{-1}{n(n-1)}\sum_{i} a_{i\pi(i)} = \frac{-1}{n(n-1)}W.$$

A similar argument applies to show

$$\frac{1}{n(n-1)}\sum_{i\neq j} a_{j\pi(i)} = \frac{-1}{n(n-1)}W.$$

Combining, we get that

$$\mathbf{E}(W'-W|\pi) = \frac{-2(n-2)}{n(n-1)}W.$$

Since this depends only on W,

$$\mathbf{E}(W'-W|W) = \frac{-2(n-2)}{n(n-1)}W = -\lambda W$$

where $\lambda = \frac{2(n-2)}{n(n-1)}$.

Now consider

$$\mathbf{E}((W'-W)^2|\pi) = \frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n}(a_{i\pi(i)} + a_{j\pi(j)} - a_{i\pi(j)} - a_{j\pi(i)})^2$$

The conditional expectation given W has a smaller variance, so bounding this will be sufficient to apply the method of exchangeable pairs.

Exercise 2: Bound the variance of this conditional expectation to obtain the complete Hoeffding combinatorial CLT.

A paper of Bolthausen's (84, Z.W.) gives a bound on the Kolmogorov distance for the Hoeffding CLT.

Theorem 24 *Consider W as above. Let $Z\sim N(0,1)$. Then*

$$\sup_t |\mathbf{P}(W\leq t) - \mathbf{P}(Z\leq t)| \leq K\frac{\sum_{i,j}|a_{ij}|^3}{n}$$

where K is a universal constant.

What is the order of the bound? Informally, we can say $\frac{1}{\sqrt{n}}$. Typically, $\frac{n^2 O(a_{ij}^2)}{n} = 1$ (since $\frac{1}{n(n-1)} \sum_{i,j} a_{ij}^2 = 1$ and $\sum_i a_{ij} = 0 = \sum_j a_{ij}$), so $a_{ij} = O(1/\sqrt{n})$. This is assuming the values are evenly spaced over the a_{ij}. So $K \frac{\sum_{i,j} |a_{ij}|^3}{n} = O(n^2 n^{-3/2})/n = O(1/\sqrt{n})$.

The method of exchangeable pairs does not allow work to obtain Bolthausen's result. It is possible to get a bound on the Wasserstein distance, but it will involve 4th powers.

14 CLT for the antivoter model

Suppose $G = (V, E)$ is an r-regular graph with n vertices (for instance, the torus). $(X_i^{(t)})_{i \in V}$ is a process of ± 1 valued rv evolving as follows. At any time t, choose a vertex i uniformly at random. Then choose a neighbor j of i uniformly at random. Let $X_i^{(t+1)} = -X_j^{(t)}$ and $X_k^{(t+1)} = X_k^{(t)}, \forall k \neq i$. So you make the opposite decision of your neighbor.

This chain will converge to a stationary distribution (supposing aperiodicity and irreducibility), although the stationary distribution is not trivial to describe. Let $(X_i)_{i \in V}$ be a random variable distributed as the stationary distribution. The problem is to show that $\sum_{i \in V} X_i$ is approximately Gaussian (after centering and scaling) if n is large and r is fixed.

Rinott and Rotar (97, AAP) got a Berry-Esséen bound, using the method of exchangeable pairs.

Let $W = \sum_{i \in V} X_i$. Construct W' by taking a step in the chain. Rinott and Rotar proved that, although the chain is not usually reversible, (W, W') is still an exchangeable pair. (Even though we do not require exchangeability if we use the bound on the Wasserstein distance given in this set of notes, it may not be possible to do so for the Rinott-Rotar Berry-Esséen bounds.)

Clearly, $W' - W \in \{-2, 0, 2\}$. Let

$$\begin{aligned} a(X) &= |\{\text{edges } (i,j) \in E \text{ s.t. } X_i = X_j = 1\}| \\ b(X) &= |\{\text{edges } (i,j) \in E \text{ s.t. } X_i = X_j = -1\}| \\ c(X) &= |\{\text{edges } (i,j) \in E \text{ s.t. } X_i \neq X_j\}| \end{aligned}$$

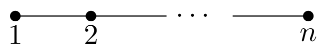
If $T(X) = |\{i \in V : X_i = 1\}| = \frac{1}{2}\left(\sum_{i \in V} X_i + n\right)$, then it can be seen that

$$T(X) = \frac{2a(X) + c(X)}{r} \Rightarrow n - T(X) = \frac{2b(X) + c(X)}{r}$$

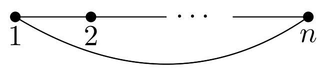
Then $\mathbf{P}(W' - W = 2 | X) = \frac{2a(X)}{rn}$ and $\mathbf{P}(W' - W = 2 | X) = \frac{2b(X)}{rn}$. So

$$\mathbf{E}(W' - W | X) = \frac{4b(X) - 4a(X)}{rn} = \frac{2(n - 2T(X))}{n} = \frac{-2}{n} W$$

Exercise 3: Consider the 1-dim Ising model on n vertices. The graph is either:



or



Let $\sigma = (\sigma_1, \ldots, \sigma_n)$, n spins (± 1 valued). The probability distribution on the spins is $\mathbf{P}(\sigma) = Z^{-1} \exp(\beta \sum_i \sigma_i \sigma_{i+1})$. Prove a CLT for $\sum_{i=1}^n \sigma_i$ using exchangeable pairs.

Lecture 10: Concentration Inequalities

Lecture date: September 19, 2007 *Scribe: Chris Haulk*

15 Concentration Inequalities

Suppose X is a random variable and m is a constant (usually the mean or median of X). We seek bounds like

$$P(X - m > t) \leq \exp(-f(t)), \qquad P(X - m < -t) \leq \exp(-g(t)).$$

Typically, $f(0) = 0$ and $\lim_{t\to\infty} f(t) = \infty$, and similarly for g. For example, if X_i are iid with $P(X_i = -1) = P(X_i = 1) = 0.5$, then

$$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i > t\right) \leq \exp\left(\frac{-nt^2}{2}\right),$$

and similarly for the lower bound, so

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > t\right) \leq 2\exp\left(\frac{-nt^2}{2}\right).$$

Obviously this provides some justification for the statement that $\frac{1}{n}\sum_{i=1}^{n} X_i$ is concentrated near 0.

Now the issue is to read off the "typical deviation" of X from m. The method is to find the range of t for which f is "like a constant," say, e.g., equal to 1. In the example above, $t = n^{-1/2}$ is the typical deviation. If $t \ll n^{-1/2}$ then $nt^2/2$ is near 0, so the bound provides no information. If $t \gg n^{-1/2}$, then $nt^2/2$ is quite large, so the upper bound on $P(X-m > t)$ is near 0, and we are left wondering whether there is a smaller neighborhood of m around which X concentrates.

Caution: Don't assume that upper bounds are sharp.

The simplest concentration inequalities come from variance bounds: the typical deviation of X from EX is $\sqrt{Var(X)}$. The following is a useful bound on the variance of a function of several random variables.

Theorem 25 *Efron Stein-Inequality (or Influence Inequality, or MG bound on Variance). Suppose that $X_1, ..., X_n$, $X_1', ..., X_n'$ are independent with $X_i' \overset{d}{=} X_i$ for all i. Let $X =$*

$(X_1, ..., X_n)$, $X^{(i)} = (X_1, ..., X_{i-1}, X_i', X_{i+1}, ...X_n)$. *Then*

$$Var(f(X)) \leq \frac{1}{2} \sum_{i=1}^{n} E\left[(f(X) - f(X^{(i)})^2 \right].$$

Proof:

Let $X' = (X_1', ..., X_n')$, $X^{[i]} = (X_1', ..., X_i', X_{i+1}, ..., X_n)$; note $X^{[0]} = X$ and $X^{[n]} = X'$.

$$\begin{aligned} Var(f(X)) &= Ef(X)^2 - (Ef(X))^2 \\ &= Ef(X)^2 - E[f(X)f(X')] \\ &= E[f(X)(f(X) - f(X'))] \\ &= \sum_{i=1}^{n} E\left[f(X)\left(f(X^{[i-1]}) - f(X^{[i]}) \right) \right] \end{aligned}$$

Fix i and note that $f(X)\left(f(X^{[i-1]}) - f(X^{[i]})\right)$ is a function of $(X_1, ...X_n, X_1', ..., X_n') =: X^*$. The distribution of X^* remains unchanged if we switch X_i and X_i'. Under this switching operation,

$$f(X)\left(f(X^{[i-1]}) - f(X^{[i]}) \right) \mapsto f(X^{(i)})\left(f(X^{[i]}) - f(X^{[i-1]}) \right),$$

so these two quantities are equal in law. It follows that

$$a = E\left[f(X)\left(f(X^{[i-1]}) - f(X^{[i]}) \right) \right] = E\left[f(X^{(i)})\left(f(X^{[i]}) - f(X^{[i-1]}) \right) \right] = b.$$

Observing that $a = b$ implies $a = b = (a+b)/2$, we obtain by Cauchy-Schwarz

$$\begin{aligned} E\left[f(X)\left(f(X^{[i-1]}) - f(X^{[i]}) \right) \right] &= \frac{1}{2} E\left[\left(f(X) - f(X^{(i)}) \right)\left(f(X^{[i-1]}) - f(X^{[i]}) \right) \right] \\ &\leq \frac{1}{2}\left(E\left[(f(X) - f(X^{(i)}))^2 \right] E\left[(f(X^{[i]}) - f(X^{[i-1]}))^2 \right] \right)^{1/2} \\ &= \frac{1}{2} E\left[\left(f(X) - f(X^{(i)}) \right)^2 \right], \end{aligned}$$

where the second step follows by noticing that

$$E[(f(X) - f(X^{(i)})^2] = E[(f(X^{[i-1]}) - f(X^{[i]}))^2]$$

(by $i-1$ applications of the switching operation). Sum over i to complete the proof.

16 Application: First Passage Percolation

We will apply the Efron-Stein inequality to study first-passage percolation. This section gives definitions; the application will be finished next class.

Consider the lattice $\mathbb{Z}^2$, with iid nonnegative random edge weights $(w_e)_{e\in E}$, E being the set of edges of the lattice. Let $t_n = \min\{\sum_{e\in p} w_e : p \text{ is a path from } (0,0) \text{ to } (n,0)\}$. In words, t_n it the first time that the vertex $(n,0)$ is reached by a liquid that is spilled onto the origin and which takes a random amount of time to flow over each edge in the graph.

Theorem[2]: $t_n/n \to \mu$ in probability where μ depends on the distribution of edge weights.

Note that the weight of a minimal path from $(0,0)$ to $(n+m,0)$ is less than or equal to the sum of the weights of the minimal paths from $(0,0)$ to $(0,n)$ and from $(0,n)$ to $(0,n+m)$. Therefore

$$E(t_{n+m}) \leq E(t_n) + E(t_m).$$

Next time we'll begin with the subadditive lemma: if $\{a_n\}$ is a sequence of real random numbers satisfying $a_{n+m} \leq a_n + a_m$ (i.e., $\{a_n\}$ is subadditive), then $\lim_n a_n/n$ exists in $[-\infty, \infty)$ and equals $\inf_n a_n/n$.

[2]Kesten, 1993 Annals of Applied Probability. *On the speed of first-passage percolation.*

Lecture 11

Lecture date: Sep 21, 2007 *Scribe: Maximilian Kasy*

For E the set of edges in the lattice $\mathbb{Z}^2$, let $(\omega_e)_{e\in E}$ be i.i.d. nonnegative edge weights. Define

$$t_n := \inf\{\sum_{e\in P}\omega_e : P \text{ is a path from}(0,0)\text{to}(n,0)\}$$

Then $E(t_{n+m}) \leq E(t_n) + E(t_m)\ \forall\ n, m$, as shown last time.

Lemma 26 (Subadditive Lemma) *If $\{a_n\}$ is a sequence of real nombers, such that $a_{n+m} \leq a_n + a_m\ \forall\ n, m$, then*

$$\lim_{n\to\infty}\frac{a_n}{n} = \inf_{n\geq 1}\frac{a_n}{n}$$

In particular the limit exists.

Proof: Fix $k \geq 1$. Take any n and choose m such that $n = mk + r$ with $0 \leq r < k$. Then $a_n \leq ma_k + a_r$. It follows that $\limsup \frac{a_n}{n} \leq \frac{a_k}{k}$, hence

$$\limsup\frac{a_n}{n} \leq \inf\frac{a_k}{k} \leq \liminf\frac{a_n}{n}$$

which concludes the proof.

Now suppose for the rest of this lecture that $E(\omega_e^2) < \infty$. It follows from the lemma that $\exists 0 \leq \mu < \infty$ such that $E(t_n)/n \to \mu$ as $n \to \infty$.
Kersten showed: If $E(\omega_e^2) < \infty$ and $P(\omega_e = 0) < p_c(d)$, where $p_c(d)$ is the critical probability for bond percolation in $\mathbb{Z}^d$, then $\mu > 0$.
We want to show: $\frac{t_n}{n} \to \mu$ in probability. Assume now

$$\exists\ a > 0 s.t. P(\omega_e > a) = 1 \tag{19}$$

We will show an inequality of the form $Var(t_n) \leq Cn$. Let l_n be the nomber of edges in a shortest minimal-weight path. Under assumption 19, $l_n \leq \frac{t_n}{a}$. By the argument of last lecture,

$$Var(t_n) \leq \frac{1}{2}\sum_{e\in E} E(t_n(\omega) - t_n(\omega^{(e)}))^2$$

where $\omega_u^{(e)} = \omega_u$ if $u \neq e$ and $\omega_e^{(e)} = \omega_e'$ where the latter is an independent copy of ω_e. By symmetry, then,

$$Var(t_n) \leq \sum_{e \in E} E\left[(t_n(\omega) - t_n(\omega^{(e)}))^2 1(\omega_e \leq \omega_e')\right] =: \otimes \text{ (say)}.$$

Now, if $\omega_e \leq \omega_e'$ and $t_n(\omega) \neq t_n(\omega^{(e)})$, then it is easy to argue that e must be in every minimal path from $(0,0)$ to $(n,0)$ under the configuration ω, and $(t_n(\omega) - t_n(\omega^{(e)})^2 \leq (\omega_e' - \omega_e)^2 \leq (\omega_e')^2$. Indeed, if the length of all minimal paths increase after increasing ω_e to ω_e', then e must belong to all minimal paths in ω, and we necessarily have $t_n(\omega^{(e)}) \leq t_n(\omega) + \omega_e' - \omega_e$. Hence

$$\begin{aligned}
\otimes &\leq \sum_e E((\omega_e')^2 1(e \in \text{every minimal path in } \omega)) \\
&= \sum_e E((\omega_e')^2) E(1(e \in \text{every minimal path in } \omega)) \\
&= E(\omega_e^2) E(\sum_e 1(e \in \text{every minimal path in } \omega)) \\
&\leq E((\omega_e)^2) E(l_n) \leq Cn
\end{aligned}$$

by assumption 19 and the fact that $E(t_n)/n \to \mu$. This concludes the argument.

It has been conjectured that actually $C_1 n^{2/3} \leq Var(t_n) \leq C_2 n^{2/3}$ in $\mathbb{Z}^2$, where C_1 and C_2 are positive constants depending on the distribution of ω_e. The best known lower bound in $C \log n$.

Let us now state two standard tools for proving concentration inequalities.

Theorem 27 (Azuma-Hoeffding inequality) *Suppose $X_1, \ldots, X_n$ are martingale differences with respect to a filtration $\{\mathscr{F}_1, \ldots, \mathscr{F}_n\}$ and there exist constants a_i, b_i such that $a_i \leq X_i \leq b_i$ almost surely. Then*

$$P\left(\max_{1 \leq k \leq n} \sum_{i=1}^k X_i \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum (b_i - a_i)^2}\right)$$

Theorem 28 (Bounded differences inequality) *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is such that there are constants $c_1, \ldots, c_n$ such that $|f(x) - f((y)| \leq c_i$ whenever x and y differ only in the ith coordinate. Suppose $X_1, \ldots, X_n$ are independent random variables. Then*

$$P(f(X_1, \ldots, X_n) - E(f(X_1, \ldots, X_n)) \geq t) \leq \exp\left(-\frac{2t^2}{\sum c_i^2}\right)$$

Sketch of a proof: Put $Y_i = E(f(X)|X_1, \ldots, X_i) - E(f(X)|X_1, \ldots, X_{i-1})$ and apply the previous theorem.

Lecture 12

Lecture date: September 24, 2007 *Scribe: Richard Liang*

17 Stein's method for concentration inequalities

The purpose of this lecture will be to prove the following theorem.

Theorem 29 *Suppose you have an exchangeable pair (X, X') of random objects. Suppose f and F are two functions such that*

(1) $F(X, X') = -F(X', X)$ *a.s.; and*

(2) $\mathbf{E}[f(X, X') \,|\, X] = f(X)$ *a.s.*

Let $v(X) = \frac{1}{2}\mathbf{E}\big[|(f(X) - f(X'))F(X, X')| \,\big|\, X\big]$. *Then*

(a) $\mathbf{E}[f(X)] = 0$ *and*

$$\begin{aligned} \mathrm{Var} f(X) &= \frac{1}{2}\mathbf{E}[(f(X) - f(X'))F(X, X')] \\ &\leq \mathbf{E}[v(X)]. \end{aligned} \tag{20}$$

(b) Suppose $\mathbf{E}(e^{\theta f(X)}|F(X, X')|)$ *is finite for all* θ. *If* B *and* C *are constants such that* $v(X) \leq Bf(X) + C$ *a.s., then*

$$\mathbf{P}\{|f(X)| > t\} \leq 2\exp\left(-\frac{t^2}{2Bt + 2C}\right). \tag{21}$$

(c) $\mathbf{E}\big[f(X)^{2k}\big] \leq (2k - 1)^k \mathbf{E}\big[v(X)^k\big]$ *for all* $k \in \mathbb{N}$.

Exercise 30 *(You may be able to do this after the proof.)*

Extend (c) to all real $k > 1/2$. *(We think that* $(2k - 1)^k$ *remains unchanged for* $k \geq 1$ *but are not sure for* $1/2 < k < 1$.*)*

Proof: First, note that $\mathbf{E}[f(X)] = \mathbf{E}[F(X, X')] = 0$ since (X, X') is an exchangeable pair and F is antisymmetric.

Further, we will assume

$$\begin{aligned} \mathbf{E}[v(X)] &< \infty \text{ for part (a); and} \qquad (22) \\ \mathbf{E}\Big[e^{\theta f(X)}|F(X, X')|\Big] &< \infty \text{ for all } \theta \text{ in part (b).} \qquad (23) \end{aligned}$$

We start by showing (a).

$$\begin{aligned} \mathrm{Var} f(X) &= \mathbf{E}\big[f(X)^2\big] \\ &= \mathbf{E}\big[f(X)F(X, X')\big] \\ &= \mathbf{E}\big[f(X')F(X', X)\big] \text{ since } (X, X') \text{ exchangeable} \\ &= -\mathbf{E}\big[f(X')F(X, X')\big] \text{ by antisymmetry of } F \\ &= \frac{1}{2}\mathbf{E}\big[\big(f(X) - f(X')\big)\, F(X, X')\big] \\ &\leq \mathbf{E}[v(X)], \end{aligned}$$

which proves (a).

(b): Let $m(\theta) = \mathbf{E}\big[e^{\theta f(X)}\big]$; then $m'(\theta) = \mathbf{E}\big[f(X)e^{\theta f(X)}\big]$. Write this as

$$\begin{aligned} m'(\theta) &= \mathbf{E}\Big[F(X, X')e^{\theta f(X)}\Big] \\ &= \frac{1}{2}\mathbf{E}\Big[F(X, X')\left(e^{\theta f(X)} - e^{\theta f(X')}\right)\Big] \end{aligned}$$

via the antisymmetry of F and the exchangeability of (X, X').

We'll use the inequality

$$|e^x - e^y| \leq \frac{1}{2}|x - y|(e^x + e^y). \qquad (24)$$

To see this, suppose $y < x$;

$$\begin{aligned} e^x - e^y &= \int_0^1 \frac{d}{dt}\left(e^{tx+(1-t)y}\right) dt \\ &= (x - y)\int_0^1 e^{tx+(1-t)y}\, dt \\ &\leq (x - y)\int_0^1 (te^x + (1 - t)e^y)\, dt \text{ by Jensen's inequality} \\ &= (x - y)\frac{1}{2}\left(e^x + e^y\right), \end{aligned}$$

and similarly for $y \geq x$.

Now

$$\begin{aligned}
|m'(\theta)| &= \Big|\mathbf{E}\Big[F(X,X')e^{\theta f(X)}\Big]\Big| \\
&\le \frac{1}{2}\Big|\mathbf{E}\Big[F(X,X')\left(e^{\theta f(X)} - e^{\theta f(X')}\right)\Big]\Big| \\
&\le \frac{|\theta|}{4}\mathbf{E}\Big[\big|F(X,X')\left(f(X) - f(X')\right)\big|\left(e^{\theta f(X)} + e^{\theta f(X')}\right)\Big] \\
&\le \frac{|\theta|}{2}\mathbf{E}\Big[\big|F(X,X')\left(f(X) - f(X')\right)\big|\, e^{\theta f(X)}\Big] \text{ by exchangeability of } (X,X') \\
&= |\theta|\mathbf{E}\Big[v(X)e^{\theta f(X)}\Big].
\end{aligned}$$

If $v(X) \le Bf(X) + C$, then the above gives

$$\begin{aligned}
|m'(\theta)| &\le |\theta|\left(B\mathbf{E}\Big[f(X)e^{\theta f(X)}\Big] + C\mathbf{E}\Big[e^{\theta f(X)}\Big]\right) \\
&= B|\theta|m'(\theta) + C|\theta|m(\theta).
\end{aligned}$$

Now, m is a convex function, with $\mathbf{E}[f(X)] = m'(0) = 0$ and taking the value $m(0) = 1$ at its minimum. Suppose $0 < \theta < 1/B$; then

$$m'(\theta)(1 - B\theta) \le C\theta m(\theta).$$

Thus

$$\frac{d}{d\theta}\log m(\theta) = \frac{m'(\theta)}{m(\theta)} \le \frac{C\theta}{1 - B\theta}$$

for $0 < \theta < 1/B$. Therefore,

$$\begin{aligned}
\log m(\theta) &= \int_0^\theta \frac{d}{dt}\log m(t)\, dt \\
&\le \int_0^\theta \frac{Ct}{1 - Bt}\, dt \\
&\le \frac{1}{1 - B\theta}\int_0^\theta Ct\, dt \\
&= \frac{C\theta^2}{2(1 - B\theta)}.
\end{aligned}$$

So, for $\theta > 0$,

$$\begin{aligned}
\mathbf{P}\{f(X) \ge t\} &= \mathbf{P}\Big\{e^{\theta f(X)} \ge e^{\theta t}\Big\} \\
&\le e^{-\theta t}m(\theta) \\
&\le \exp\left(-\theta t + \frac{C\theta^2}{2(1 - B\theta)}\right) \text{ if } 0 < \theta < 1/B.
\end{aligned}$$

Taking

$$\theta = \frac{t}{C + Bt} \in \left(0, \frac{1}{B}\right),$$

we get the desired upper bound, and $\mathbf{P}(f(X) \le -t)$ can be bounded similarly.

(c): Using a similar manipulation to that in the proof of (a),

$$\begin{aligned} \mathbf{E}\Big[f(X)^{2k}\Big] &= \mathbf{E}\Big[f(X)^{2k-1}F(X, X')\Big] \\ &= \frac{1}{2}\mathbf{E}\Big[\Big(f(X)^{2k-1} - f(X')^{2k-1}\Big)\, F(X, X')\Big]. \end{aligned} \tag{25}$$

Also, similarly to (24), we can show

$$\Big|x^{2k-1} - y^{2k-1}\Big| \le \frac{2k-1}{2}|x - y|\,\Big|x^{2k-2} + y^{2k-2}\Big|.$$

Plugging this into (25), we get

$$\mathbf{E}\Big[f(X)^{2k}\Big] \le (2k-1)\mathbf{E}\Big[v(X)f(X)^{2k-2}\Big].$$

Applying Hölder's inequality with $1/p = 1 - 1/k$ and $1/q = 1/k$ gives

$$\mathbf{E}\Big[f(X)^{2k}\Big] \le (2k-1)\left(\mathbf{E}\Big[f(X)^{2k}\Big]\right)^{(k-1)/k}\left(\mathbf{E}\Big[v(X)^k\Big]\right)^{1/k}$$

and so

$$\left(\mathbf{E}\Big[f(X)^{2k}\Big]\right)^{1/k} \le (2k-1)\left(\mathbf{E}\Big[v(X)^k\Big]\right)^{1/k}.$$

This completes the proof of (c). □

Example 31 *Suppose we have the Curie-Weiss model on n spins: for $\sigma = (\sigma_1, \ldots, \sigma_n)$,*

$$\mathbf{P}\{\sigma\} = \frac{1}{Z_\beta}\exp\left(\frac{\beta}{n}\sum_{1\le i<j\le n}\sigma_i\sigma_j\right).$$

Let

$$m(\sigma) = \frac{1}{n}\sum_i \sigma_i.$$

Construct X' by taking one step in the Gibbs sampler (also known as the Glauber dynamics). Set $F(X, X') = \sigma_I - \sigma'_I$ where I is the updated index. Then

$$\begin{aligned} f(X) &= \mathbf{E}\big[F(X, X') \,|\, X\big] \\ &\approx m(\sigma) - \tanh(\beta m(\sigma)). \end{aligned}$$

We'll do this in the next lecture.

Lecture 13

Lecture date: September 26, 2007 *Scribe: Joel Mefford*

Recall the following theorem from the previous lecture.

Theorem 32 *Suppose (X, X') is an exchangeable pair,*
$F(X, X')$ is an anti-symmetric function,
$f(X) = \mathbf{E}(F(X, X') \mid X)$,
and let

$$v(X) = \frac{1}{2}\mathbf{E}\big(|(f(X) - f(X'))F(X, X')|\big|X\big)$$

If $v(X) \leq Bf(X) + C$, for $B, C \geq 0$
then,

$$\mathbf{P}(|f(X)| \geq t) \leq 2\exp\left(-\frac{t^2}{2Bt + 2C}\right) \tag{26}$$

For example:
Suppose $\{a_{i,j}\}_{i,j=1}^n$ are constants in $[0, 1]$ (after translation and scaling).

Let π be a uniform random permutaton of $(1, 2, \ldots, n)$.

Let

$$X = \sum_{i=1}^n a_{i,\pi(i)}, \text{ (Hoeffding statistics).}$$

Here we could want concentration inequalities for the random variable X.

Theorem 33

$$\mathbf{P}\left(|X - \mathbf{E}(X)| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2t + 4\,\mathbf{E}(X)}\right) \tag{27}$$

Interpretation: for small $\mathbf{E}(X)$, the density of X has an exponential tail, for large $\mathbf{E}(X)$, the density of X has a Gaussian tail.

Proof: Let $\pi' = \pi \circ (I, J)$ where (I, J) is a random transposition. We allow $I = J$. Let

$$X' = \sum_{i=1}^n a_{i,\pi'(i)}$$

and

$$F(X,X') = \tfrac{1}{2}n(X - X') \text{ (clearly anti} - \text{symmetric).}$$

Now,

$$X - X' = \frac{n}{2}\left(a_{I,\pi(I)} + a_{J,\pi(J)} - a_{I,\pi(J)} - a_{J,\pi(I)}\right).$$

It follows that

$$\begin{aligned}
\mathbf{E}(F(X,X') \mid \pi) &= \frac{n}{2}\sum_{i,j}\left(a_{i,\pi(i)} + a_{j,\pi(j)} - a_{i,\pi(j)} - a_{j,\pi(i)}\right) \times \frac{1}{n^2} \\
&= \frac{1}{2n}\left[2n\sum_{i} a_{i,\pi(i)} - 2\sum_{i,j} a_{i,j}\right] \\
&= X - \mathbf{E}(X).
\end{aligned}$$

Since the right hand side depends only on X, therefore $f(X) = \mathbf{E}(F(X,X') \mid X) = X - \mathbf{E}(X)$. From this it follows that

$$\begin{aligned}
\tfrac{1}{2}\,\mathbf{E}\left(\left|(f(X) - f(X'))F(X,X')\right| \mid \pi\right) &= \frac{n}{4}\,\mathbf{E}\left((X - X')^2 \mid \pi\right) \\
&= \frac{n}{4}\cdot\frac{1}{n^2}\sum_{i,j}\left(a_{I,\pi(I)} + a_{J,\pi(J)} - a_{I,\pi(J)} - a_{J,\pi(I)}\right)^2 \\
&\quad (a_{i,j} \in [0,1]) \\
&\le \frac{1}{4n}\sum_{i,j}\left(a_{I,\pi(I)} + a_{J,\pi(J)} + a_{I,\pi(J)} + a_{J,\pi(I)}\right) \\
&= X + \mathbf{E}(X) \\
&= f(X) + 2\,\mathbf{E}(X).
\end{aligned}$$

Combining, we get

$$v(X) \le f(X) + 2\,\mathbf{E}(X). \tag{28}$$

Thus, by using Theorem 1, with $B = 1$ and $C = 2\,\mathbf{E}(X)$,

$$\mathbf{P}\left(|X - \mathbf{E}(X)| \ge t\right) \le 2\exp\left(-\frac{t^2}{2t + 4\,\mathbf{E}(X)}\right)$$

□

Theorem 2 gives a Bernstein-type inequality. Compare the above to Azuma-Hoeffding bounds:

If $X \sim \text{Binomial}(n,p)$, then

$$\mathbf{P}\left(|X - \mathbf{E}(X)|\right) = \mathbf{P}\left(|X - np|\right) \leq 2\exp\left(-\frac{t^2}{2n}\right)$$

This shows that the deviations will be of order $\sqrt{n}$. However, this bound does not hold for very small p. See also, Bennett's inequality.

The Curie-Weiss Model

In the Curie-Weiss model, the object of interest is a vector of n spins,

$$\sigma = (\sigma_1, \ldots, \sigma_n) \in \{-1, 1\}^n.$$

The probability distribution of the spin assignments follows a Gibbs measure,

$$\mathbf{P}(\sigma) = Z_\beta^{-1} \exp\left(\frac{\beta}{n} \sum_{i<j\leq n} \sigma_i \sigma_j\right), \beta \geq 0, Z_\beta = \text{normalization constant.} \tag{29}$$

For $\beta = 0$, there will be a uniform distribution on spin configurations. Otherwise, configurations will have different energies, or probabilities based on the Gibbs measure, depending on the number of $(+1)$ and (-1) spins. For a particular setting of the numbers of $(+1)$ and (-1) spins, there may be many configurations. So, the entropy is defined as the log of the number of such configurations, given the numbers of spins. This model is equivalent to an Ising model on a complete graph.

The magnetization, $m(\sigma)$, is defined as the average of the spins,

$$m(\sigma) = \frac{1}{n} \sum_{i=1}^{n} \sigma_i \in [-1, 1]. \tag{30}$$

The physical interpretation of the model is to treat the spins as atom or orientable units of a magnetic material and β as related to temperature. More specifically,

$$\beta = \frac{1}{kT},$$

where T is the temperature and $\frac{1}{k}$ is the Curie temperature. For $\beta \leq 1$, or high temperatures, the magnetization of the material will be small, $m(\sigma) \approx 0$, with high probability. However, for $\beta > 1$ there will likely be a preponderance of $(+1)$ or (-1) spins, yielding a substantial magnetization such that $m - \tanh(\beta m) \approx 0$ with high probability.

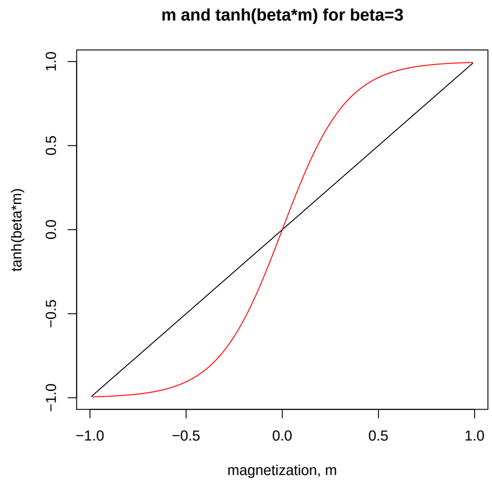


Figure 1: Magnetization m and $\tanh(\beta m)$ versus m for $\beta = 3$.

From an initial spin configuration σ, we can construct a new configuration σ' through a Gibb's sampling proceedure. First an index $I, 1 \leq I \leq n$ is chosen at random. Then σ_I is replaced by $\sigma_I^{'}$, where $\sigma_I^{'}$ is drawn from the conditional distribution given $(\sigma_j)_{j \neq I}$.

Exercise 34 *Show that $(\sigma, \sigma^{'})$ is an exchangeable pair.*

Let $F(\sigma, \sigma^{'}) = \sigma_I - \sigma_I^{'} = n(m(\sigma) - m(\sigma^{'}))$. So,

$$\begin{aligned}
\mathbf{E}\left(F(\sigma, \sigma^{'}) \mid \sigma\right) &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}(\sigma_i - \sigma_i^{'} \mid \sigma) \\
&= m(\sigma) - \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}(\sigma_i^{'} \mid \sigma).
\end{aligned}$$

Exercise 35 *Show,*

$$\mathbf{E}(\sigma_i^{'} \mid \sigma) = \tanh\left(\frac{\beta}{n}\sum_{j, j \neq i} \sigma_j\right)$$

.

Use, $\mathbf{P}(\sigma_i^{'} \mid \sigma) \propto \exp\left(\frac{\beta}{n}\sum_{j, j \neq i} \sigma_j\right)$ *and* $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

Let

$$m_i(\sigma) = \frac{1}{n} \sum_{j, j \neq i} \sigma_j = m(\sigma) - \frac{\sigma_i}{n}. \tag{31}$$

Note that $|m - m_i| = \frac{1}{n}, \forall i$. Now,

$$\begin{aligned} f(\sigma) &\equiv \mathbf{E}(F(\sigma, \sigma^{'}) \mid \sigma) \\ &= m(\sigma) - \frac{1}{n} \sum_{i=1}^{n} \tanh(\beta m_i(\sigma)). \end{aligned}$$

Exercise 36 *Show that*

$$\left| f(\sigma) - f(\sigma^{'}) \right| \leq \frac{2(1+\beta)}{n}.$$

Hint: use the fact that $\tanh(x)$ *is Lipschitz.*

Continuing, using $\left| F(\sigma, \sigma^{'} \right| \leq 2$, we have

$$\begin{aligned} v(\sigma) &= \tfrac{1}{2} \mathbf{E} \left(\left| f(\sigma) - f(\sigma^{'}) \right| \left| F(\sigma, \sigma^{'}) \right| \mid \sigma \right) \\ &\leq \tfrac{1}{2} \cdot 2 \cdot \frac{2(1+\beta)}{n} = \frac{2(1+\beta)}{n} \\ &\leq \frac{2(1+\beta)}{n}. \end{aligned}$$

Thus, using Theorem 1 with $B = 0$ and $C = \frac{2(1+\beta)}{n}$, we get

$$\mathbf{P}(|f(\sigma)| \geq t) \leq 2 \exp \left(- \frac{nt^2}{4(1+\beta)} \right).$$

It is easy to see that

$$|(m(\sigma) - \tanh(\beta m(\sigma))) - f(\sigma)| \leq \frac{\beta}{n}.$$

Thus,

$$\mathbf{P} \left(|m(\sigma) - \tanh(\beta m(\sigma))| \geq \frac{\beta}{n} + t \right) \leq 2 \exp \left(- \frac{nt^2}{4(1+\beta)} \right), \forall t \geq 0.$$

Lecture 14

Lecture date: Sept 28, 2007 Scribe: John Zhu

18 Some final remarks about the Curie-Weiss concentration

In the Curie Weiss model, if $m(\sigma) = \frac{1}{n}\sum \sigma_i$, then

$$P\left(|m - \tanh(\beta m)| \geq \frac{\beta}{n} + t\right) \leq \exp\left(-\frac{nt^2}{4(1+\beta)}\right).$$

This implies that $m - \tanh(\beta m) = O(1/\sqrt{n})$ which is the optimal result when $\beta \neq 1$.

At $\beta = 1, m - \tanh(\beta m) \asymp n^{-3/4}$.

For $\beta < 1$ and all x, $|x - \tanh(\beta x)| \geq (1-\beta)|x|$ and thus

$$|m(\sigma)| \leq \frac{|m(\sigma) - \tanh(\beta m(\sigma))|}{1-\beta} = O(1/\sqrt{n}).$$

In particular $P_\beta(|m(\sigma)| > \beta/n + t) \leq 2\exp(-\frac{n(1-\beta)^2 t^2}{4(1+\beta)})$.

When $\beta = 1, P(|m(\sigma)| \geq t) \leq C\exp(-cnt^4)$ where C, c are not dependent on n. This implies $m = O(n^{-1/4})$. Using Stein's method one can further show that $n^{1/4}m(\sigma)$ converges in distribution to $Ce^{-x^4/12}$.

Exercise 1: Prove the above inequality using the exchangeable pair theorem.

Sketch: Use part (c) of the theorem. Use $P(|m| \geq t) \leq \frac{E(m^{2k})}{t^{2k}}$ and optimize over k. Recall: $f(\sigma) = m(\sigma) - 1/n\sum\tanh(\beta m(\sigma))$, show that when $\beta = 1, |f(\sigma) - f(\sigma')| \leq cm(\sigma)^2/n + c/n^2$. Also, $|m(\sigma)|^3 \leq C|f(\sigma)| + C|m(\sigma)|/n$. Now combine.

19 KMT Strong Embedding

Theorem. (Komlós-Major-Tusnády) *Suppose $\epsilon_1, \epsilon_2, \ldots$ are i.i.d. with finite moment generating functions in a neighborhood of* 0 *with mean* 0 *and variance* 1. *Let $S_n = \sum_{i=1}^n \epsilon_i$. We* *can construct a version of* $(S_k)_{k\geq 0}$ *and a standard Brownian motion* B *on the same space such that for all* n*, and* $t \geq 0$

$$P(\max_{1\leq k\leq n}|S_k - B_k| \geq C\log n + t) \leq Ke^{-lt}$$

where C, K, l *depend only on the distribution of* ϵ_1*.*

We will prove this result for the simple random walk using ideas from Stein's method. We proceed in a series of lemmas.

Lemma 37 *Let* n *be a positive integer and suppose* A *is a continuous map from* $\mathbb{R}^n$ *to the set of* $n \times n$ *positive semidefinite matrices. Suppose the* $\|A\|$ *is bounded by a* $b < \infty$*. Then there exists a probability measure* μ *such that if random variable* $X \sim \mu$ *then for all* $\theta \in \mathbb{R}^n$,

$$\mathbf{E}\exp\langle\theta, X\rangle \leq \exp(b\|\theta\|^2) \ \textit{and}\ \ \mathbf{E}\langle X, \nabla f(X)\rangle = \mathbf{E}\,\mathrm{Tr}(A(X)\mathrm{Hess}f(X))$$

for all $f \in C^2(\mathbb{R}^n)$ *such that* $\mathbf{E}\,|f(X)|^2, \mathbf{E}\,\|\nabla f(X)\|^2, \mathbf{E}\,|\mathrm{Tr}(A(X)\mathrm{Hess}f(X)| < \infty$.

Proof: Let K denote the set of all probability measures μ on $\mathbb{R}^n$ such that $\int_{\mathbb{R}^n} x d\mu = 0$ and $\int \exp\langle\theta, x\rangle d\mu \leq \exp(b\|\theta\|^2)$ for all $\theta \in \mathbb{R}^n$. By Skorokhod representation and Fatou's lemma K is a nonempty, compact, and convex subset of the space V of finite signed measures on $\mathbb{R}^n$.

Aside: K is closed, and compactness follows from tightness.

We now use the following:

Schauder-Tychonoff Fixed Point Theorem: A continuous map from a nonempty, convex, compact subset K of a locally convex topological space into K has a fixed point.

To be continued.

Lecture 15

Lecture date: Oct 1, 2007 *Scribe: Arnab Sen*

Proof of Lemma 1 (contd.) Recall that K is set of all probability measure μ on $\mathbb{R}^n$ such that

$$\int x\mu(x) = 0 \text{ and } \int \exp\langle\theta, x\rangle\mu(dx) \le \exp(b\|\theta\|^2)\ \forall\theta \in \mathbb{R}^n.$$

We have proved that K is a nonempty, compact and convex subset of the space of all finite signed measures which is a locally convex topological vector space.

Now fix $h \in (0,1)$. Define a map $T_h : K \to V$ as follows. Given μ, let X and Z be independent random vectors with $X \sim \mu$ and $Z \sim$ standard gaussian law on $\mathbb{R}^n$. Let $T_h(\mu)$ be the law of

$$(1-h)X + \sqrt{2hA(X)}Z,$$

where $\sqrt{A(X)}$ is the positive definite square root of the matrix $A(X)$. Recall that, for any nonnegative definite B, its positive definite square root is defined as

$$\sqrt{B} = U\sqrt{\Lambda}U^T, \text{ where } B = U\Lambda U^T \text{ is a spectral decomposition of } B.$$

Linear algebra tells us that the transformation $B \mapsto \sqrt{B}$ is continuous. In fact, $\|\sqrt{B_1} - \sqrt{B_2}\| \le \|B_1 - B_2\|^{1/2}$.

Claim : If $0 < h < 1$, then $T_h(K) \subseteq K$.

$\mathbf{E}\left((1-h)X + \sqrt{2hA(X)}Z\right) = 0$ implies $\int xT_h(\mu)(dx) = 0$. For any $\theta \in \mathbb{R}^n$,

$$\begin{aligned}
\int \exp\langle\theta, x\rangle T_h(\mu)(dx) &= \mathbf{E}\exp\langle\theta, (1-h)X + \sqrt{2hA(X)}Z\rangle \\
&= \mathbf{E}\exp(\langle\theta, (1-h)X\rangle + h\langle\theta, A(X)\theta\rangle) \\
&\le \exp(bh\|\theta\|^2)\mathbf{E}\exp\langle\theta, (1-h)X\rangle \\
&\le \exp(bh\|\theta\|^2 + b(1-h)^2\|\theta\|^2) \le \exp(b\|\theta\|^2)
\end{aligned}$$

where the last step is a consequence of the easy inequality $1 - h + h^2 \le 1$ for $h \in (0,1)$. Thus the claim has been proved.

Since, $x \mapsto \sqrt{A(x)}$ is a continuous map, by continuous mapping theorem, $T_h : K \to K$ is continuous (in weak* topology). Hence, by the Schauder-Tychonoff fixed point theorem, T_h has a fixed point μ_h in K.

Suppose $X_h \sim \mu_h$. Let $Y_h := -hX + \sqrt{2hA(X)}Z$. Thus $X_h \stackrel{d}{=} X_h + Y_h$. Take any $f \in C^2(\mathbb{R}^n)$ with ∇f and Hess f bounded and uniformly continuous. Fix $h \in (0,1)$, and

note that

$$\mathbf{E}f(X_h + Y_h) = \mathbf{E}f(X_h)$$

Also, by Taylor approximation, we have

$$f(X_h + Y_h) = f(X_h) + \langle Y_h, \nabla f(X_h)\rangle + \frac{1}{2}\langle Y_h, \text{Hess} f(X_h) Y_h\rangle + \mathcal{R}_h \tag{32}$$

where $\mathcal{R}_h$ is the remainder term. Now, $\mathbf{E}\langle Y_h, \nabla f(X_h)\rangle = -h\mathbf{E}\langle X_h, \nabla f(X_h)\rangle$. On the other hand,

$$\begin{aligned}\mathbf{E}\langle Y_h, \text{Hess} f(X_h) Y_h\rangle &= 2h\mathbf{E}\left(Z^T\sqrt{A(X_h)}(\text{Hess} f(X_h))\sqrt{A(X_h)}Z\right) + O(h^{3/2})\\ &= 2h\mathbf{E}\text{Tr}(A(X_h)\ \text{Hess}\ f(X_h)) + O(h^{3/2})\end{aligned}$$

Also, from the conditions on function f, one can show that $\lim_{h\to 0} h^{-1}\mathbf{E}|\mathcal{R}_h| = 0$.

Since the collection $\{\mu_h\}_{0<h<1} \subseteq K$, and K is compact, it has a limit point $\mu \in K$ as $h \to 0$. Let $X \sim \mu$. From (32), after taking expectation and dividing by h, we have

$$\mathbf{E}\langle X_h, \nabla f(X_h)\rangle = \mathbf{E}\text{Tr}(A(X_h)\text{Hess} f(X_h)) + O(h^{1/2}) + h^{-1}\mathbf{E}\mathcal{R}_h.$$

Now, letting $h \to 0$, we get, by uniform integrability,

$$\mathbf{E}\langle X, \nabla f(X)\rangle = \mathbf{E}\text{Tr}(A(X)\text{Hess} f(X)). \tag{33}$$

Then, the probability μ satisfies the criteria of the theorem for 'nice' functions $f \in C^2(\mathbb{R}^n)$ with ∇f and Hess f bounded and uniformly continuous. The extension to more general f, as required in the lemma, can be done via standard approximation arguments. □

Next lemma tells us that if X and A as above, then the deviation of X_i and X_j can be controlled by entries of A. An intelligent choice of A will thus lead to a greater control over difference $|X_i - X_j|$.

Lemma 2 *Let $A(x) = ((a_{ij}(x)))$ and X be as in Lemma 1. Take any $1 \le i < j \le n$. Let*

$$v_{ij}(x) := a_{ii}(x) + a_{jj}(x) - 2a_{ij}(x).$$

Then for all $\theta \in \mathbb{R}^n$,

$$\mathbf{E}\exp(\theta|X_i - X_j|) \le 2\mathbf{E}\exp(2\theta^2 v_{ij}(x)).$$

Proof. Take any positive integer k. Define $f : \mathbb{R}^n \to \mathbb{R}$ as

$$f(x) = (x_i - x_j)^{2k}.$$

Then a simple calculation shows that

$$\langle X, \nabla f(X)\rangle = 2k(x_i - x_j)^{2k}.$$

and

$$\mathrm{Tr}(A(X)\mathrm{Hess}f(X)) = 2k(2k-1)(x_i - x_j)^{2k-2}v_{ij}(x).$$

By poitive definiteness of A, it is easy to see that $v_{ij}(x) \geq 0$. An application of Hölder inequality gives

$$\mathbf{E}|\mathrm{Tr}(A(X)\mathrm{Hess}f(X))| \leq 2k(2k-1)\left(\mathbf{E}(X_i - X_j)^{2k}\right)^{\frac{k-1}{k}}\left(\mathbf{E}v_{ij}(X)^k\right)^{\frac{1}{k}}.$$

From the identity (33), we can now, after a simple computation, conclude that

$$\mathbf{E}(X_i - X_j)^{2k} \leq (2k-1)^k \mathbf{E}v_{ij}(X)^k.$$

Now using $e^{|x|} \leq e^x + e^{-x} = 2\sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}$, we get

$$\begin{aligned}
\mathbf{E}\exp(\theta|X_i - X_j|) &\leq 2\sum_{k=0}^{\infty} \frac{\theta^{2k}\mathbf{E}(X_i - X_j)^{2k}}{(2k)!} \\
&\leq 2\sum_{k=0}^{\infty} \frac{\theta^{2k}(2k-1)^k \mathbf{E}v_{ij}(X)^k}{(2k)!} \\
&\leq 2\sum_{k=0}^{\infty} \frac{\theta^{2k}2^k \mathbf{E}v_{ij}(X)^k}{k!} = 2\mathbf{E}\exp\left(2\theta^2 v_{ij}(X)\right)
\end{aligned}$$

where in the above step we use the following inequality

$$\frac{(2k-1)^k}{(2k)!} \leq \frac{2^k}{k!}.$$

This completes the proof. □

Lemma 3 *Let ρ be a probability density function on $\mathbb{R}$ which is positive everywhere. Suppose $\int_{-\infty}^{\infty} x\rho(x)dx = 0$ and $\int_{-\infty}^{\infty} x^2\rho(x)dx < \infty$. Define,*

$$h(x) = \frac{\int_x^{\infty} y\rho(y)dy}{\rho(x)}.$$

Let $X \sim \rho$. Then

$$\mathbf{E}X\varphi(X) = \mathbf{E}h(X)\varphi'(X) \tag{34}$$

for each absolutely continuous function φ such that both sides are well defined and $\mathbf{E}|h(X)\varphi(X)| < \infty$. Moreover, if h_1 is another function satisfying (34) for all Lipschitz φ, then $h_1 = h$ a.e.

Conversely, if Y is a random variable such that (34) holds with Y in place of X, for all absolutely continuous φ such that $|\varphi(x)|, |x\varphi(x)|$ and $|h(x)\varphi'(x)|$ are uniformly bounded, then $Y \stackrel{d}{=} X$.

Proof. We will only prove the first claim of the theorem (that too partially !) which is a direct application of integration by parts. Let $u(x) := h(x)\rho(x)$. Note that since $\int_{-\infty}^{\infty} x\rho(x)dx = 0$, u can be written as $u(x) = \int_x^{\infty} y\rho(y)dy = -\int_{-\infty}^{x} y\rho(y)dy$. Thus, $u(x) > 0$ for all $x \in \mathbb{R}$. Also, $\lim_{|x|\to\infty} u(x) = 0$ and by Fubini, it is easy to verify that $\int_{-\infty}^{\infty} u(x)dx = \mathbf{E}h(X) = \mathbf{E}X^2 < \infty$. Then for any bounded Lipschitz function φ,

$$\begin{aligned}\int_{-\infty}^{\infty} x\varphi(x)\rho(x)dx &= \varphi(x)(-u(x))\,\big|_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} \varphi'(x)(-u(x))dx \\ &= \int_{-\infty}^{\infty} \varphi'(x)h(x)\rho(x)dx.\end{aligned}$$

This proves one part of the lemma.

Exercise 38 Finish the rest of the proof.

Lecture 16

Lecture date: Oct 3, 2007 *Scribe: Anand Sarwate*

20 The general plan of attack

Recall the setup from the last lecture: we have iid ± 1 symmetric random variables $\varepsilon_1, \varepsilon_2, \ldots,$ and

$$S_n = \sum_{i=1}^{n} \varepsilon_i \tag{35}$$

is the simple random walk on the integers. Our goal is to construct a version of $(S_n)_{n \geq 0}$ and Brownian motion $(B_t)_{t \geq 0}$ on the same probability space such that for all n

$$\max_{i \leq k \leq n} |S_k - B_k| = O(\log n) \ . \tag{36}$$

A cartoon of the coupling is shown in Figure 20.

Figure 2: Coupling between the simple random walk and Brownian motion. At time n the variance is $\sqrt{n}$ but the maximum deviation between the two processes is at most $O(\log n)$.

Figure 3: A tighter coupling between a simple random walk and the Brownian bridge, conditioned on $S_n = B_n$.

The original KMT proof used an explicit construction of the coupling between the random walk and Brownian motion, whereas we rely on the Schauder-Tychonoff fixed point theorem. Our line of attack to construct the coupling is to couple a conditional process to the Brownian bridge. Figure 2 shows this coupling.

The induction hypothesis we will use is the following: given a possible value S_n, we can construct a random walk $S_0, S_1, \ldots, S_n$ with S_n having that value and a Brownian motion $(B_t)_{t \leq n}$ conditioned to have $B_n = S_n$ such that for all $\lambda < \lambda_0$ we have

$$\mathbf{E} \exp\left(\lambda \max_{i \leq n} |S_i - B_i|\right) \leq \exp\left(C \log n + \frac{K\lambda^2 S_n^2}{n}\right) \ , \tag{37}$$

where C, K, and λ_0 must be chosen appropriately.

To see how to use this hypothesis, fix n and the value of S_n. Take $n/3 \leq k \leq 2n/3$ and assume that the induction hypothesis holds for all $n' < n$. The main step is to do a pointwise coupling of S_k with B_k such that the deviation is at most $O(1)$. Then we condition on a value for S_k and use the induction hypothesis to couple $(S_i)_{i\leq k}$ with $(B_i)_{i\leq k}$ and $(S_i)_{k\leq i\leq n}$ with $(B_i)_{k\leq i\leq n}$. Then by piecing it together you get that the result holds for n as well.

21 Pointwise coupling

The central question we have to answer is this : how do we couple something that is almost Gaussian to something that is exactly Gaussian such that we obtain exponentially decaying tails?

Suppose X is a random variable with density ρ, $\mathbf{E}\,X = 0$ and $\mathbf{E}\,X^2 < \infty$. Let

$$h(x) = \frac{\int_x^\infty y\rho(y)dy}{\rho(x)} \; . \tag{38}$$

We know that for all well-behaved φ,

$$\mathbf{E}[X\varphi(X)] = \mathbf{E}[h(X)\varphi'(X)] \; . \tag{39}$$

The idea is that if $h(X) \approx \sigma^2$ with high probability, then X is approximately Gaussian with variance σ^2. Thus our objective is to construct a joint distribution on (X, Z) with marginals $X \sim \rho$ and $Z \sim \mathcal{N}(0, \sigma^2)$ such that the difference $X - Z$ is controlled by $h(X) - \sigma^2$.

Let

$$A(x_1, x_2) = \begin{pmatrix} h(x_1) & \sigma\sqrt{h(x_1)} \\ \sigma\sqrt{h(x_1)} & \sigma^2 \end{pmatrix} \; . \tag{40}$$

Note that $A(x_1, x_2)$ does not depend on x_2, and that it is positive semidefinite. By Lemma 1 we can construct a probability measure μ on $\mathbb{R}^2$ such that if $(x_1, x_2)^T \sim \mu$ then

$$\mathbf{E}\left[\left\langle \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \nabla f \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \right\rangle\right] = \mathbf{E}\left[\mathrm{Tr}\left(A(X_1, X_2)\,\mathrm{Hess}\, f(X_1, X_2)\right)\right] \tag{41}$$

for all suitable f. Rewriting this a bit:

$$\mathbf{E}\left[X_1 \frac{\partial f}{\partial X_1} + X_2 \frac{\partial f}{\partial X_2}\right] = \mathbf{E}\left[h(X_1)\frac{\partial^2 f}{\partial X_1^2} + 2\sigma\sqrt{h(X_1)}\frac{\partial^2 f}{\partial X_1 \partial X_2} + \sigma^2 \frac{\partial^2 f}{\partial X_2^2}\right] \; . \tag{42}$$

Now take $\varphi : \mathbb{R} \to \mathbb{R}$ such that $|\varphi(x)|$, $|x\varphi(x)|$ and $|h(x)\varphi'(x)|$ are uniformly bounded, and let Φ be an antiderivative of φ so that $\Phi' = \varphi$.

Consider $f(x_1, x_2) = \Phi(x_1)$. Then all the x_2 terms in (42) vanish, so

$$\mathbf{E}[X_1 \varphi(X_1)] = \mathbf{E}[h(X_1)\varphi'(X_1)] \ , \tag{43}$$

and by the previous lemma $X_1 \sim \rho$. Similarly, taking $f(x_1, x_2) = \Phi(x_2)$, we get

$$\mathbf{E}[X_2 \varphi(X_2)] = \sigma^2 \, \mathbf{E}[\varphi'(X_2)] \ , \tag{44}$$

so $X_2 \sim \mathcal{N}(0, \sigma^2)$. Note that the off-diagonal terms in (40) vanish for these two choices of f. If we set those terms to 0 then X_1 and X_2 would be independent.

By an earlier Lemma, we now have the bound:

$$\begin{aligned}
\mathbf{E} \exp\left(\theta |X_1 - X_2|\right) &\leq 2 \, \mathbf{E} \exp\left(2\theta^2 v_{12}(x_1, x_2)\right) && (45)\\
&= 2 \, \mathbf{E} \exp\left(2\theta^2 (a_{11}(x_1, x_2) + a_{22}(x_1, x_2) - 2a_{12}(x_1, x_2))\right) && (46)\\
&= 2 \, \mathbf{E} \exp\left(2\theta^2 (h(x_1) + \sigma^2 - 2\sigma\sqrt{h(x_1)})\right) && (47)\\
&= 2 \, \mathbf{E} \exp\left(2\theta^2 (\sqrt{h(x_1)} - \sigma)^2\right) \ . && (48)
\end{aligned}$$

Remark: In a sense, the choice of A in (40) is the tightest coupling possible. Consider the following alternate choice for A:

$$A(x_1, x_2) = \begin{pmatrix} h(x_1) & h(x_1) \wedge \sigma^2 \\ h(x_1) \wedge \sigma^2 & \sigma^2 \end{pmatrix} \ . \tag{49}$$

For this A we get $v_{12}(x_1, x_2) = |h(x_1) - \sigma^2|$. In coupling $h(S_n) = n + O(\sqrt{n})$, so $v_{12}(x_1, x_2) = O(\sqrt{n})$. This choice corresponds to the Skorohod embedding, which gives

$$\max_{1 \leq k \leq n} |S_k - B_k| = O(n^{1/4}) \ . \tag{50}$$

This bound on the deviation is the best possible for summands with finite 4-th moment. For finite p-th moment we can get $O(n^{1/p})$. The assumptions in the KMT are that the moment generating function is finite in a neighborhood of 0, which gives a $O(\log n)$ deviation.

Why do we choose this particular function $h(\cdot)$? Suppose $X_1, X_2, \ldots, X_n$ are iid, distributed according to the density ρ, with mean 0 and unit variance, and define $h(\cdot)$ as in (38). Let

$$S = \sum_{i=1}^{n} X_i \ . \tag{51}$$

What is the function $h_S(\cdot)$ corresponding to the density of S? If we define as before

$S_i = S - X_i$, we can calculate, using (39) :

$$\mathbf{E}[S\varphi(S)] = \sum_{i=1}^{n} \mathbf{E}\left[X_i \varphi\left(S_i + X_i\right)\right] \tag{52}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{E}\left[\varphi'\left(S_i + X_i\right) h(X_i)\right] \tag{53}$$

$$= \mathbf{E}\left[\varphi'(S)\left(\sum_{i=1}^{n} h(X_i)\right)\right] \ . \tag{54}$$

Now, it is easy to check that $\mathbf{E}\, h(X_i) = \mathbf{E}\, X_i^2 = 1$. Therefore, using (39) again we see that

$$h_S(S) = \mathbf{E}\left[\sum_{i=1}^{n} h(X_i) \;\middle|\; S\right] = n + O(\sqrt{n}). \tag{55}$$

We now show the $O(1)$ bound on the coupling:

$$(\sqrt{h_S(S)} - \sigma)^2 = n\left(\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right)^{1/2} - 1\right)^2 \tag{56}$$

$$= n\left(1 + O\left(\frac{1}{\sqrt{n}}\right) + \ldots - 1\right)^2 \tag{57}$$

$$= O(1) \ . \tag{58}$$

Lecture 17

Lecture date: Oct. 5, 2007 *Scribe: Partha S. Dey*

Lemma 3 *Suppose W is a random variable with $\mathbf{E}W = 0$ and $\mathbf{E}W^2 < \infty$. Let T be another random variable defined on the same probability space as W, satisfying*

$$\mathbf{E}W\varphi(W) = \mathbf{E}T\varphi'(W)$$

for all Lipschitz function φ. Suppose $|T|$ is a.s. bounded by a constant. Then, given any $\sigma^2 > 0$, we can construct a version of W and a $N(0, \sigma^2)$ random variable Z on the same probability space such that for any $\theta \in \mathbb{R}$,

$$\mathbf{E}\exp(\theta\,|W - Z|) \leq 2\mathbf{E}\exp(2\theta^2\sigma^{-2}(T - \sigma^2)^2).$$

Exercise 39 *Relax the condition that "$|T|$ is a.s. bounded by a constant".*

Before going to the proof of lemma 3, let us state some exercises which can probably be solved using Stein's method.

Exercise 40 *In the Curie-Weiss model for $\beta < 1$, prove a CLT for the magnetization $n^{-1}\sum_{i=1}^n \sigma_i$.*

Exercise 41 *In the Curie-Weiss model for $\beta < 1$, prove a version of Tusnády's lemma for the sum of spins.*

Exercise 42 *For the Ising model on the cycle of n points, prove a version of Tusnády's lemma for the sum of spins.*

Exercise 43 *For the Ising model on the cycle (or chain) of n points, consider the process $(S_k)_{k=1}^n$ where $S_k = \sum_{i=1}^k \sigma_i$. Prove a version of the KMT theorem for this process.*

Now let us prove lemma 3.

Proof: Recall that by assumption we have, $\mathbf{E}W\varphi(W) = \mathbf{E}T\varphi'(W) = \mathbf{E}(\mathbf{E}(T|W)\varphi'(W))$ for all Lipschitz φ. We'll prove the lemma in two steps.

First assume that W has a density ρ with respect to the lebesgue measure which is positive and continuous everywhere. Define the function h by

$$h(x) = \frac{\int_x^\infty y\rho(y)dy}{\rho(x)}.$$

Then by uniqueness of h (see the second assertion of Lemma 3), we have

$$h(x) = \mathbf{E}(T|W = x).$$

Note that h is nonnegative by definition. So we can define the function A from $\mathbb{R}^2$ into the set of 2×2 nonnegative definite matrices as

$$A(x_1, x_2) = \begin{pmatrix} h(x_1) & \sigma\sqrt{h(x_1)} \\ \sigma\sqrt{h(x_1)} & \sigma^2 \end{pmatrix}.$$

Continuity of ρ implies that h is continuous. Since T is bounded a.s., so is h (because $h = \mathbf{E}(T|W)$). Hence using lemma 1 we can construct a random vector $X = (X_1, X_2)$ such that

$$\mathbf{E}\langle X, \nabla f(X)\rangle = \mathbf{E}\ \mathrm{Tr}(A(X)\mathrm{Hess}f(X))$$

for all $f \in C^2(\mathbb{R}^2)$ such that the expectations $\mathbf{E}\,|f(X)|^2$, $\mathbf{E}\,\|\nabla f(X)\|^2$ and $\mathbf{E}\,|\mathrm{Tr}(A(X)\mathrm{Hess}f(X))|$ are finite.

By the observation made in the last lecture we have $X_1 \sim \rho$ and $X_2 \sim \mathrm{N}(0, \sigma^2)$. Also by lemma 2 we have for all $\theta \in \mathbb{R}$,

$$\mathbf{E}\exp(\theta\,|X_1 - X_2|) \leq 2\mathbf{E}\exp(2\theta^2 \upsilon_{12}(X_1, X_2))$$

where

$$\begin{aligned}
\upsilon_{12}(x_1, x_2) &= a_{11}(x_1, x_2) + a_{22}(x_1, x_2) - 2a_{12}(x_1, x_2) \\
&= h(x_1) + \sigma^2 - 2\sigma\sqrt{h(x_1)} \\
&= (\sqrt{h(x_1)} - \sigma)^2 = \frac{(h(x_1) - \sigma^2)^2}{(\sqrt{h(x_1)} + \sigma)^2} \leq \frac{(h(x_1) - \sigma^2)^2}{\sigma^2}.
\end{aligned}$$

The last inequality holds because $h(x) \geq 0$. Since $h(X_1)$ has the same distribution as $h(W)$ and $h(W) = \mathbf{E}(T|W)$ using Jensen's inequality for conditional expectation we have

$$(\mathbf{E}h(X_1) - \sigma^2)^2 = (\mathbf{E}(T|W) - \sigma^2)^2 \leq \mathbf{E}((T - \sigma^2)^2|W)$$

and

$$\exp(2\theta^2\sigma^{-2}\mathbf{E}((T - \sigma^2)^2|W)) \leq \mathbf{E}(\exp(2\theta^2\sigma^{-2}(T - \sigma^2)^2)|W).$$

Hence, if W has a density ρ which is continuous and positive everywhere we can construct a version of W and a $\mathrm{N}(0, \sigma^2)$ random variable Z on the same probability space such that for any $\theta \in \mathbb{R}$,

$$\mathbf{E}\exp(\theta\,|W - Z|) \leq 2\mathbf{E}\exp(2\theta^2\sigma^{-2}(T - \sigma^2)^2).$$

Let us now drop the assumption that W has a density ρ which is continuous and positive everywhere. Take any $\epsilon > 0$. Let $W_\epsilon = W + \epsilon Y$ where Y is an independent standard gaussian random variable. If ν denotes the law of W on the real line, then W_ϵ has the probability density function

$$\rho_\epsilon(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\epsilon} e^{-\frac{(x-y)^2}{2\epsilon^2}} d\nu(y).$$

ρ_ϵ is positive and continuous everywhere by the above representation and DCT. Again, note that for any Lipschitz function φ,

$$\begin{aligned} \mathbf{E}(W_\epsilon \varphi(W_\epsilon)) &= \mathbf{E}(W\varphi(W + \epsilon Y)) + \epsilon \mathbf{E}(Y\varphi(W + \epsilon Y)) \\ &= \mathbf{E}(T\varphi'(W + \epsilon Y)) + \epsilon^2 \mathbf{E}(\varphi'(W + \epsilon Y)) \\ &= \mathbf{E}((T + \epsilon^2)\varphi'(W_\epsilon)) \end{aligned}$$

Thus, by what we have already proved, we can construct a version of W_ϵ and a $\mathrm{N}(0, \sigma^2 + \epsilon^2)$ random variable Z_ϵ on the same probability space such that for all $\theta \in \mathbb{R}$,

$$\mathbf{E}\exp(\theta\,|W_\epsilon - Z_\epsilon|) \leq 2\mathbf{E}\exp\left(\frac{2\theta^2((T + \epsilon^2) - (\sigma^2 + \epsilon^2))^2}{\sigma^2 + \epsilon^2}\right) \leq 2\mathbf{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right).$$

Let μ_ϵ be the law of the pair (W_ϵ, Z_ϵ) on $\mathbb{R}^2$. The family $\{\mu_\epsilon\}_{\epsilon>0}$ is tight by the bound on the moment generating function in lemma 1. Let μ_0 be a limit point as $\epsilon \to 0$ and let $(W_0, Z_0) \sim \mu_0$. Then W_0 has the same law as W, and Z_0 is standard gaussian. By Skorohod representation and Fatou's lemma we have

$$\mathbf{E}\exp(\theta\,|W_0 - Z_0|) \leq \liminf_{\epsilon\to 0} \mathbf{E}\exp(\theta\,|W_\epsilon - Z_\epsilon|) \leq 2\mathbf{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right).$$

This completes the proof. □

Exercise 44 Show that the assumption that $\mathbf{E}W\varphi(W) = \mathbf{E}T\varphi'(W)$ for all Lipschitz function φ implies that $W\mathbf{1}_{\{W\neq 0\}}$ has a density w.r.t the lebesgue measure on $\mathbb{R}$. (Note that, in fact, W can have positive mass at 0. Consider the random variable $W = ZI$ where $Z \sim \mathrm{N}(0,1)$ and $I \sim \mathrm{Bin}(1,p)$ are independent. Then $\mathbf{E}W\varphi(W) = \mathbf{E}I\varphi'(W)$ for all Lipschitz φ, but $W = 0$ with probability $1-p$.)

Remark 45 The matrix $A(x_1, x_2)$ used in the proof of lemma 3 has a different interpretation. It is possible to define two stochastic processes $(X_t)_{t\geq 0}, (Z_t)_{t\geq 0}$ such that $(X_t)_{t\geq 0}$ has stationary distribution with density ρ and $(Z_t)_{t\geq 0}$ is an Ornstein-Uhlenbeck process with stationary distribution $\mathrm{N}(0, \sigma^2)$ and the matrix $A(X_t, Z_t)$ is the volatility matrix of the process $(X_t, Z_t)_{t\geq 0}$. But it is difficult to get similar bound on m.g.f using usual stochastic process methods.

Lecture 18

Lecture date: Oct 8, 2007 *Scribe: Guy Bresler*

22 Tusnády's Lemma

Lemma 46 (Tusnády's Lemma, Lemma 3.10 in handout) *Let $\epsilon_1, \ldots, \epsilon_n$ be i.i.d. symmetric ± 1 random variables and $S_n = \sum_{i=1}^n \epsilon_i$ be the simple random walk. Then there exist universal constants κ and $\rightarrow > 0$ such that it is possible to construct S_n and Z_n ($Z_n \sim \mathcal{N}(0,1)$) on the same probability space so that for all $|\theta| \leq \rightarrow$, $E\exp(\theta|S_n - Z_n|) \leq \kappa$.*

Implication: $P(|S_n - Z_n| \geq t) \leq e^{-\rightarrow t}\kappa$. (We are not worried about dependence on θ, only care about fixed $\rightarrow$.)

We will use Lemma 3.4 from the previous lecture:

Lemma 47 (Lemma 3.4 in handout) *If W is a random variable with mean 0, finite variance, and $\mathbf{E}W\varphi(W) = ET\varphi'(W)$ for all Lipschitz φ, where $|T|$ is a.s. bounded, then for any $\sigma > 0$ it is possible to construct $Z \sim \mathcal{N}(0, \sigma^2)$ such that $\mathbf{E}\exp(\theta|W - Z|) \leq 2\mathbf{E}\exp(2\theta^2(T - \sigma^2)^2/\sigma^2)$ for all $\theta > 0$.*

The idea of the proof is as follows. Take $\widetilde{W} = S_n + Y$, where $Y \sim \text{Unif}[-1,1]$ independent of S_n, and show that

$$\mathbf{E}\widetilde{W}\varphi(\widetilde{W}) = \mathbf{E}T\varphi'(\widetilde{W}), \tag{59}$$

where $T = n - S_nY + (1 - Y^2)/2$. Supposing we can show this, take $\sigma^2 = n$. Then

$$\frac{(T - \sigma^2)^2}{\sigma^2} = \frac{(S_nY - \frac{1-Y^2}{2})^2}{n} \leq \frac{2S_n^2 + \frac{1}{2}}{n},$$

where the inequality follows from the fact that $(a+b)^2 \leq 2(a^2 + b^2)$ and $Y \in [-1,1]$. This allows us to apply Lemma 3.4 to get that we can construct $\widetilde{W}$ and $Z \sim \mathcal{N}(0,1)$ on the same space such that

$$\mathbf{E}\exp(\theta|\widetilde{W} - Z|) \leq 2\mathbf{E}\exp\left(2\theta^2\left(\frac{2S_n^2 + \frac{1}{2}}{n}\right)\right). \tag{60}$$

"Famous Trick": We know the moment generating function for $\xi \sim \mathcal{N}(0,1)$ is $\mathbf{E}e^{t\xi} = e^{t^2/2}$. It follows (by first conditioning on S_n) that

$$\mathbf{E}e^{\alpha S_n^2/2n} = \mathbf{E}e^{\sqrt{\alpha}\xi S_n/\sqrt{n}},$$

for any $\alpha > 0$. Next, conditioning on ξ, we get that the above is equal to

$$\mathbf{E}\left(\mathbf{E}\left(e^{\sqrt{\alpha}\xi\epsilon_1/\sqrt{n}}\right)^n\right) = \mathbf{E}\left(\cosh^n\left(\frac{\sqrt{\alpha}\xi}{\sqrt{n}}\right)\right).$$

It is an easy exercise to show that $\cosh(x) \leq e^{x^2/2}$ from the series expansion. Using this, we have that for $\alpha < 1$

$$\mathbf{E}\left(e^{\alpha S_n^2/2n}\right) \leq \mathbf{E}\left[\left(e^{\frac{\alpha\xi^2}{2n}}\right)^n\right] = \mathbf{E}e^{\alpha\xi^2/2} < \infty.$$

Together with equation (60) this completes the proof of Tusnády's lemma, assuming (59). We proceed with showing that (59) holds. We will use the following lemma.

Lemma 48 (Lemma 3.7 in handout) *Suppose X, Y are independent random variables, $X \sim Unif\{\pm 1\}$, $Y \sim Unif[-1, 1]$. Then for all Lipschitz φ,*

$$\mathbf{E}X\varphi(X+Y) = \mathbf{E}[(1-XY)\varphi(X+Y)]$$

and

$$\mathbf{E}Y\varphi(X+Y) = \mathbf{E}\left[\left(\frac{1-Y^2}{2}\right)\varphi'(X+Y)\right]$$

Proof: From the densities of X and Y we have

$$\mathbf{E}[(1-XY)\varphi'(X+Y)] = \frac{1}{4}\int_{-1}^{1}(1+y)\varphi'(-1+y)dy + \frac{1}{4}\int_{-1}^{1}(1-y)\varphi'(1+y)dy.$$

The proof follows by applying integration by parts to both terms. The other identity is proved similarly, and the proof is omitted. □

For ease of notation, write

$$S = S_n \quad S^- = S_{n-1} = \sum_{i=1}^{n-1}\epsilon_i \quad X = \epsilon_n.$$

Let $\mathbf{E}^-$ denote the conditional expectation given $\epsilon_1, \ldots, \epsilon_{n-1}$. By Lemma 48,

$$\mathbf{E}^-[X\varphi(\widetilde{W})] = \mathbf{E}^-X\varphi(S^- + X + Y)] = \mathbf{E}^-[(1-XY)\varphi'(S^- + X + Y)],$$

and taking expectations gives

$$\mathbf{E}\epsilon_n\varphi(\widetilde{W}) = \mathbf{E}[(1-\epsilon_n Y)\varphi'(\widetilde{W})].$$

By symmetry of $\epsilon_1, \dots, \epsilon_n$ we can take a sum over the previous equation to get

$$\mathbf{E}S\varphi(\widetilde{W}) = \mathbf{E}[(n - SY)\varphi'(\widetilde{W})].$$

Applying Lemma 48 once again, we have

$$\mathbf{E}Y\varphi(\widetilde{W}) = \mathbf{E}Y\varphi(S^- + X + Y) = \mathbf{E}\left[\left(\frac{1-Y^2}{2}\right)\varphi'(\widetilde{W})\right].$$

Adding these last two equations gives the desired result of equation (59), i.e.

$$\mathbf{E}\widetilde{W}\varphi(\widetilde{W}) = \mathbf{E}(S+Y)\varphi(\widetilde{W}) = \mathbf{E}\left[\left(n - S_nY + \frac{(1-Y^2)}{2}\right)\varphi'(\widetilde{W})\right] = \mathbf{E}T\varphi'(\widetilde{W}).$$

This concludes the proof of Tusnady's Lemma.

23 Bounds on a Pinned Random Walk

Lemma 49 (Lemma 3.4 in handout) *Suppose $\epsilon_1, \dots, \epsilon_n$ are each in $\{-1, 1\}$. Let π be a uniform random permutation of $\{1, \dots, n\}$ and for each k, let $S_k = \sum_{i=1}^k \epsilon_{\pi(i)}$. Let $W_k = S_k - \frac{k}{n}S_n$. Then for all $\theta \in \mathbb{R}$, for all $k, 1 \le k \le n$, and for all possible values of S_n,*

$$\mathbf{E}\exp\left(\frac{\theta W_k}{\sqrt{k}}\right) \le e^{\theta^2}.$$

The proof is by the method of exchangeable pairs. We only need this lemma in order to prove the next lemma, so the proof is omitted here. Note that the bound does not depend on S_n; this is because W_k fluctuates the most when $S_n = 0$. When $S_n = n$, for example, $\epsilon_i = 1$ for all i.

Lemma 50 (Lemma 3.6 in handout) *There exists a universal constant α_0 such that for all n and all possible values of S_n, any k with $\frac{n}{3} \le k \le \frac{2n}{3}$, and any $\alpha \le \alpha_0$,*

$$\mathbf{E}\exp\left(\frac{\alpha S_n^2}{k}\right) \le \exp\left(1 + \frac{3}{4}\frac{\alpha S_n^2}{n}\right).$$

Proof: Letting $Z \sim \mathcal{N}(0,1)$, we have

$$\begin{aligned}
\mathbf{E}\exp\left(\frac{\alpha S_k^2}{k}\right) &\overset{(a)}{=} \mathbf{E}\exp\left(\sqrt{\frac{2\alpha}{k}} Z S_k\right) \\
&\overset{(b)}{=} \mathbf{E}\exp\left(\sqrt{\frac{2\alpha}{k}} Z W_k + \sqrt{\frac{2\alpha}{k}} Z \frac{k S_n}{n}\right) \\
&\overset{(c)}{\leq} \mathbf{E}\exp\left(2\alpha Z^2 + \sqrt{\frac{2\alpha}{k}} Z \frac{k S_n}{n}\right) \\
&\overset{(d)}{=} \frac{1}{\sqrt{1-4\alpha}} \exp\left(\frac{\alpha k S_n^2}{(1-4\alpha)n^2}\right),
\end{aligned}$$

where the steps follow by (a) same trick (moment generating function of standard normal) as presented earlier; (b) definition of W_k; (c) Lemma 49; (d) taking expectation with respect to Z, as long as $\alpha < 1/4$. The desired result follows by noting the range of possible k and choosing α_0 sufficiently small. □

Lecture 19

Lecture date: Oct. 10, 2007 *Scribe: Laura Derksen*

24 Tsunády's lemma for a given total sum

Let $\epsilon_1, ..., \epsilon_n$ be i.i.d. symmetric random variables taking values in $\{-1, 1\}$ and let $S_n = \sum_{i=1}^n \epsilon_i$. Let π be a uniformly random permutation, and let $S_k = \sum_{i=1}^k \epsilon_{\pi(i)}$ and $W_k = S_k - \frac{kS_n}{n}$.

Lemma 51 (Lemma 3.8 in handout) *There exist universal constants $c > 0$ and $\theta_0 > 0$ such that for any n, any value of S_n, and any fixed k such that $\frac{n}{3} \leq k \leq \frac{2n}{3}$, it is possible to construct W_k and Z_k (where $Z_k \sim N(0, k(n-k)/n)$) on the same probability space such that for any θ with $|\theta| < \theta_0$*

$$\mathbf{E}[\exp(\theta|W_k - Z_k|)] \leq \exp\left(1 + \frac{c\theta^2 S_n^2}{n}\right).$$

Proof: Fix k, and for now denote W_k by W. Define $\widetilde{W} = W + Y$ where Y is a uniform random variable on $[-1, 1]$ and is independent of all of the other random variables we'll use. We would like to show that

$$\mathbf{E}(\widetilde{W}\varphi(\widetilde{W})) = \mathbf{E}(T\varphi'(\widetilde{W}) \tag{61}$$

for any Lipschitz function φ and

$$T = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=k+1}^{n} a_{ij} + (1 - Y^2)/Z \tag{62}$$

where

$$a_{ij} = 1 - \epsilon_{\pi(i)}\epsilon_{\pi(j)} - (\epsilon_{\pi(i)} - \epsilon_{\pi(j)})Y. \tag{63}$$

We can expect that $T \approx k(n-k)/n$. In fact, it can be shown that if $\sigma^2 = k(n-k)/n$ then

$$\frac{(T - \sigma^2)^2}{\sigma^2} \leq C\left(\frac{S_k^2}{k} + \frac{S_n^2}{n} + 1\right) \tag{64}$$

where C is a universal constant.

By Lemma 3.6, we know that

$$\mathbf{E}[\exp(\theta S_k^2/k)] \leq A\exp(c\theta S_n^2/n) \tag{65}$$

where A and c are constants.

The proof can now be completed using this and Lemma 3.4. Now let's show (61).

Let us write

$$W = S_k - \frac{kS_n}{n} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=k=1}^{n}(\epsilon_{\pi(i)} - \epsilon_{\pi(j)}) \tag{66}$$

and notice that if we fix i and j and condition on $(\pi(l))_{l\notin\{i,j\}}$, then the conditional expectation of $(\epsilon_{\pi(i)} - \epsilon_{\pi(j)})$ must be zero.

So, we fix i and j such that $i \le k \le j$ and condition on $(\pi(l))_{l\notin\{i,j\}}$. Denote S_n by S and let $S^- = \sum_{l\notin\{i,j\}} \epsilon_{\pi(l)}$. Let $\mathbf{E}^-$ denote conditional expectation, and consider

$$\mathbf{E}^-[(\epsilon_{\pi(i)} - \epsilon_{\pi(j)})\varphi(\widetilde{W})]. \tag{67}$$

Case 1: if $S \neq S^-$, that is, $\epsilon_{\pi(i)} = \epsilon_{\pi(j)}$, then (67) $= 0$.

Case 2: if $S = S^-$, that is, $\epsilon_{\pi(i)} = -\epsilon_{\pi(j)}$, then let $X = \frac{1}{2}(\epsilon_{\pi(i)} - \epsilon_{\pi(j)}) = \epsilon_{\pi(i)}$. So, $W = W^- + X$ where

$$W^- = \sum_{l=1}^{k}\epsilon_{\pi(l)} - \epsilon_{\pi(i)} - \frac{kS}{n}.$$

Now, using Lemma 3.7 we obtain

$$\begin{aligned}
\mathbf{E}^-[(\epsilon_{\pi(i)} - \epsilon_{\pi(j)})\varphi(\widetilde{W})] &= \mathbf{E}^-[2X\varphi(W^- + X + Y)] \\
&= 2\mathbf{E}^-[(1 - XY)\varphi'(\widetilde{W})] \\
&= \mathbf{E}^-[(2 - (\epsilon_{\pi(i)} - \epsilon_{\pi(j)})Y)\varphi'(\widetilde{W})].
\end{aligned}$$

If we define a_{ij} as in (63) above, then

$$a_{ij} = \begin{cases} 2 - (\epsilon_{\pi(i)} - \epsilon_{\pi(j)})Y & \text{if } S = S^-; \\ 0 & \text{if } S \neq S^-. \end{cases}$$

Finally we obtain

$$\mathbf{E}^-[(\epsilon_{\pi(i)} - \epsilon_{\pi(j)})\varphi(\widetilde{W})] = \mathbf{E}^-(a_{ij}\varphi'(\widetilde{W})] \tag{68}$$

and replace $\mathbf{E}^-$ by $\mathbf{E}$ by taking expectations on both sides. If we again apply Lemma 3.7 we obtain (61). □

Exercise 52 *Prove Lemma 3.5.*

Lemma 53 **(Lemma 3.9 in handout)** *For any n and any possible value of S_n we can construct $W_0, W_1, ..., W_n$ and $Z_0, Z_1, ..., Z_n$ where the Z_i's are jointly Gaussian with mean zero and*

$$Cov(Z_i, Z_j) = \frac{(i \vee j)(n - (i \vee j))}{n}$$

such that for all $\lambda \leq \lambda_0$,

$$\mathbf{E}[\exp(\lambda \max_{i \leq n} |W_i - Z_i|)] \leq \exp\left(C \log n + \frac{K\lambda^2 S_n^2}{n}\right) \tag{69}$$

where C, K, and λ_0 are universal constants.

The proof is by induction: use $\exp(x \vee y) \leq e^x + e^y$. *To be continued.*

Lecture 20

Lecture date: Oct 12, 2007 *Scribe: Tanya Gordeeva*

25 Proof of Lemma 3.9

Recall lemma 3.9 from the handout:

Lemma 54 (Lemma 3.9) *There exist universal constants C, K, and λ_0 such that, for any n and for any possible value of S_n, we can construct a centered Gaussian process $(Z_0, \ldots, Z_n)$ with $\mathrm{Cov}(Z_i, Z_j) = \frac{(i\wedge j)(n-(i\vee j))}{n}$ such that*

$$\forall \lambda < \lambda_0, \mathbf{E}\exp(\lambda \max_{i\le n} |W_i - Z_i|) \le \exp\left(C\log n + \frac{K\lambda^2 S_n^2}{n}\right)$$

where $W_i = S_i - \frac{iS_n}{n}$.

Sketch of the proof:

Assume the result is true for all $n' < n$. Fix k with $\frac{n}{3} \le k \le \frac{2n}{3}$.

1) Given S_n, construct (S, Z) where S has the distribution of S_k given S_n and Z follows $N(0, \frac{k(n-k)}{n})$ such that $\mathbf{E}\exp\left(\theta\left|\left(S - \frac{kS_n}{n}\right) - Z\right|\right) \le \exp\left(1 + \frac{c\theta^2 S_n^2}{n}\right)$ for all $\theta < \theta_0$. This is possible by lemma 3.8.

Now construct $(S_0, \ldots, S_k)$ as a simple random walk with $S_k = S$. Independently generate a random walk $(S'_0, \ldots, S'_{n-k})$ with $S'_{n-k} = S_n - S$.

If $(U_0, \ldots, U_n)$ is defined as

$$U_i = \begin{cases} S_i & \text{if } i \le k \\ S + S'_{i-k} & \text{if } n \ge i \ge k \end{cases}$$

then $(U_0, \ldots, U_n)$ is a SRW conditioned to be S_n at n.

2) The next step is to generate the Brownian bridge. Let $(Z_0, \ldots, Z_k)$ and $(Z'_0, \ldots, Z'_{n-k})$ be two independent Brownian bridges. Define $(Y_0, \ldots, Y_n)$ by

$$Y_i = \begin{cases} Z_i + \frac{i}{k}Z & \text{if } i \le k \\ Z'_{i-k} + \frac{n-i}{n-k}Z & \text{if } i > k \end{cases}$$

The candidates for the coupling will be $(U_0, \ldots, U_n)$ and $(Y_0, \ldots, Y_n)$.

Let $\mathbf{E}^*$ denote the conditional expectation given (S, Z). Apply induction to ensure that

$$\mathbf{E}^* \exp\left(\lambda \max_{i \le k} \left| \left(S_i - \frac{i}{k} S\right) - Z_i \right| \right) \le \exp\left(C \log k + \frac{K\lambda^2 S^2}{k}\right)$$

and

$$\mathbf{E}^* \exp\left(\lambda \max_{i \le n-k} \left| \left(S'_i - \frac{i}{n-k}(S_n - S)\right) - Z'_i \right| \right) \le \exp\left(C \log(n-k) + \frac{K\lambda^2 (S_n - S)^2}{n-k}\right)$$

Let $T_L = \max_{i \le k} \left|\left(S_i - \frac{i}{k} S\right) - Z_i\right|$ and $T_R = \max_{i \le n-k} \left|\left(S'_i - \frac{i}{n-k}(S_n - S)\right) - Z'_i\right|$. Let $T = |S - Z|$. Verify that $\max_{i \le n} |U_i - Y_i| \le \max\{T_L + T, T_R + T\}$.

So $\mathbf{E} \exp(\lambda \max_{i \le n} |U_i - Y_i|) \le \mathbf{E} \exp(\lambda(T_L + T)) + \mathbf{E} \exp(\lambda(T_R + T))$. Now,

$$\mathbf{E} \exp(\lambda(T_L + T)) = \mathbf{E}\left[\mathbf{E}^*(\exp(\lambda T_L)) e^{\lambda T}\right].$$

We bound the conditional expectation using induction, and apply the Cauchy-Schwarz inequality. So

$$\begin{aligned}\mathbf{E} \exp(\lambda(T_L + T)) &\le \mathbf{E}\left(\exp\left(C \log k + \frac{K\lambda^2 S^2}{k}\right) e^{\lambda T}\right) \\ &\le \exp(C \log k) \sqrt{\mathbf{E}\left(\exp\left(\frac{2K\lambda^2 S^2}{k}\right)\right) \mathbf{E}(e^{2\lambda T})}\end{aligned}$$

Note that this does not depend on Z. This is how this results from our use of the induction hypothesis: For each n, and each possible value s of S_n, let ρ^n_s be a joint density of the coupled process satisfying the induction hypothesis. Given $(S, Z) = (s, z)$, the distribution of $((S_0, \ldots, S_k), (Z_0, \ldots, Z_k))$ is simply ρ^n_s, which does not involve Z. So, conditional on S, $((S_0, \ldots, S_k), (Z_0, \ldots, Z_k))$ is independent of Z.

Now, by a previous lemma, we have that $\mathbf{E} e^{2\lambda T} \le \exp\left(1 + \frac{4c\lambda^2 S_n^2}{n}\right)$. By the induction hypothesis and lemma 3.6, we have

$$\mathbf{E} \exp\left(\frac{2K\lambda^2 S^2}{k}\right) \le \exp\left(1 + \frac{2K\lambda^2 S_n^2}{n} \frac{3}{4}\right)$$

If K is sufficiently large compared to c, we can guarantee that $\frac{3}{4} 2K + 4c \le 2K$. Choosing such K and combining,

$$\mathbf{E} \exp(\lambda(T_L + T)) \le \exp\left(C \log k + \frac{K\lambda^2 S_n^2}{n} + 1\right)$$

Similarly,

$$\mathbf{E}\exp(\lambda(T_R + T)) \le \exp\left(C\log(n-k) + \frac{K\lambda^2 S_n^2}{n} + 1\right)$$

Now $C\log k = C\log n - C\log\frac{n}{k} \le C\log n - C\log\frac{3}{2}$ since $\frac{n}{3} \le k \le \frac{2n}{3}$. Similarly, $C\log(n-k) \le C\log n - C\log\frac{3}{2}$.

Combining,

$$\mathbf{E}\exp(\lambda\max_{i\neq n}|U_i - Y_i|) \le 2\exp\left(C\log n - C\log\frac{3}{2} + \frac{K\lambda^2 S_n^2}{n} + 1\right)$$

Choose C such that $C\log\frac{3}{2} > \log 2 + 1$ to complete the proof.

This gives the coupling for the Brownian bridge, but it requires some more work to move to Brownian motion and move away from finite time. See handout for details.

Lecture 21

Lecture date: Oct 15, 2007 *Scribe: Chris Haulk*

Exercise

Let $X = (X_{i,j})_{1 \leq i < j \leq n}$ denote the Erdös-Renyi random graph $G(n, p)$. Let

$$f(X) = \#\text{triangles}(X) - E(\#\text{triangles}(X)).$$

Obtain concentration inequalities for $f(X)$ by explicitly constructing an antisymmetric F and an exchangeable pair (X, X') such that

$$E[F(X, X')|X] = f(X).$$

and using part (b) the concentration theorem (lecture 12, Sept 24). Hints: independently regenerate a random edge. f is boolean, so it's a polynomial: use polynomials. Thoughts: this will probably give sharp results when p is fixed, $n \uparrow \infty$, but will probably not give sharp results when $p \downarrow 0, n \uparrow \infty$. A serious research problem would be to use part (c) of the theorem and optimize over k to get the correct tail bound in this latter case[3].

26 Spin Glasses: Sherrington-Kirkpatrick Model

This lecture will cover definitions concerning the SK model of spin glasses and state results to be proven in later lectures. Let $g = (g_{ij})_{1 \leq i < j \leq N}$ be a collection of iid $N(0, 1)$ random variables. Collectively, these random variables are known as the *disorder*. Given the disorder, the spin configuration $\sigma = (\sigma_1, ..., \sigma_N) \in \{0, 1\}^N$ follows a Gibbs distribution with conditional density proportional to

$$\exp\left(\frac{\beta}{\sqrt{N}} \sum_{1 \leq i < j \leq N} g_{ij}\sigma_i\sigma_j + h\sum_{i=1}^{N} \sigma_i\right)$$

Let

$$Z_N(\beta, h, g) = \sum_{\sigma \in \{-1,1\}^N} \exp\left(\frac{\beta}{\sqrt{N}} \sum_{i<j} g_{ij}\sigma_i\sigma_j + h\sum_{i=1}^{N} \sigma_i\right)$$

[3]Optimal results are known only up to logarithmic factors, for details check Kim and Vu 2004, *Divide and conquer martingales and the number of triangles in a random graph*, or Janson, Oleszkiewicz, and Rucinski 2004, *Upper tails for subgraph counts in random graphs.*

be the normalizing constant for this Gibbs measure. For any function f of the spins σ and the disorder g, the *quenched law* (or distribution) of $f(g,\sigma)$ is the conditional distribution of f *given* g. The *annealed law* is the unconditional expectation of f. Usually, the quenched expectation of a function $f(g,\sigma)$ is denoted by $\langle f\rangle$:

$$\langle f\rangle = (Z_N(\beta,h,g))^{-1} \sum_{\sigma\in\{-1,1\}^N} f(g,\sigma)\exp\left(\frac{\beta}{\sqrt{N}}\sum_{i<j} g_{ij}\sigma_i\sigma_j + h\sum_{i=1}^N \sigma_i\right).$$

Then the annealed expectation is just $E\langle f\rangle$, which will often be shortened to $\nu(f)$.

Suppose we generate two independent configurations $\sigma^1, \sigma^2 \in \{-1,1\}^N$ from the same Gibbs measure. That is, sample g, and then take two (conditionally) independent samples σ^1, σ^2 from the distribution of spins given g. The overlap between σ^1 and σ^2 is defined to be

$$R_{1,2} = \frac{1}{N}\sum_{i=1}^N \sigma_i^1\sigma_i^2.$$

Let $p_N(\beta,h) = N^{-1}E(\log Z_N(\beta,h))$.

Theorem 55 *Replica-symmetric solution of the S-K model.*

There is $\beta_0 > 0$ such that $\forall\beta\in(0,\beta_0)$, $\forall h$

$$\lim_{N\to\infty} p_N(\beta,h) = \log(2) + E\log\cosh(\beta z\sqrt{q}+h)) + \frac{\beta^2(1-q)^2}{4}$$

where $z\sim N(0,1)$ and $q = q(\beta,h)$ is the unique solution of $q = E[\tanh^2(\beta z\sqrt{q}+h)]$.

$$\frac{\log Z_N}{N} - p_N(\beta,h) \xrightarrow{P} 0$$

(*Note that this is* not *equal to* $\lim_{N\to\infty}\log(\frac{E(Z_N)}{N})$.)

27 History Lesson

Sherrington and Kirkpatrick claimed that the statement above was true for *all* β, h, however it soon became clear that this is false. Talagrand proved the theorem above in 1998 (see also Shcherbina 1997). Parisi conjectured a "broken replica symmetry solution" - this was recently proven to be correct. If $h = 0$, the model is more tractable. In that case,

$$\lim_{N\to\infty} p_N(\beta,h) = \frac{\beta^2}{4} + \log(2) \qquad (0\le\beta<1)$$

and there is a phase transition at $\beta = 1$. This was derived by Aizenmann, Lebowitz and Ruelle in 1987. One outstanding and important conjecture is that the replica symmetric solution holds for all (β, h) satisfying

$$\beta^4 E\left[\frac{1}{\cosh^4(\beta z\sqrt{q} + h)}\right] < 1$$

where q and z are as before. The boundary of this region is known as the Almeida-Thouless line (AT line). Talagrand showed that the replica-symmetric solution is invalid beyond the AT line.

Lecture 22

Lecture date: Oct 17, 2007 *Scribe: Maximilian Kasy*

Repetition of the setup for the Sherrington Kirkpatrick model:
A so called disorder $(g_{i,j})_{i\leq j\leq N}$ is drawn from an i.i.d. $N(0,1)$ distribution, conditional on g, $\sigma = (\sigma_1, \ldots, \sigma_N)$ has density proportional to

$$\exp\left(\frac{\beta}{\sqrt{N}}\sum g_{ij}\sigma_i\sigma_j + h\sum \sigma_i\right)$$

Definition 56 (Overlap) *Pick σ^1, σ^2 independently from the Gibbs-measure (i.e. the conditional distribution of σ given g). The overlap is defined as the random variable*

$$R_{12} = \frac{1}{N}\sum_i \sigma_i^1\sigma_i^2$$

Proposition 57 $\exists\beta_0 > 0$*: if $\beta < \beta_0$ then "with high probability" $R_{12} \simeq q$ where $q = q(\beta, h)$ solves*

$$q = E[\tanh^2(\beta z\sqrt{q} + h)]$$

where $z \sim N(0,1)$. More precisely:

$$E\left\langle (R_{12} - q)^{2k}\right\rangle = \frac{C(k)}{N^k}\ \forall\, k, N$$

(Recall: $\langle\cdot\rangle$ denotes conditional expectation given g)

Corollary 58 *Let*

$$RS(\beta, h) := \log 2 + E[\log\cosh(\beta z\sqrt{(q)} + h)] + \frac{\beta^2(1-q)^2}{4}$$

where q is defined as above. Then, for $\beta \leq \beta_0$,

$$\frac{\log Z_N}{N} \to RS(\beta, h)$$

in probability as $N \to \infty$, where $Z_N(\beta, h, g)$ denotes the normalizing constant of the Gibbs measure.

The proof of this Corollary proceeds in two steps:

1. Show that
$$\frac{\log Z_N}{N} - E\left[\frac{\log Z_N}{N}\right] \to 0$$
in probability. This is the easier part.

2. Show that
$$E\left[\frac{\log Z_N}{N}\right] \to RS(\beta, h)$$
This latter step is in turn subdivided into two parts: Showing that the lim sup of the left hand side is bounded by the right hand side (which is somewhat easier, and holds for all β, h), and showing inversely that the lim inf is bigger than the right hand side, which only is true for a certain range of β, h.

28 The Gaussian Poincare inequality

1. In 1 dimension: If $Z \sim N(0,1)$ then for all continuous f we have $Varf(Z) \leq E[f'(Z)^2]$

2. If $Z_1, \ldots, Z_n$ i.i.d. $N(0,1)$, then $Varf(Z_1, \ldots, Z_n) \leq \sum E[\left(\frac{\partial f}{\partial z_i}\right)^2]$

Proof:
$1 \to 2$: If X, Y iid then $Varf(x) = \frac{1}{2}E[(f(x) - f(y))^2]$. By Efron-Stein
$$Varf(Z_1, \ldots, Z_n) \leq$$
$$\frac{1}{2}\sum E[(f(Z_1, \ldots, Z_n) - f(Z_1, \ldots, Z_{i-1}, Z_i', Z_{i+1}, \ldots, Z_n))^2] =$$
$$\sum E[Varf(Z_1, \ldots, Z_n|_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n)] \leq \sum E[\left(\frac{\partial f}{\partial z_i}\right)^2]$$
where the last inequality follows from 1.
1, by Stein's method: Given an absolutely continuous f find g such that $g'(x) - xg(x) = f(x) - Ef(Z)$. W.l.o.g. we assume $Ef(Z) = 0$. Then
$$Varf(Z) = Ef(Z)^2 = E((g'(Z) - Zg(Z))f(Z)) = -Eg(Z)f'(Z)$$
(The last equality follows from integration by parts). This implies by Cauchy Schwartz
$$Ef(Z)^2 \leq \sqrt{Ef'(Z)^2 Eg(Z)^2}$$
Now $f'(x) = \frac{\partial}{\partial x}(g' - xg) = g'' - xg' - g$ Hence
$$Ef(Z)^2 = -E(g(Z)(g''(Z) - Zg'(Z) - g(Z))) = E(g'(Z)^2 + g(Z)^2)$$

Thus $Eg(Z)^2 \leq Ef(Z)^2$. Combining we get the result. This generalizes: If $v : R \to R$ is strictly convex and X is a R.V. with density proportional to $\exp(-v(x))$, then

$$Var f(X) \leq E\left(\frac{f'(X)^2}{v''(X)}\right)$$

Proof: find a g such that $g' + \frac{v'}{v}g = f - Ef$.
Exercise: More generally, If $v : \mathbb{R}^n \to \mathbb{R}$ is strictly convex and X is a R.V. with density proportional to $\exp(-v(x))$, then for all absolutely continuous $f : \mathbb{R}^n \to \mathbb{R}$:

$$Var f(X) \leq E\left[(\nabla f(X))^t (Hess\, V(X))^{-1} \nabla f(X)\right]$$

A deep generalization of this is the so called Helffer-Sjöstrand machinery.

Now, $\frac{\log Z_N}{N}$ is a function of g, hence

$$Var\left(\frac{\log Z_N}{N}\right) \leq \sum_{i<j} E\left(\frac{\partial}{\partial g_{ij}} \frac{\log Z_N}{N}\right)^2$$

Now,

$$\frac{\partial}{\partial g_{ij}} \frac{\log Z_N}{N} = \frac{1}{N Z_N} \frac{\partial Z_N}{\partial g_{ij}} =$$

$$\frac{1}{N Z_N} \sum_{\sigma} \frac{\beta}{\sqrt{N}} \sigma_i \sigma_j \exp\!\left(\frac{\beta}{\sqrt{N}} \sum g_{ij} \sigma_i \sigma_j + h \sum \sigma_i\right) = \frac{\beta}{N^{3/2}} \left\langle \sigma_i \sigma_j \right\rangle.$$

Hence

$$\sum_{i<j} E\left(\frac{\partial}{\partial g_{ij}} \frac{\log Z_N}{N}\right)^2 = \frac{1}{N^3} E \sum_{i<j} \left\langle \sigma_i \sigma_j \right\rangle^2 \leq \frac{1}{2N}.$$

It follows that

$$Var\left(\frac{\log Z_N}{N}\right) \leq \frac{1}{2N}$$

This also shows

$$\begin{aligned} Var\left(\frac{\log Z_N}{N}\right) &\leq \frac{\beta^2}{2N^3} E \sum_{i<j} \left\langle \sigma_i \sigma_j \right\rangle^2) \\ &= \frac{\beta^2}{2N^3} E \left\langle \sum_{i<j} \sigma_i^1 \sigma_j^1 \sigma_i^2 \sigma_j^2 \right\rangle \\ &= \frac{\beta^2}{2N} E \left\langle R_{12}^2 \right\rangle \end{aligned}$$

where σ^1, σ^2 are i.i.d. draws from the Gibbs measure. We will show later that when $\beta < 1$ and $h = 0$, the above bound is of order N^{-2}.

Lecture 23

Lecture date: Oct. 19, 2007 *Scribe: Partha Dey*

Recall that, in the Sherrington Kirkpatrick model, the probability of a configuration $\boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in \{-1,+1\}^N$ is

$$\mathbf{P}(\boldsymbol{\sigma}) = Z_N^{-1} \exp\left(\frac{\beta}{\sqrt{N}} \sum_{i<j} g_{ij}\sigma_i\sigma_j + h\sum_{i=1}^N \sigma_i \right)$$

where $(g_{ij})_{1\le i<j\le N}$ are i.i.d. standard gaussian random variables and

$$Z_N = \sum_{\boldsymbol{\sigma}\in\{-1,+1\}^N} \exp\left(\frac{\beta}{\sqrt{N}} \sum_{i<j} g_{ij}\sigma_i\sigma_j + h\sum_{i=1}^N \sigma_i \right)$$

is the normalizing constant. Suppose $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2$ are i.i.d. configurations from this Gibbs measure given $\boldsymbol{g} = (g_{ij})_{i<j}$. The overlap between $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2$ is defined as

$$R_{12} = \frac{1}{N}\sum_{i=1}^N \sigma_i^1\sigma_i^2.$$

Suppose $\langle\cdot\rangle$ denotes conditional expectation w.r.t. the Gibbs measure given $\boldsymbol{g}$ and ν denotes unconditional expectation, *i.e.* $\nu(f) = \mathbf{E}\langle f\rangle$. Then we have the following result.

Theorem 59 $\exists\, \beta_0 > 0$ *such that for all* $\beta \in [0,\beta_0]$ *and for all* h

$$\frac{\log Z_N}{N} \to \log 2 + \mathbf{E}\log\cosh(\beta z\sqrt{q} + h) + \frac{\beta^2(1-q)^2}{4}$$

where q *satisfies* $q = \mathbf{E}\tanh^2(\beta z\sqrt{q}+h)$ *and* $z \sim N(0,1)$.

Idea of the proof: Choose any arbitary number $q \in [0,1]$. Consider the alternative Gibbs measure $\propto \exp(\sum_{i=1}^N(\beta z_i\sqrt{q}+h)\sigma_i)$ where $z_1, z_2, \ldots, z_N$ are i.i.d. N(0,1) random variables independent of $\boldsymbol{g}$. Let ν_0 be the unconditional law of this Gibbs measure. Note that σ_i's are independent under this Gibbs measure (both conditionally and unconditionally) and this measure is easier to handle. Also

$$\frac{\beta}{\sqrt{N}} \sum_{1\le i<j\le N} g_{ij}\sigma_i\sigma_j + h\sum_{i=1}^N \sigma_i = \sum_{i=1}^N \left(\frac{\beta}{2} l_i + h\right)\sigma_i$$

where $l_i = \frac{1}{\sqrt{N}} \sum_{j=1, j\neq i}^N g_{ij}\sigma_j$. The main idea is to show that for β sufficiently small, with a proper choice of q one can compare ν_0 and ν "in some sense". In the last lecture we proved that $N^{-1}(\log Z_N - \mathbf{E}\log Z_N) \to 0$ in probability. Today we'll prove that

$$\mathbf{E}\left(\frac{\log Z_N}{N}\right) \leq \log 2 + \mathbf{E}\log\cosh(\beta z\sqrt{q} + h) + \frac{\beta^2(1-q)^2}{4} \text{ for all } q \in [0,1], \beta \geq 0, h \in \mathbb{R}.$$

Lemma 60 (Gaussian Interpolation) *Suppose* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *and* $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ *are two centered gaussian random vectors independent of each other. Let* $F : \mathbb{R}^n \to \mathbb{R}$ *be a* C^2 *function and let*

$$\varphi(t) = \mathbf{E}F(\sqrt{t}\boldsymbol{X} + \sqrt{1-t}\boldsymbol{Y}).$$

Then we have

$$\varphi'(t) = \frac{1}{2}\sum_{i,j=1}^n (\mathbf{E}(X_iX_j) - \mathbf{E}(Y_iY_j)) \cdot \mathbf{E}\left(\frac{\partial^2 F}{\partial x_i \partial y_j}(\sqrt{t}\boldsymbol{X} + \sqrt{1-t}\boldsymbol{Y})\right).$$

In particular we have

$$\mathbf{E}F(\boldsymbol{X}) - \mathbf{E}F(\boldsymbol{Y}) = \int_0^1 \varphi'(t)dt.$$

Proof: Exercise. □

For each $\boldsymbol{\sigma} \in \{-1, +1\}^N$, let

$$u_{\boldsymbol{\sigma}} = \frac{\beta}{\sqrt{N}}\sum_{i<j} g_{ij}\sigma_i\sigma_j \text{ and } v_{\boldsymbol{\sigma}} = \beta\sqrt{q}\sum_{i=1}^N z_i\sigma_i.$$

Then the normalizing constants in the S-K model and in the alternative model are

$$Z_N = \sum_{\boldsymbol{\sigma}} \exp(u_{\boldsymbol{\sigma}} + h\sum \sigma_i) \quad \text{and} \quad Z_N^0 = \sum_{\boldsymbol{\sigma}} \exp(v_{\boldsymbol{\sigma}} + h\sum \sigma_i)$$

respectively. So if we define a function $Z : \mathbb{R}^{\{-1,+1\}^N} \to \mathbb{R}$ as

$$Z(\boldsymbol{x}) = \sum_{\boldsymbol{\sigma} \in \{-1,+1\}^N} w_{\boldsymbol{\sigma}} \exp(x_{\boldsymbol{\sigma}})$$

where $\boldsymbol{x} = (x_{\boldsymbol{\sigma}})_{\boldsymbol{\sigma}\in\{-1,+1\}^N}$ and $w_{\boldsymbol{\sigma}} = \exp(h\sum_{i=1}^N \sigma_i)$, we have $Z_N = Z(\boldsymbol{u})$, $Z_N^0 = Z(\boldsymbol{v})$ where $\boldsymbol{u} = \{u_{\boldsymbol{\sigma}}\}_{\boldsymbol{\sigma}\in\{-1,+1\}^N}$ and $\boldsymbol{v} = \{v_{\boldsymbol{\sigma}}\}_{\boldsymbol{\sigma}\in\{-1,+1\}^N}$. Let

$$F(\boldsymbol{x}) = \frac{\log Z(\boldsymbol{x})}{N} \text{ for } \boldsymbol{x} \in \mathbb{R}^{\{-1,+1\}^N}$$

and $\varphi(t) = \mathbf{E}F(\sqrt{t}\boldsymbol{u} + \sqrt{1-t}\boldsymbol{v})$. We are interested in

$$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt.$$

Clearly we have

$$\begin{aligned}
\frac{\partial F}{\partial \boldsymbol{x_\sigma}} &= \frac{\partial}{\partial \boldsymbol{x_\sigma}}\left(\frac{\log Z(\boldsymbol{x})}{N}\right) = \frac{1}{NZ(\boldsymbol{x})} w_{\boldsymbol{\sigma}} \exp(x_{\boldsymbol{\sigma}}) \\
\text{and } \frac{\partial^2 F}{\partial \boldsymbol{x_\tau} \partial \boldsymbol{x_\sigma}} &= -\frac{1}{N(Z(\boldsymbol{x}))^2} w_{\boldsymbol{\tau}} w_{\boldsymbol{\sigma}} \exp(x_{\boldsymbol{\sigma}} + x_{\boldsymbol{\tau}}) + \frac{1}{NZ(\boldsymbol{x})} w_{\boldsymbol{\sigma}} \exp(x_{\boldsymbol{\sigma}}) \cdot \mathbf{1}_{\{\boldsymbol{\sigma}=\boldsymbol{\tau}\}}.
\end{aligned}$$

Let $U(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \frac{1}{2}\mathbf{E}(u_{\boldsymbol{\sigma}^1} u_{\boldsymbol{\sigma}^2} - v_{\boldsymbol{\sigma}^1} v_{\boldsymbol{\sigma}^2})$. Then

$$\begin{aligned}
\varphi'(t) &= \sum_{\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2} U(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) \mathbf{E}\left(\frac{\partial^2 F}{\partial \boldsymbol{x}_{\boldsymbol{\sigma}^1} \partial \boldsymbol{x}_{\boldsymbol{\sigma}^2}} (\sqrt{t}\boldsymbol{u} + \sqrt{1-t}\boldsymbol{v})\right) \\
&= \frac{1}{N} \sum_{\boldsymbol{\sigma}} U(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \frac{w_{\boldsymbol{\sigma}} \exp(\sqrt{t} u_{\boldsymbol{\sigma}} + \sqrt{1-t} v_{\boldsymbol{\sigma}})}{Z_t} \\
&\quad - \frac{1}{N} \sum_{\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2} U(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) \frac{w_{\boldsymbol{\sigma}^1} w_{\boldsymbol{\sigma}^2} \exp(\sqrt{t}(u_{\boldsymbol{\sigma}^1} + u_{\boldsymbol{\sigma}^2}) + \sqrt{1-t}(v_{\boldsymbol{\sigma}^1} + v_{\boldsymbol{\sigma}^2}))}{Z_t^2}
\end{aligned}$$

where $Z_t = \sum_{\boldsymbol{\sigma}} w_{\boldsymbol{\sigma}} \exp(\sqrt{t} u_{\boldsymbol{\sigma}} + \sqrt{1-t} v_{\boldsymbol{\sigma}})$. For each $t \in [0,1]$ we have a gibbs measure $\propto \exp(\sqrt{t} u_{\boldsymbol{\sigma}} + \sqrt{1-t} v_{\boldsymbol{\sigma}} + h \sum \sigma_i)$ where $u_{\boldsymbol{\sigma}} = \frac{\beta}{\sqrt{N}} \sum_{i<j} g_{ij} \sigma_i \sigma_j$ and $v_{\boldsymbol{\sigma}} = \beta\sqrt{q} \sum_{i=1}^N z_i \sigma_i$. Let $\langle \cdot \rangle_t$ denote the expectation w.r.t. this gibbs measure. Let ν_t denote the unconditional expectation. Then

$$\varphi'(t) = \frac{1}{N}(\mathbf{E}\langle U(\boldsymbol{\sigma}, \boldsymbol{\sigma})\rangle_t - \mathbf{E}\langle U(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)\rangle_t).$$

Now

$$\begin{aligned}
U(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) &= \frac{\beta^2}{2N} \mathbf{E}\left((\sum_{i<j} g_{ij} \sigma_i^1 \sigma_j^1)(\sum_{i<j} g_{ij} \sigma_i^2 \sigma_j^2)\right) - \frac{\beta^2 q}{2} \mathbf{E}\left((\sum_{i=1}^N Z_i \sigma_i^1)(\sum_{i=1}^N Z_i \sigma_i^2)\right) \\
&= \frac{\beta^2}{2N} \sum_{i<j} \sigma_i^1 \sigma_i^2 \sigma_j^1 \sigma_j^2 - \frac{\beta^2 q}{2} \sum_{i=1}^N \sigma_i^1 \sigma_i^2 \\
&= \frac{\beta^2}{4N}\left((\sum_{i=1}^N \sigma_i^1 \sigma_i^2)^2 - N\right) - \frac{\beta^2 q}{2} \sum_{i=1}^N \sigma_i^1 \sigma_i^2 \\
&= \frac{\beta^2 N}{4}\left(R_{12}^2 - \frac{1}{N}\right) - \frac{\beta^2 q N}{2} R_{12} \\
\implies \frac{1}{N} U(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) &= \frac{\beta^2}{4}\left(R_{12}^2 - \frac{1}{N}\right) - \frac{\beta^2 q}{2} R_{12}.
\end{aligned}$$

Note that

$$\frac{1}{N}U(\boldsymbol{\sigma},\boldsymbol{\sigma}) = \frac{\beta^2}{4}\left(1-\frac{1}{N}\right) - \frac{\beta^2 q}{2}.$$

Now plugging in the values of $U(\boldsymbol{\sigma},\boldsymbol{\sigma}), U(\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2)$ we have

$$\varphi'(t) = \left(\frac{\beta^2}{4} - \frac{\beta^2 q}{2}\right) - \left(\frac{\beta^2}{4}\mathbf{E}\langle R_{12}^2\rangle_t + \frac{\beta^2 q}{2}\mathbf{E}\langle R_{12}\rangle_t\right) = -\frac{\beta^2}{4}\mathbf{E}\langle (R_{12}-q)^2\rangle_t + \frac{\beta^2}{4}(1-q)^2.$$

This gives, in particular,

$$\varphi(1) \le \varphi(0) + \frac{\beta^2(1-q)^2}{4} \quad \forall\ 0 \le q \le 1.$$

Now note that

$$\begin{aligned}
\varphi(0) &= \frac{1}{N}\mathbf{E}\log\Bigg(\sum_{\boldsymbol{\sigma}\in\{-1,+1\}^N}\prod_{i=1}^{N}\exp((\beta z_i\sqrt{q}+h)\sigma_i)\Bigg) \\
&= \frac{1}{N}\mathbf{E}\log\prod_{i=1}^{N}(\exp(\beta z_i\sqrt{q}+h)+\exp(-\beta z_i\sqrt{q}-h)) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\log\left(2\cosh(\beta z_i\sqrt{q}+h)\right) = \log 2 + \mathbf{E}\log\cosh(\beta z\sqrt{q}+h)
\end{aligned}$$

where $z \sim N(0,1)$. So for any $0 \le q \le 1$, we have

$$\mathbf{E}\left(\frac{\log Z_N}{N}\right) \le \log 2 + \mathbf{E}\log\cosh(\beta Z\sqrt{q}+h) + \frac{\beta^2(1-q)^2}{4}.$$

This inequality is called **Guerra's inequality** and this holds for all $\beta \ge 0, h \in \mathbb{R}$.

Exercise 61 *Prove that the R.H.S. of Guerra's inequality is minimized when*

$$q = \mathbf{E}\tanh^2(\beta z\sqrt{q}+h).$$

Lecture 24

Lecture date: October 22, 2007 *Scribe: Joel Mefford*

Continuing with the Sherrington-Kirkpatrick model.
$\forall q \in [0,1], \forall \beta \geq 0, \forall h \in \mathbb{R}$

$$\mathbf{E}\left(\frac{1}{N}\log Z_N\right) \leq \inf_q \left\{\log 2 + \mathbf{E}\log\cosh\left(\beta z\sqrt{q}+h\right) + \frac{\beta^2(1-q)^2}{4}\right\}, \tag{70}$$

where $z \sim N(0,1)$.

Exercise 62 *(From the previous lecture) Show that the right hand side of Equation 1 is minimized when*

$$q = \mathbf{E}\tanh^2\left(\beta z\sqrt{q}+h\right).$$

Now,

$$\varphi'(t) = -\frac{\beta^2}{4}\mathbf{E}\left\langle (R_{12}-q)^2\right\rangle_t + \frac{\beta^2}{4}(1-q)^2,$$

where $R_{12} = \frac{1}{N}\sum_1^N \sigma_i^1\sigma_i^2$. This implies that, $\forall 0 \leq q \leq 1$,

$$\frac{1}{N}\mathbf{E}\log Z_N = \varphi(1) \leq \varphi(0) + \frac{\beta^2(1-q)^2}{4},$$

and

$$\varphi(0) = \log 2 + \mathbf{E}\log\cosh\left(\beta z\sqrt{q}+h\right),$$

where $z \sim N(0,1)$.

Thus,

$$\frac{1}{N}\mathbf{E}\log Z_N \leq \inf_q \left\{\log 2 + \mathbf{E}\log\cosh\left(\beta z\sqrt{q}+h\right) + \frac{\beta^2(1-q)^2}{4}\right\}. \tag{71}$$

From exercise 1, the right hand side of this equation is minimized when $q = \mathbf{E}\tanh^2\left(\beta z\sqrt{q}+h\right)$.

Exercise 63 *Show that if $h > 0$, equation (2) has exactly one root.*

From the formula for $\varphi'(t)$, we see that approximate equality for eqation 2 holds if and only if q is such that $\mathbf{E}\left\langle (R_{12}-q)^2\right\rangle_t \approx 0$ for $0 \leq t \leq 1$.

Conversely, if $\mathbf{E}\left\langle (R_{12}-q)^2\right\rangle_t \approx 0, \forall 0 \leq t \leq 1$ holds, then we must have approximate equality in Equation 2. Therefore, q must satisfy $q = \mathbf{E}\tanh^2\left(\beta z\sqrt{q}+h\right)$.

Let us take $q = q(\beta, h)$ such that $q = \mathbf{E}\tanh^2\left(\beta z\sqrt{q}+h\right)$.

First, we observe that at $t=0$, the coordinates are independent under $\langle\cdot\rangle_0$.

$$\begin{aligned}\langle\sigma_i\rangle_0 &= 1\cdot\mathbf{P}(\sigma_i=1)+(-1)\cdot\mathbf{P}(\sigma_i=-1)\\ &= \tanh\left(\beta z_i\sqrt{2}+h\right)\\ \left\langle\sigma_i^1\sigma_i^2\right\rangle_0 &= \left\langle\sigma_i^1\right\rangle_0\left\langle\sigma_i^2\right\rangle_0\\ &= \langle\sigma_i\rangle_0^2\\ &= \tanh^2\left(\beta z_i\sqrt{q}+h\right)\end{aligned}$$

Here, σ^1 and σ^2 are i.i.d. from the Gibbs measure, and

$$R_{12} = \frac{1}{N}\sum_{i=1}^N \sigma_i^1\sigma_i^2 \approx \langle R_{12}\rangle_0 = \frac{1}{N}\sum_{i=1}^N \tanh^2\left(\beta z_i\sqrt{q}+h\right).$$

Since $z_1, z_2, \ldots, z_N \overset{i.i.d}{\sim} N(0,1)$,

$$\frac{1}{N}\sum_{i=1}^N \tanh^2\left(\beta z_i\sqrt{q}+h\right) \approx \mathbf{E}\tanh^2\left(\beta z_i\sqrt{q}+h\right) = q.$$

Thus, under $\langle\cdot\rangle_0$, we have $\mathbb{R}_{12}\approx q$ with high probability.

In fact, for $\lambda < \frac{1}{2}$,

$$\nu_0\left(\exp\left(\lambda N(R_{12}-q)^2\right)\right) \leq \frac{1}{\sqrt{1-2\lambda}}. \tag{72}$$

Exercise 64 *Derive equation (3).*

Latala's Proof of the concentration of the overlap

Theorem 65 *If $\beta < \frac{1}{2}$, for $2s < 1-4\beta^2$, $\nu = \nu_1$*

$$\nu\exp\left(sN(R_{12}-q)^2\right) \leq \frac{1}{\sqrt{1-2s-4\beta^2}}.$$

Thus,

$$\nu\left(R_{12}-q\right)^{2k} \leq \frac{(Ck)^k}{N^k}.$$

Proof: Take any function $f = f(\sigma^1,\sigma^2)$ of a pair of spin configurations.

Claim

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\nu_t(f) &\equiv \nu_t'(f) \\
&= \frac{N\beta^2}{2}\left[\nu_t\left((R_{12}-q)^2 f\right) - 2\nu_t\left((R_{34}-q)^2 f\right) - 2\nu\left((R_{23}-q)^2 f\right) + 3\nu_t\left((R_{34}-q)^2 f\right)\right]
\end{aligned}
$$

If $f \geq 0$ everywhere,

$$\frac{\mathrm{d}}{\mathrm{d}t}\nu_t(f) \leq \frac{N\beta^2}{2}\left[\nu_t\left((R_{12}-q)^2 f\right) + 3\nu_t\left((R_{34}-q)^2\right)\right].$$

Taking $f = (R_{12}-q)^{2k}$,

$$\nu_t'(R_{12}-q)^{2k} \quad \leq \quad \frac{N\beta^2}{2}\left[\nu_t\left(R_{12}-q\right)^{2k+2} + 3\nu_t\left((R_{34}-q)^2(R_{12}-q)^{2k}\right)\right].$$

Now,

$$
\begin{aligned}
\nu_t\left((R_{34}-q)^2(R_{12}-q)^{2k}\right) &\leq \left(\nu_t\left(R_{34}-q\right)^{2k+2}\right)^{\frac{k}{k+1}}\left(\nu_t\left(R_{12}-q\right)^{2k+2}\right)^{\frac{1}{k+1}} \\
&= \nu_t\left(R_{12}-q\right)^{2k+2}.
\end{aligned}
$$

Combining, we get

$$\nu_t'\left(R_{12}-q\right)^{2k} \quad \leq \quad 2N\beta^2\nu_t\left(R_{12}-q\right)^{2k+2}.$$

Multiplying both sides by $\frac{\lambda^k N^k}{k!}$ and summing over k, we have

$$\nu_t'\left(\exp\left(\lambda N(R_{12}-q)^2\right)\right) \leq 2N\beta^2\nu_t\left((R_{12}-q)^2 e^{\lambda N(R_{12}-q)^2}\right).$$

It follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\nu_t \exp\left(\left(\lambda - 2t\beta^2\right)N(R_{12}-q)^2\right) \leq 0.$$

In partiicular,

$$\nu_1\left(\exp\left\{\left(\lambda - 2\beta^2\right)N\left(R_{12}-q\right)\right\}\right) \leq \nu_0\left(e^{\lambda N(R_{12}-q)^2}\right).$$

So, for $\lambda = s + 2\beta^2 < \frac{1}{2}$, we get our results: For $\beta < \frac{1}{2}$, for $2s < 1 - 4\beta^2$,

$$\nu \exp\left(sN(R_{12}-q)^2\right) \leq \frac{1}{\sqrt{1-2s-4\beta^2}}.$$

□

Lecture 25

Lecture date: Oct 24, 2007 *Scribe: John Zhu*

Latala's Result: If $\beta < 1/2$ then for $2s < 1 - 4\beta^2$

$$\nu(\exp(sN(R_{1,2}-q)^2)) \leq \frac{1}{\sqrt{1-2s-4\beta^2}} \Rightarrow \nu((R_{1,2}-q)^{2k}) \leq \frac{(Ck)^k}{N^k} \text{ for all } k$$

so the overlap is concentrated.

Exercise 1: Using this result show that for any fixed p,

$$\mathbf{E}(\langle\sigma_1\sigma_2\cdots\sigma_p\rangle - \langle\sigma_1\rangle\langle\sigma_2\rangle\cdots\langle\sigma_p\rangle)^2 \leq K(p)\nu((R_{1,2}-q)^2) \leq \frac{K(p)}{N}$$

Exercise 2: The above expectation goes to 0 even if p grows with N. How fast can it grow?

The 1st exercise means that any collection of spins at p locations are approximately independent.

Hints for Exercise 1: use induction on p and note if $\sigma^1,\ldots,\sigma^4$ are 4 configurations then $R_{1,3} - R_{1,4} - R_{2,3} + R_{3,4} = \frac{(\sigma^1-\sigma^2)\cdot(\sigma^3-\sigma^4)}{N}$.

Exercise 3: (A Talagrand research problem) Show the total variation distribution distance between the joint law of $\sigma_1,\ldots\sigma_p$ and the product of marginals $\to 0$ as $N\to\infty$.

$\frac{d}{dt}\mathbf{E}\frac{\log Z_t}{N} = -\frac{\beta^2}{4}\nu_t((R_{1,2}-q)^2) + \frac{\beta^2}{4}(1-q)^2$. Since $\nu_t((R_{1,2}-q)^2) = O(\frac{1}{N})$ $\forall\ 0 \leq t \leq 1$ thus,

$$\mathbf{E}\frac{\log Z_N}{N} = \varphi(1) = \varphi(0) + \frac{\beta^2(1-q)^2)}{4} + O(\frac{1}{N}) = \log 2 + \mathbf{E}(\log\cosh(\beta Z\sqrt{q}+h))$$

Thouless - Anderson - Palmer Equations: The random quantities $\langle\sigma_1\rangle,\langle\sigma_2\rangle,\ldots\langle\sigma_N\rangle$ satisfy an approximate system of equations

$$\langle\sigma_i\rangle \approx \tanh(\frac{\beta}{\sqrt{N}}\sum_{j=1,j\neq i}^{N} g_{ij}\langle\sigma_j\rangle + h - \beta^2(1-q)\langle\sigma_i\rangle) \quad i=1,2,\ldots,N$$

Talagrand in 2003 gave the first rigorous proof.

Suppose $\beta < 1/2, h = 0$. Then $q = 0$, and so $\mathbf{E}\langle R_{1,2}\rangle \le \frac{c}{N}$. Let $l_i = \frac{1}{\sqrt{N}}\sum_{j=1,j\neq i}^N g_{ij}\sigma_j$ be the "local field" at site i. We will look at the annealed (i.e. unconditional) distribution of l_1.

Take any smooth function $f : \mathbb{R} \to \mathbb{R}$

$$\nu(l_1 f(l_1)) = \frac{1}{\sqrt{N}}\sum_{j=2}^N \mathbf{E}[g_{1j}\langle \sigma_j f(l_1)\rangle] = \frac{1}{\sqrt{N}}\sum_{j=2}^N \mathbf{E}[\frac{\partial}{\partial g_{1j}}\langle \sigma_j f(l_1)\rangle]$$

$$\frac{\partial}{\partial g_{1j}}\langle \sigma_j f(l_1)\rangle = \frac{\partial}{\partial g_{1j}} \frac{\sum_\sigma \sigma_j f(l_1(\sigma)) \exp(\frac{\beta}{\sqrt{N}}\sum_{r<s} g_{rs}\sigma_r\sigma_s)}{\sum_r \exp(\frac{\beta}{\sqrt{N}}\sum_{r<s} g_{rs}\sigma_r\sigma_s)}$$

$$= \frac{\sum_\sigma \left[\sigma_j f'(l_1(\sigma))\frac{\sigma_j}{\sqrt{N}}\exp(\cdots) + \sigma_j f(l_1(\sigma))\frac{\beta}{\sqrt{N}}\sigma_1\sigma_j\exp(\cdots)\right]}{\sum_\sigma \exp(\cdots)}$$

$$-\frac{\sum_\sigma f(l_1(\sigma))\exp(\cdots)}{(\sum \exp(\cdots))^2}\left(\sum_r \frac{\beta}{\sqrt{N}}\sigma_1\sigma_j\exp(\cdots)\right)$$

$$= \frac{\langle f'(l_1)\rangle}{\sqrt{N}} + \frac{\beta}{\sqrt{N}}\langle \sigma_1 f(l_1)\rangle - \frac{\beta}{\sqrt{N}}\langle \sigma_j f(l_1)\rangle\langle\sigma_1\sigma_j\rangle$$

Thus

$$\nu(l_1 f(l_1)) = \frac{N-1}{N}\nu(f'(l_1)) - \frac{\beta(N-1)}{N}\nu(\sigma_1 f(l_1)) - \frac{\beta}{N}\sum_{j=2} N\,\mathbf{E}\langle\sigma_j f(l_1)\rangle\langle\sigma_1\sigma_j\rangle$$

and

$$\langle \sigma_j f(l_1)\rangle\langle\sigma_1\sigma_j\rangle = \langle f(l_1(\sigma^1))\sigma_j^1\sigma_j^2\sigma_1^2\rangle.$$

So

$$\frac{1}{N}\sum_{j=2}^N \langle\sigma_j f(l_1)\rangle\langle\sigma_1\sigma_j\rangle = \langle f(l_1(\sigma^1))\sigma_1^2 R_{1,2}\rangle + O(\frac{1}{N}).$$

Exercise 4: Under $\langle\ \rangle$, what is the conditional expectation of σ_1 given $\sigma_2, \ldots \sigma_N$?

l_1 is a function of $\sigma_2, \ldots \sigma_N$, so $\langle \sigma_1 f(l_1)\rangle = \langle \tanh(\beta l_1) f(l_1)\rangle$. Combining the steps we get,

$$\nu\left(f'(l_1) - (l_1 - \beta\tanh(\beta l_1))\, f(l_1)\right) = O(\frac{1}{\sqrt{N}})$$

For any suitable probability density ρ, if $X \sim \rho$ then for any suitable f we have $\mathbf{E}(f'(X) + \frac{\rho'(X)}{\rho(X)}f(X)) = 0$ (integration by parts), and the converse is also true. Thus, the annealed distribution of l_1 must be close to the distribution with density ρ that satisfies

$$\frac{d}{dx}\log\rho(x) = -(x - \beta\tanh\beta x).$$

This implies

$$\log \rho(x) = \text{ const } - \frac{x^2}{2} + \log \cosh(\beta x)$$

$$\Rightarrow \rho(x) = \text{ Const } \cosh(\beta x) e^{-x^2/2} = \text{ Const } (e^{-(x-\beta)^2/2} + e^{-(x+\beta)^2/2}).$$

Thus, as $N \to \infty$, the annealed distribution of l_1 tends to the symmetric mixture of $N(\beta, 1)$ and $N(-\beta, 1)$.

Lecture 26

Lecture date: October 26, 2007 *Scribe: Richard Liang*

29 Annealed CLT for the Hamiltonian of the Sherrington-Kirkpatrick model

We begin by recalling the setting of the previous lecture: let $\beta < 1/2$, $h = 0$, and

$$l_i = \frac{1}{\sqrt{N}} \sum_{\substack{j=1 \\ j \neq i}}^{N} g_{ij} \sigma_j.$$

We saw that the annealed distribution of l_1 approaches

$$\frac{1}{2} N(\beta, 1) + \frac{1}{2} N(-\beta, 1)$$

as $N \to \infty$. (In fact this can be extended to the regime $\beta < 1$, $h = 0$; we'll return to this later.) We also saw that for any smooth function f,

$$\nu\big(f'(l_1) - (l_1 - \beta \tanh(\beta l_1))\, f(l_1)\big) \to 0$$

(here ν is the annealed distribution).

Exercise 66 *Develop Stein's method for*

$$\frac{1}{2} N(\beta, 1) + \frac{1}{2} N(-\beta, 1)$$

and get a total variation bound for the above CLT.

The next exercise should have been given earlier, but is an important result.

Exercise 67 *Recall Lemma 3.4 from the proof of the KMT embedding: if W is a random variable such that $\mathbf{E}W = 0$ and $\mathbf{E}\big[W^2\big] < \infty$, and T is a random variable such that*

$$\mathbf{E}[W\varphi(W)] = \mathbf{E}\big[T\varphi'(W)\big] \text{ for all } \varphi \text{ Lipschitz,} \tag{73}$$

then for all $\sigma^2 > 0$, there exists a pair of random variables (W', Z) on the same probability space such that W' is a version of W, $Z \sim N(0, \sigma^2)$ and

$$\mathbf{E}\big[\exp(\theta\,|W' - Z|)\big] \le 2\mathbf{E}\left[\exp\left(\frac{2\theta^2\,(T-\sigma^2)^2}{\sigma^2}\right)\right].$$

Prove a multivariate version of this result.

((73) can be alternately written as $\mathbf{E}[W\varphi'(W)] = \mathbf{E}[T\varphi''(W)]$ for all φ. This generalizes to $\mathbf{E}[W \cdot \nabla\varphi(W)] = \mathbf{E}[Tr(T\,Hess(\varphi(W)))]$, where $W = (W_1, \dots, W_d)$, $\varphi : \mathbb{R}^d \to \mathbb{R}$, and $T = (T_{ij})_{i,j\le d}$ is a positive definite random matrix. Replace σ^2 by the positive definite matrix $\Sigma = (\sigma_{ij})_{i,j=1}^d$ and replace θ by $\theta \in \mathbb{R}^d$.)

In the regime $\beta < 1/2$, $h = 0$, we'll compute the limiting annealed distribution of the Hamiltonian

$$\sum_{i<j} g_{ij}\sigma_i\sigma_j.$$

Specifically, we will consider

$$H = \frac{1}{N}\sum_{i<j} g_{ij}\sigma_i\sigma_j - \frac{\sqrt{N}\beta}{2}.$$

The appropriate scaling and centering come from the following ("very simple") exercise.

Exercise 68 *Let*

$$p_N(\beta) = \mathbf{E}\left[\frac{\log Z_N(\beta)}{N}\right].$$

We saw that $p_N(\beta) \to \log 2 + \beta^2/4$. Show

$$p_N'(\beta) = \mathbf{E}\left\langle \frac{1}{N^{3/2}}\sum_{i<j} g_{ij}\sigma_i\sigma_j \right\rangle$$
$$p_N''(\beta) = \mathbf{E}\mathrm{Var}\left(\frac{1}{N}\sum_{i<j} g_{ij}\sigma_i\sigma_j \,\middle|\, g\right).$$

Take any smooth f: we have

$$\begin{aligned}\nu(Hf(H)) &= \mathbf{E}\left\langle \frac{1}{N}\sum_{i<j} g_{ij}\sigma_i\sigma_j f(H)\right\rangle - \frac{\sqrt{N}\beta}{2}\mathbf{E}\langle f(H)\rangle \\ &= \frac{1}{N}\sum_{1\le i<j\le N}\mathbf{E}[g_{ij}\langle\sigma_i\sigma_j f(H)\rangle] - \frac{\sqrt{N}\beta}{2}\mathbf{E}\langle f(H)\rangle. \qquad (74)\end{aligned}$$

We use integration by parts to deal with the first term in (74):

$$\begin{aligned}
&\frac{1}{N}\sum_{1\le i<j\le N}\mathbf{E}[g_{ij}\langle\sigma_i\sigma_j f(H)\rangle]\\
&=\frac{1}{N}\sum_{i<j}\mathbf{E}\left[\frac{\partial}{\partial g_{ij}}\langle\sigma_i\sigma_j f(H)\rangle\right]\\
&=\frac{1}{N^2}\sum_{i<j}\mathbf{E}\left\langle f'(H)\right\rangle-\frac{\beta}{N^{3/2}}\sum_{i<j}\mathbf{E}[\langle f(H)\sigma_i\sigma_j\rangle\langle\sigma_i\sigma_j\rangle]+\frac{\beta}{N^{3/2}}\sum_{i<j}\mathbf{E}\langle f(H)\rangle. \qquad (75)
\end{aligned}$$

The first term in (75) equals

$$\frac{N(N-1)}{2N^2}\nu\big(f'(H)\big)\approx\frac{1}{2}\nu\big(f'(H)\big)$$

and similarly the third is approximately

$$\frac{\sqrt{N}\beta}{2}\nu(f(H)),$$

which takes care of the second term in (74). The second term of (75) is approximately

$$\frac{\beta}{2N^{3/2}}\mathbf{E}\left[\sum_{i,j=1}^{N}\langle f(H)\sigma_i\sigma_j\rangle\langle\sigma_i\sigma_j\rangle\right].$$

We can write the summand as $\left\langle f\big(H\big(g,\sigma^1\big)\big)\,\sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2\right\rangle$, where σ^1 and σ^2 are two spin configurations independently drawn from the quenched distribution $\langle\cdot\rangle$. Doing this, the above gives

$$\begin{aligned}
&\frac{\beta\sqrt{N}}{2}\mathbf{E}\left\langle f\big(H\big(g,\sigma^1\big)\big)\,\frac{1}{N^2}\sum_{i,j=1}^{N}\sigma_i^1\sigma_j^1\sigma_i^2\sigma_j^2\right\rangle\\
&=\frac{\beta\sqrt{N}}{2}\mathbf{E}\left\langle f\big(H\big(g,\sigma^1\big)\big)\,R_{12}^2\right\rangle.
\end{aligned}$$

If f is bounded, then

$$\left|\mathbf{E}\left\langle f(H)R_{12}^2\right\rangle\right|\le|f|_\infty\mathbf{E}\left\langle R_{12}^2\right\rangle\le\frac{C|f|_\infty}{N}.$$

Putting these above calculations back into (74), we get

$$\nu(Hf(H))=\frac{1}{2}\nu\big(f'(H)\big)+O\bigg(\frac{1}{\sqrt{N}}\bigg),$$

which shows that $H\Longrightarrow N(0,1/2)$.

Exercise 69 *Prove an annealed CLT for* $\sum_{i<j}g_{ij}\sigma_i\sigma_j$ *when* $\beta<1/2,h\neq 0$.

30 Quenched laws

Example 70 *For the Hamiltonian, we showed that for any f,*

$$\nu\left(\frac{1}{2}f'(H) - Hf(H)\right) = \mathbf{E}\left\langle \frac{1}{2}f'(H) - Hf(H)\right\rangle \to 0.$$

For a quenched CLT, we have to show that for all f,

$$\left\langle \frac{1}{2}f'(H) - Hf(H)\right\rangle \xrightarrow{p} 0.$$

Exercise 71 *Suppose that for "all" f,*

$$\left\langle \frac{1}{2}f'(H) - Hf(H)\right\rangle \xrightarrow{p} 0.$$

Show that "for all" φ, $\langle \varphi(H)\rangle \xrightarrow{p} \mathbf{E}[\varphi(Z)]$, where $Z \sim N(0, 1/2)$.

Exercise 72 *Suppose Exercise 71 holds. Let μ_N denote the (random) distribution of H under the Gibbs measure (that is, the "quenched" or conditional distribution given g). Show that $\mu_N \xrightarrow{p} N(0, 1/2)$ on the space of probability measures.*

STAT C206A / MATH C223A : Stein's method and applications 1

Lecture 27

Lecture date: Oct 29, 2007 *Scribe: Arnab Sen*

In the previous lectures, for the SK model with $h = 0, \beta < 1/2$ (though the result actually holds for $h = 0, \beta < 1$), we proved that the annealed law of l_1 converges to the mixture of Gaussian $1/2N(\beta, 1) + 1/2N(-\beta, 1)$ by showing,

$$\nu\left(f'(l_1) - (l_1 - \beta\tanh(\beta l_1))f(l_1)\right) \overset{N\to\infty}{\to} 0 \quad \text{for every smooth } f.$$

We may go even further and can actually show by Stein's method that

$$TV\left(\mathcal{L}(l_1), \mathcal{L}(1/2N(\beta,1) + 1/2N(-\beta,1)\right) \le c/\sqrt{N}$$

where TV stands for the total variation distance. Now, let $\nu(\cdot|\mathbf{g})$ denote the quenched distribution of l_1, i.e. the conditional distribution of l_1 given $\mathbf{g} = (g_{ij})$. We want to show

$$\nu(\cdot|\mathbf{g}) \to 1/2N(\beta,1) + 1/2N(-\beta,1) \quad \text{in probability.}$$

In other words, the random measure $\nu(\cdot|\mathbf{g})$ converges in probability to nonrandom probability measure $1/2N(\beta,1)+1/2N(-\beta,1)$ in the space of all probability measures (equipped with metric for convergence in distribution).

It suffices to show for all bounded measurable h,

$$\langle h(l_1)\rangle \overset{P}{\to} \mathbf{E}h(Z) \quad \text{where } Z \sim 1/2N(\beta,1) + 1/2N(-\beta,1).$$

By standard Stein method arguments, it is enough to prove

$$\langle f'(l_1) - (l_1 - \beta\tanh(\beta l_1))f(l_1)\rangle \overset{P}{\to} 0, \tag{76}$$

for all nice f's which come as a bounded solution of the differential equations $f'(x) - (x - \tanh(\beta x))f(x) = h(x) - Eh(Z)$ for bounded h.

We pause for a moment to remark that showing (76) is a rather delicate problem as many of the standard tools, e.g. Poincaré inequality, fail in this case.

Before going to the proof, let us have a quick recap of what we did in the annealed case.

We started with

$$\langle l_1 f(l_1)\rangle = N^{-1/2}\sum_{j=2}^{N} g_{1j}\langle\sigma_j f(l_1)\rangle = \sum_{j=2}^{N} g_{1j}h_j \text{ where } h_j := N^{-1/2}\langle\sigma_j f(l_1)\rangle.$$

Using integration by parts, we had that

$$\mathbf{E}\sum_{j=2}^{N} g_{1j}h_j = \mathbf{E}\sum_{j=2}^{N} \frac{\partial h_j}{\partial g_{1j}}. \tag{77}$$

It turned out, using $\langle R_{12}^2 \rangle \approx 0$, that

$$\sum_{j=2}^{N} \frac{\partial h_j}{\partial g_{1j}} \approx \langle f'(l_1) + \beta \tanh(\beta l_1) f(l_1) \rangle.$$

For the quenched CLT, we would like to show that (77) holds approximately even without the expectation, i.e.

$$\sum_{j=2}^{N} g_{1j}h_j \approx \sum_{j=2}^{N} \frac{\partial h_j}{\partial g_{1j}} \quad \text{with high probability.}$$

So, we will need the following approximation lemma.

Lemma 73 (approximation lemma) *Suppose* $g_1, g_2, \ldots, g_m$ *are i.i.d.* $N(0,1)$ *and* $h_1, h_2, \ldots, h_m$ *are functions of* $g = (g_1, g_2, \ldots, g_m)$*. Then*

$$\mathbf{E}\left(\sum_{j=1}^{m} g_j h_j - \sum_{j=1}^{m} \frac{\partial h_j}{\partial g_j}\right)^2 = \sum_{j=1}^{m} \mathbf{E}h_j^2 + \sum_{j=1}^{m}\sum_{k=1}^{m} \mathbf{E}\frac{\partial h_j}{\partial g_k}\frac{\partial h_k}{\partial g_j}. \tag{78}$$

Proof. Let $u = \sum_{j=1}^{m} g_j h_j - \sum_{j=1}^{m} \frac{\partial h_j}{\partial g_j}$. Then $\frac{\partial u}{\partial g_j} = h_j + \sum_{k=1}^{m} g_k \frac{\partial h_k}{\partial g_j} - \sum_{k=1}^{m} \frac{\partial^2 h_k}{\partial g_j \partial g_k}$.

$$\begin{aligned}
L.H.S. = \mathbf{E}u^2 &= \mathbf{E}u\left(\sum_{j=1}^{m} g_j h_j - \sum_{j=1}^{m} \frac{\partial h_j}{\partial g_j}\right)\\
&= \sum_{j=1}^{m} \mathbf{E}u g_j h_j - \sum_{j=1}^{m} \mathbf{E}u\frac{\partial h_j}{\partial g_j}\\
&= \sum_{j=1}^{m} \mathbf{E}u\frac{\partial h_j}{\partial g_j} + \sum_{j=1}^{m} \mathbf{E}h_j\frac{\partial u}{\partial g_j} - \sum_{j=1}^{m} \mathbf{E}u\frac{\partial h_j}{\partial g_j} \quad \text{by integration by parts}\\
&= \sum_{j=1}^{m} \mathbf{E}h_j\frac{\partial u}{\partial g_j}\\
&= \sum_{j=1}^{m} \mathbf{E}h_j^2 + \sum_{j=1}^{m}\sum_{k=1}^{m} \mathbf{E}h_j g_k\frac{\partial h_k}{\partial g_j} - \sum_{j=1}^{m}\sum_{k=1}^{m} \mathbf{E}h_j\frac{\partial^2 h_k}{\partial g_j \partial g_k}\\
&= \sum_{j=1}^{m} \mathbf{E}h_j^2 + \sum_{j=1}^{m}\sum_{k=1}^{m} \mathbf{E}\frac{\partial h_j}{\partial g_k}\frac{\partial h_k}{\partial g_j}.
\end{aligned}$$

where the last step follows from integration by parts on the second term. □

Estimation of R.H.S. of (78):

$$\begin{aligned}\sum_{j=2}^{N}\mathbf{E}h_j^2 &= N^{-1}\mathbf{E}\sum_{j=2}^{N}\langle\sigma_j f(l_1)\rangle^2\\ &= N^{-1}\mathbf{E}\sum_{j=2}^{N}\langle f(l_1(\sigma^1)f(l_1(\sigma^2)\sigma_j^1\sigma_j^2\rangle \quad \text{where } \sigma^1,\sigma^2 \overset{ind}{\sim} \text{ Gibbs measure given } \mathbf{g}\\ &= \mathbf{E}\langle f(l_1(\sigma^1)f(l_1(\sigma^2)R_{12}\rangle + O(\frac{1}{N})\\ &= O(\frac{1}{\sqrt{N}}) \quad \text{in the high temperature phase.}\end{aligned}$$

Lemma 74 *In SK model (any h), for any function $v = v(\mathbf{g},\sigma)$ of $\mathbf{g},\sigma$,*

$$\frac{\partial}{\partial g_{ij}}\langle v\rangle = \langle\frac{\partial v}{\partial g_{ij}}\rangle + \frac{\beta}{\sqrt{N}}\langle v\sigma_i\sigma_j\rangle - \frac{\beta}{\sqrt{N}}\langle v\rangle\langle\sigma_i\sigma_j\rangle.$$

Proof. Exercise. □

Using lemma 74, we have

$$\frac{\partial h_k}{\partial g_{1j}} = \frac{1}{N}\langle f'(l_1)\sigma_k\sigma_j\rangle + \frac{\beta}{N}\langle f(l_1)\sigma_1\sigma_k\sigma_j\rangle - \frac{\beta}{N}\langle f(l_1)\sigma_k\rangle\langle\sigma_1\sigma_j\rangle.$$

Thus,

$$\frac{\partial h_k}{\partial g_{1j}} = \frac{1}{N}\langle\sigma_j v_k\rangle = \frac{1}{N}\langle\sigma_k w_j\rangle,$$

where $v_k = v_k(\mathbf{g},\sigma) := f'(l_1)\sigma_k + \beta f(l_1)\sigma_1\sigma_k - \beta\langle f(l_1)\sigma_k\rangle\sigma_1$ and $w_j = w_j(\mathbf{g},\sigma) := f'(l_1)\sigma_j + \beta f(l_1)\sigma_1\sigma_j - \beta f(l_1)\langle\sigma_1\sigma_j\rangle$.

$$\begin{aligned}\sum_{j=2}^{N}\sum_{k=2}^{N}\mathbf{E}\frac{\partial h_j}{\partial g_{1k}}\frac{\partial h_k}{\partial g_{1j}} &= \frac{1}{N^2}\sum_{j=2}^{N}\sum_{k=2}^{N}\mathbf{E}\langle v_j(\mathbf{g},\sigma^1)w_j(\mathbf{g},\sigma^2)\sigma_k^1\sigma_k^2\rangle\\ &= \frac{1}{N}\sum_{j=2}^{N}\mathbf{E}\langle v_j(\mathbf{g},\sigma^1)w_j(\mathbf{g},\sigma^2)R_{12}\rangle + O(\frac{1}{N})\\ &= O(\frac{1}{\sqrt{N}}) \text{ since } v_j, w_j \text{ bounded, } \mathbf{E}|\langle R_{12}\rangle| = O(\frac{1}{\sqrt{N}}) \text{ for } \beta < 1/2, h = 0.\end{aligned}$$

This completes the proof.

Exercise 75 *Prove the quenched CLT for the hamiltonian in $\beta < 1/2, h = 0$.*

Lecture 28

Lecture date: Oct 31, 2007 *Scribe: Anand Sarwate*

31 A recap

For $\beta < 1/2$ in the Sherrington-Kirkpatric (S-K) model, we showed a bound on the overlap R_{12}:

$$\mathbf{E}\left\langle (R_{12} - q)^{2k} \right\rangle \leq \frac{(Ck)^k}{N^k} , \tag{79}$$

where $q = \mathbf{E}[\tanh^2(\beta Z \sqrt{q} + h)]$ and $Z \sim \mathcal{N}(0,1)$. This means that the overlap is concentrated. When $h = 0$ this implies that

$$\mathbf{E}\left\langle R_{12}^{2k} \right\rangle \leq \frac{(Ck)^k}{N^k} , \tag{80}$$

so R_{12} is close to 0 in this case. This result was crucial in showing that for $h = 0$ the quantity

$$l_1 = \frac{1}{\sqrt{N}} \sum_{j=2}^{N} g_{1j} \sigma_j \tag{81}$$

has a limiting annealed distribution

$$\frac{1}{2}\mathcal{N}(\beta, 1) + \frac{1}{2}\mathcal{N}(-\beta, 1) . \tag{82}$$

We also proved that the quenched distribution of l_1 converges in probability to (82) by showing that

$$\left\langle f'(l_1) - (l_1 - \beta \tanh(\beta l_1)) f(l_1) \right\rangle \xrightarrow{P} 0 . \tag{83}$$

Note that this convergence is in probability on the space of probability measures.

Finally, we also found a CLT for the Hamiltonian $\sum g_{ij}\sigma_i\sigma_j$ when $h = 0$. When $h \neq 0$ the limiting distribution of the Hamiltonian is not known.

32 The TAP equations

Today we will start looking at the Thouless-Anderson-Palmer (TAP) equations, which are a collection of self-consistent equations for the quenched average value for $i = 1, 2, \ldots, N$:

$$\langle \sigma_i \rangle \approx \tanh\left(\frac{\beta}{\sqrt{N}} \sum_{j=1, j\neq i}^{N} g_{ij} \langle \sigma_j \rangle + h - \beta^2 (1-q) \langle \sigma_i \rangle \right) \tag{84}$$

Furthermore, it is true that

$$\langle \sigma_i \rangle \xrightarrow{d} \tanh(\beta z \sqrt{q} + h) \tag{85}$$

where $z \sim \mathcal{N}(0,1)$. Moreover, $\langle \sigma_1 \rangle, \langle \sigma_2 \rangle, \ldots, \langle \sigma_p \rangle$ are asymptotically independent for fixed p as $N \to \infty$.

The concentration of the overlaps implies that

$$\langle \sigma_1 \sigma_j \rangle \cong \langle \sigma_i \rangle \langle \sigma_j \rangle \ , \tag{86}$$

which in turn implies

$$\frac{1}{N} \sum_{i=1}^{N} \sigma_i \cong \left\langle \frac{1}{N} \sum_{i=1}^{N} \sigma_i \right\rangle \tag{87}$$

$$= \frac{1}{N} \langle \sigma_i \rangle \tag{88}$$

$$\xrightarrow{P} \mathbf{E}[\tanh(\beta z \sqrt{q} + h)] \ . \tag{89}$$

Similarly, since $R_{12} \to q$,

$$R_{12} \cong \langle R_{12} \rangle \tag{90}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \langle \sigma_i \rangle^2 \tag{91}$$

$$\xrightarrow{P} \mathbf{E}[\tanh^2(\beta z \sqrt{q} + h)] \ . \tag{92}$$

This shows why q must satisfy

$$q = \mathbf{E}[\tanh^2(\beta z \sqrt{q} + h)] \ . \tag{93}$$

For simplicity of notation, let us define

$$r_i = \frac{1}{\sqrt{N}} \sum_{j=1, j\neq i}^{N} g_{ij} \langle \sigma_j \rangle - \beta(1-q) \langle \sigma_i \rangle \tag{94}$$

so that the TAP equations say

$$\langle \sigma_i \rangle \cong \tanh(\beta r_i + h) \ . \tag{95}$$

Note that r_i is a function of g only. It can be shown that $r_i \xrightarrow{d} \mathcal{N}(0, q)$.

33 A sketch of the proof

First, note that the conditional expectation of σ_1 given $\sigma_2, \ldots, \sigma_N$ is just $\tanh(\beta l_1 + h)$, so

$$\langle \sigma_1 \rangle = \langle \tanh(\beta l_1 + h) \rangle \ . \tag{96}$$

Now the goal is to approximate the distribution of l_1.

We first reparameterize Gaussian mixtures. Given a, b, μ, σ^2, let ψ_{a,b,μ,σ^2} denote the probability density on $\mathbb{R}$ proportional to

$$\cosh(ax+b) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \ . \tag{97}$$

Exercise Show that ψ_{a,b,μ,σ^2} is the same as $p\varphi_{\mu_1,\sigma^2} + (1-p)\varphi_{\mu_2,\sigma^2}$, where φ_{μ,σ^2} is the density $\mathcal{N}(\mu, \sigma^2)$, $\mu_1 = \mu + a\sigma^2$, $\mu_2 = \mu - a\sigma^2$, and

$$p = \frac{\exp(a\mu + b)}{\exp(a\mu+b) + \exp(-a\mu - b)} \ . \tag{98}$$

If $X \sim \psi_{a,b,\mu,\sigma^2}$, then

$$\begin{aligned} \mathbf{E}[\tanh(aX+b)] &= \tanh(a\mathbf{E}[X] + b - (2p-1)a^2\sigma^2) & (99) \\ &= \tanh(a\mu + b) \ . & (100) \end{aligned}$$

The term $-(2p-1)a^2\sigma^2$ is called the Onsager correction term, and is what allows us to move the expectation inside the tanh. The quenched distribution of l_1 is approximately $\psi_{\beta,h,r_1,1-q}$, and so

$$\langle \tanh(\beta l_1 + h) \rangle \cong \tanh(\beta r_1 + h) \tag{101}$$

and $r_1 = \langle l_1 \rangle - \beta(1-q)\langle \sigma_1 \rangle$. The quenched distribution is a random distribution with parameter r_1.

The Stein characterizing operator for ψ_{a,b,μ,σ^2} is

$$Tf(x) = f'(x) - \left(\frac{x-\mu}{\sigma^2} - a\tanh(ax+b) \right) f(x) \tag{102}$$

To see this, look at

$$f'(x) + \left(\frac{d}{dx} \log \psi_{a,b,\mu,\sigma^2}(x) \right) f(x) \ . \tag{103}$$

Recall that for the characteristic operator, if $X \sim \psi_{a,b,\mu,\sigma^2}$ then $\mathbf{E}[Tf(x)] = 0$ for all f and conversely.

We have to show that

$$\mathbf{E}\left\langle f'(l_1) - \left(\frac{l_1 - r_1}{1-q} - \beta \tanh(\beta l_1 + h)\right) f(l_1)\right\rangle^2 \longrightarrow 0 \ . \tag{104}$$

It is instructive to consider the contrast with the annealed equation. If $\mathbf{E}\langle\cdot\rangle \to 0$ then we've proved nothing. This comes from r_1 not being a constant. However, a quenched equation implies a distributional result because r_1 is a constant, given g.

The remaining steps are then

1. Start with

$$h_j(g) = \frac{1}{\sqrt{N}} \langle(\sigma_j - \langle\sigma_j\rangle) f(l_1)\rangle \ . \tag{105}$$

2. Then use the approximation Lemma to show that

$$\sum_{j=2}^{N} g_{1j} h_j \cong \sum_{j=2}^{N} \frac{\partial h_j}{\partial g_{1j}} \ . \tag{106}$$

3. Recognize, after some computation and using $R_{12} \cong q$, that (104) and (106) are the same.

The full details of these arguments can be found in the paper (S. Chatterjee, *Spin Glasses and Stein's Method*, arXiv:0706.3500v1 [math.PR]).

Lecture 29

Lecture date: Nov 2, 2007 *Scribe: Tanya Gordeeva*

When $h = 0$, a phase transition occurs at $\beta = 1$. We saw that in the high temperature phase, $R_{12} = O(N^{-1/2})$. Parisi conjectures that at $\beta = 1$ and $h = 0$, R_{12} is of order $N^{-1/3}$. Guerra proved that at $\beta = 1$ and $h = 0$, $\mathbf{E}\langle R_{12}^2\rangle \leq C/\sqrt{N}$ for some constant C, ie, $R_{12} = O(N^{-1/4})$ at most. Talagrand has another proof in his book, but it is complicated. Nothing better is known.

Using Stein's method, we will show that $\mathbf{E}\langle |R_{12}|^3\rangle \geq C/N$ for some $C > 0$.

Proof: Let $\psi(x)$ be the probability density

$$\frac{\cosh(x)e^{-x^2/2}}{\sqrt{2\pi e}}$$

(i.e. the symmetric mixture of $N(1,1)$ and $N(-1,1)$). Hopefully, this is the distribution of the local field as $N \to \infty$.

For any bounded, measurable f, let $Mf = \int_{-\infty}^{\infty} f(x)\psi(x)\,dx$. Define an operator U as

$$Uf(x) = \frac{e^{x^2/2}}{\cosh x}\int_{-\infty}^{\infty}\cosh(t)e^{-t^2/2}(f(t) - Mf)\,dt$$

and let $Tf(x) = f'(x) - (x - \tanh(x))f(x)$. Verify that $TUf = f - Mf$, so U is the inversion of the Stein operator.

Lemma 76 *$||Uf||_\infty \leq C||f||_\infty$ and $||(Uf)'||_\infty \leq C||f||_\infty$ for some universal constant C.*

Lemma 77 (Expansion lemma) *Fix a bounded, measurable $f : \mathbb{R} \to \mathbb{R}$, let $b_1, \ldots, b_m$ be arbitrary functions of σ. Assume that b_1 does not depend on σ_1. Then*

$$\begin{aligned}\mathbf{E}(\langle f(l_1)b_1\rangle\langle b_2\rangle\cdots\langle b_m\rangle) =& (Mf)\mathbf{E}(\langle b_1\rangle\cdots\langle b_m\rangle)\\ &- \sum_{r=2}^{m}\frac{1}{N}\sum_{j=2}^{N}\mathbf{E}(\langle \sigma_j Uf(l_1)b_1\rangle\langle b_2\rangle\cdots\langle b_r\sigma_1\sigma_j\rangle\langle b_{r+1}\rangle\cdots\langle b_m\rangle)\\ &+ \frac{m}{N}\sum_{j=2}^{N}\mathbf{E}(\langle\sigma_j Uf(l_1)b_1\rangle\langle b_2\rangle\cdots\langle b_m\rangle) + \frac{1}{N}\mathbf{E}(\langle (Uf)'(l_1)b_1\rangle\langle b_2\rangle\cdots\langle b_m\rangle)\\ &+ \frac{1}{N}\mathbf{E}(\langle\sigma_j Uf(l_1)b_1\sigma_1\rangle\langle b_2\rangle\cdots\langle b_m\rangle)\end{aligned}$$

Now take $f(x) = \tanh(x)$. Then $\langle \sigma_1\sigma_2 \rangle = \langle f(l_1)\sigma_2 \rangle$ (since $f(l_1)$ is the conditional expectation of σ_1 given the rest of the spins). Thus $\mathbf{E}\langle \sigma_1\sigma_2 \rangle^2 = \mathbf{E}(\langle f(l_1)\sigma_2 \rangle \langle \sigma_1\sigma_2 \rangle)$.

Let $b_1 = \sigma_2$, $b_2 = \sigma_1\sigma_2$. Since ψ is a symmetric density and tanh is odd, $Mf = 0$. Let $h = Uf$. Applying the expansion lemma,

$$\begin{aligned}
\mathbf{E}\langle \sigma_1\sigma_2 \rangle^2 =& 0 - \frac{1}{N}\sum_{j=2}^{N} \mathbf{E}(\langle \sigma_1 h(l_1)\sigma_2 \rangle \langle \sigma_2\sigma_j \rangle) \\
&+ \frac{2}{N}\sum_{j=2}^{N} \mathbf{E}(\langle \sigma_j h(l_1)\sigma_2 \rangle \langle \sigma_2\sigma_j \rangle \langle \sigma_1\sigma_j \rangle) \\
&+ \frac{1}{N}\mathbf{E}(\langle h'(l_1)\sigma_2 \rangle \langle \sigma_1\sigma_2 \rangle) \\
&+ \frac{1}{N}\mathbf{E}(\langle h(l_1)\sigma_1\sigma_2 \rangle \langle \sigma_1\sigma_2 \rangle)
\end{aligned}$$

To simplify the above, let *Rem* denote any term that is bounded by $CN^{-1}\sqrt{\mathbf{E}\langle R_{12} \rangle^2}$ for some constant C.

Since h' is bounded, so

$$\begin{aligned}
\frac{1}{N}\mathbf{E}(\langle h'(l_1)\sigma_2 \rangle \langle \sigma_1\sigma_2 \rangle) &\leq \frac{C}{N}\mathbf{E}|\langle \sigma_1\sigma_2 \rangle| \leq \frac{C}{N}\sqrt{\mathbf{E}\langle \sigma_1\sigma_2 \rangle^2} \\
&\leq \frac{C'}{N}\sqrt{\mathbf{E}\frac{1}{N^2}\sum_{i,j}\langle \sigma_i\sigma_j \rangle^2} = \frac{C'}{N}\sqrt{\mathbf{E}\langle R_{12} \rangle^2}.
\end{aligned}$$

So the fourth term is *Rem*. The last term is also *Rem*. Any single term in the sum in the third term is also *Rem*. In the 2nd term, for $j = 2$, we get $\frac{1}{N}\mathbf{E}\langle h(l_1) \rangle$. All other terms are *Rem*. So

$$E\langle \sigma_1\sigma_2 \rangle^2 = -\frac{1}{N}\mathbf{E}\langle h(l_1) \rangle - \mathbf{E}(\langle h(l_1)\sigma_2\sigma_3 \rangle \langle \sigma_2\sigma_3 \rangle) - 2\mathbf{E}(\langle h(l_1)\sigma_2\sigma_3 \rangle \langle \sigma_1\sigma_2 \rangle \langle \sigma_1\sigma_3 \rangle) - Rem.$$

This is the first expansion. To complete the proof, we apply the expansion lemma to each of the above terms. It will be enough to show:

$$-\frac{\mathbf{E}\langle h(l_1) \rangle}{N} = \frac{1}{N} + Rem,$$

$$-\mathbf{E}(\langle h(l_1)\sigma_2\sigma_3 \rangle \langle \sigma_2\sigma_3 \rangle) = E\langle \sigma_1\sigma_2 \rangle^2 + T_1,$$

where $|T_1| \leq C\mathbf{E}\langle |R_{12}|^3 \rangle$, and

$$|\mathbf{E}(\langle h(l_1)\sigma_2\sigma_3 \rangle \langle \sigma_1\sigma_2 \rangle \langle \sigma_1\sigma_3 \rangle)| \leq C\mathbf{E}\langle |R_{12}|^3 \rangle + Rem.$$

If we can show the above, then

$$\mathbf{E}\langle\sigma_1\sigma_2\rangle^2 = \frac{1}{N} + \mathbf{E}\langle\sigma_1\sigma_2\rangle^2 + T_1 + T_2$$

where T_1, T_2 are bounded by $C\mathbf{E}\langle|R_{12}|^3\rangle + Rem$. So

$$\frac{1}{N} \le C\mathbf{E}\langle|R_{12}|^3\rangle + Rem \le C\mathbf{E}\langle|R_{12}|^3\rangle + \frac{C'\sqrt{\mathbf{E}\langle R_{12}^2\rangle}}{N}$$

Now $\sqrt{\mathbf{E}\langle R_{12}^2\rangle} \le (\mathbf{E}\langle|R_{12}|^3\rangle)^{1/3}$. Suppose $\mathbf{E}\langle|R_{12}|^3\rangle \le \frac{1}{2CN}$. Then we get $1/N \le 1/2N + (C'/N)(1/2CN)^{1/3}$, a contradiction when N is large enough. So for N large enough, $\mathbf{E}\langle|R_{12}|^3\rangle \ge \frac{1}{2CN}$.

To prove the above three statements:

Apply the approximation lemma, and only the first terms will matter. Let $w = Uh$ (so we invert the Stein operator again). Verify that $Mh = -1$. By the expansion lemma,

$$\begin{aligned}\mathbf{E}\langle h(l_1)\rangle &= -1 + \frac{1}{N}\sum_{j=2}^{N}\mathbf{E}(\langle\sigma_j w(l_1)\rangle\langle\sigma_1\sigma_j\rangle) + \frac{\mathbf{E}\langle w'(l_1)\rangle + \mathbf{E}\langle\sigma_1\sigma_j\rangle}{N}\\ &= -1 + Rem + O(1/N).\end{aligned}$$

Using the expansion lemma on the second term, $\mathbf{E}(\langle h(l_1)\sigma_2\sigma_3\rangle\langle\sigma_2\sigma_3\rangle) = (Mh)\mathbf{E}\langle\sigma_2\sigma_3\rangle^2 = -\mathbf{E}\langle\sigma_2\sigma_3\rangle^2 = -\mathbf{E}\langle\sigma_1\sigma_2\rangle^2$ with some remainder terms. The third term can also be bounded through the expansion lemma.

Lastly, a sketch of the proof of the expansion lemma:

We want to find $\mathbf{E}\langle f(l_1)b_1\rangle\cdots\langle b_m\rangle$. Let $h = Uf$ so that $h'(x) - (x - \tanh(x))h(x) = f(x) - Mf$. Replace $f(l_1) - Mf$ by $h'(l_1) - l_1 h(l_1) + \tanh(l_1)h(l_1)$ and apply integration by parts on the terms arising from $l_1 h(l_1)$.

□

Exercise 78 *Get an upper bound for* $\mathbf{E}\langle|R_{12}|^3\rangle$.

Exercise 79 *Evaluate*

$$\lim_{N\to\infty}\mathbf{E}(N\langle R_{12}R_{23}R_{31}\rangle)$$

or, alternatively,

$$\lim_{N\to\infty}\mathbf{E}(N\langle\sigma_1\sigma_2\rangle\langle\sigma_2\sigma_3\rangle\langle\sigma_3\sigma_1\rangle)$$

You can use Guerra's result.

Lecture 30

Lecture date: Nov. 5, 2007 *Scribe: Laura Derksen*

We will again consider the SK model with parameters β and h. Let

$$l_1 = \frac{1}{\sqrt{N}} \sum_{j=2}^{N} g_{1j}\sigma_j,$$

and

$$r_1 = \frac{1}{\sqrt{N}} \sum_{j=2}^{N} g_{1j}\langle\sigma_j\rangle - \beta(1-q)\langle\sigma_1\rangle.$$

We previously sketched that l_1 is approximately equal in distribution to a random variable with density proportional to

$$\cosh(\beta x + h) \exp\left(\frac{-(x-r_1)^2}{2(1-q)}\right). \tag{107}$$

More precisely, let ν_1 be the random probability measure associated with the above density. Then for any bounded measurable function f,

$$\mathbf{E}\Big(\langle f(l_1)\rangle - \int f(x)\nu_1(dx)\Big)^2 \leq \frac{c\|f\|_\infty}{\sqrt{N}}. \tag{108}$$

Exercise 80 *Improve $\sqrt{N}$ to N.*

Exercise 81 *Get a bound on $\mathbf{E}[(\langle f(l_1)\rangle - \int f(x)\nu_1(dx))^{2k}]$. It should look something like $\frac{c\|f\|_\infty^k}{N^k}$ or $\frac{c\|f\|_\infty^k}{N^{k/2}}$.*

Exercise 82 *Get results for multiple local fields $l_1, l_2, ..., l_p$ where p is fixed. Even the case $p = 2$ would be interesting. When N is large, these become almost independent. Get something like $\langle f_1(l_1)f_2(l_2)...f_p(l_p)\rangle \simeq \prod_{i=1}^p \int f(x)\nu_i(dx)$.*

Next class we will look at TAP equations for some other models.

We saw that if W is a random variable with $\mathbf{E}(W^2) < \infty$ such that $\mathbf{E}[W\varphi(W)] = \mathbf{E}[T\varphi'(W)]$ for some bounded random variable T, then we can prove Tusnády's lemma for W. Many other things can be derived from this equation, the following lemma for example.

Lemma 83 *Suppose* $\mathbf{E}[W\varphi(W)] = \mathbf{E}[T\varphi'(W)]$ *for all Lipschitz* φ. *Then for any* σ^2,

$$d_{TV}(W, N(0,\sigma^2)) \leq \frac{2\,\mathbf{E}\,|T-\sigma^2|}{\sigma^2}, \tag{109}$$

where d_{TV} *denotes the total variation distance. In particular, if* $\sigma^2 = \mathrm{Var}(W)$, *the upper bound is*

$$d_{TV}(W, N(0,\sigma^2)) \leq \frac{2\sqrt{\mathrm{Var}(T)}}{\sigma^2}. \tag{110}$$

Proof: Taking $\varphi(x) \equiv 1$ we get $\mathbf{E}(W) = 0$, and taking $\varphi(x) = x$ gives $\mathbf{E}(T) = \mathrm{Var}(W)$. Given a bounded measurable function f such that $0 \leq f \leq 1$, find φ such that

$$\sigma^2\varphi'(x) - x\varphi(x) = f(x) - \mathbf{E}[f(Z)]$$

where Z is $N(0,\sigma^2)$. Then

$$\begin{aligned}\mathbf{E}[f(W)] - \mathbf{E}[f(Z)] &= \sigma^2\,\mathbf{E}[\varphi'(W)] - \mathbf{E}[W\varphi(W)] \\ &= \mathbf{E}[(\sigma^2 - T)\varphi'(W)] \\ &\leq \frac{2\,\mathbf{E}\,|T-\sigma^2|}{\sigma^2}\end{aligned}$$

since for $0 \leq f \leq 1$, $\|\varphi'\| \leq \frac{2}{\sigma^2}$. □

Suppose $W = f(X_1, X_2, ..., X_n)$ where $X_1, X_2, ..., X_n$ are i.i.d. standard Gaussian. Then there is a generic way to obtain T: let $X = (X_1, X_2, ..., X_n)$ and let $Y = (Y_1, Y_2, ..., Y_n)$ be independent of X and identically distributed. Then take

$$T = \int_0^1 \frac{1}{2\sqrt{t}} \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X)\frac{\partial f}{\partial x_i}(\sqrt{t}X + \sqrt{1-t}Y)\,dt. \tag{111}$$

For T we can also take the expectation of the above expression conditioned on X or W. With this T, we have $\mathbf{E}[W\varphi(W)] = \mathbf{E}[T\varphi'(W)]$ for all absolutely continuous φ.

An application: take $\varphi(x) = x$ to see

$$\mathrm{Var}(W) = \mathbf{E}(T) = \int_0^1 \frac{1}{2\sqrt{t}}\,\mathbf{E}\Big[\sum_{i=1}^n \frac{\partial f}{\partial x_i}(X)\frac{\partial f}{\partial x_i}(\sqrt{t}X + \sqrt{1-t}Y)\Big]\,dt.$$

Let $X_t = \sqrt{t}X + \sqrt{1-t}Y$. Then

$$\begin{aligned}\mathbf{E}\Big[\sum_{i=1}^n \frac{\partial f}{\partial x_i}(X)\frac{\partial f}{\partial x_i}(X_t)\Big] &\leq \mathbf{E}\left[\sqrt{\sum_{i=1}^n\left(\frac{\partial f}{\partial x_i}(X)\right)^2 \sum_{i=1}^n\left(\frac{\partial f}{\partial x_i}(X_t)\right)^2}\right] \\ &\leq \sqrt{\mathbf{E}[\|\nabla f(X)\|^2]\,\mathbf{E}[\|\nabla f(X_t)\|^2]} \\ &= \mathbf{E}[\|\nabla f(X)\|^2].\end{aligned}$$

Suppose f is Lipschitz with constant A, that is, $\|f(x) - f(y)\| \leq A\|x - y\|$ for all x and y. Then $\|\nabla f(x)\| \leq A$ for all x. Thus, with T defined as in (111), we have $|T| \leq A^2$ almost surely.

Let $m(\theta) = \mathbf{E}(e^{\theta W})$. Then

$$m'(\theta) = \mathbf{E}(We^{\theta W}) = \theta\, \mathbf{E}(Te^{\theta W}) \leq \theta A^2 m(\theta),$$

so

$$\frac{d}{d\theta} \log m(\theta) = \frac{m'(\theta)}{m(\theta)} < A^2\theta$$

and

$$\log m(\theta) \leq \log m(0) + \frac{A^2\theta^2}{2} = \frac{A^2\theta^2}{2}.$$

This gives

$$m(\theta) \leq \exp(A^2\theta^2/2),$$

therefore

$$\mathbf{P}(|W| \geq t) \leq \exp(-\theta t + A^2\theta^2/2) \tag{112}$$

for any $\theta > 0$.

Theorem 84 (Gaussian Concentration Inequality) *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is an A-Lipschitz function, and $X = (X_1, X_2, ..., X_n)$ is a random vector made up of i.i.d. standard Gaussian entries. Then*

$$\mathbf{P}(|f(X) - \mathbf{E}[f(X)]| \geq t) \leq 2e^{-t^2/2A^2} \tag{113}$$

for all $t \geq 0$.

Now, let's prove that our T in (111) works if we assume $\mathbf{E}[f(X)] = 0$.

$$\begin{aligned} \mathbf{E}[W\varphi'(W)] &= \mathbf{E}[\varphi(f(X))(f(X) - f(Y))] \\ &= \int_0^1 \mathbf{E}[\varphi(f(X))\frac{d}{dt}f(X_t)]\,dt \\ &= \int_0^1 \mathbf{E}\left[\varphi(f(X))\sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_t)\left(\frac{X_i}{2\sqrt{t}} - \frac{Y_i}{2\sqrt{1-t}}\right)\right] dt. \end{aligned}$$

Now fix some i and t and let $X_{t,i} = \sqrt{t}X_i + \sqrt{1-t}Y_i$. Let $Z_t = \sqrt{1-t}X - \sqrt{t}Y$, and $Z_{t,i} = \sqrt{1-t}X_i - \sqrt{t}Y_i$. Then $\sqrt{t}X_{t,i} + \sqrt{1-t}Z_{t,i} = X_i$, and $\text{Cov}(X_{t,i}, Z_{t,i}) = 0$, so because they are Gaussian, $X_{t,i}$ and $Z_{t,i}$ are independent. Therefore,

$$\begin{aligned} &\mathbf{E}\left[\varphi(f(X))\frac{\partial f}{\partial x_i}(X_t)\left(\frac{X_i}{2\sqrt{t}} - \frac{Y_i}{2\sqrt{1-t}}\right)\right] \\ &= \frac{1}{2\sqrt{t(1-t)}}\,\mathbf{E}\left[\varphi(f(\sqrt{t}X_t + \sqrt{1-t}Z_t))\frac{\partial f}{\partial x_i}(X_t)Z_{t,i}\right]. \end{aligned}$$

Integrate by parts to obtain

$$= \frac{1}{2\sqrt{t(1-t)}}\,\mathbf{E}\left[\varphi'(f(X))\frac{\partial f}{\partial x_i}(X)\sqrt{1-t}\frac{\partial f}{\partial x_i}(X_t)\right].$$

This completes the proof.

Lecture 31

Lecture date: Nov. 7, 2007 *Scribe: Anand Sarwate*

34 Gaussian concentration recap

If (W, T) is a pair of random variables such that

$$\mathbf{E}[W\varphi(W)] = \mathbf{E}[T\varphi'(W)] \tag{114}$$

for all Lipschitz φ, then for any $\sigma^2 > 0$

$$d_{TV}\left(\mathcal{L}(W), \mathcal{N}(0, \sigma^2)\right) \leq \frac{2\mathbf{E}|T - \sigma^2|}{\sigma^2} , \tag{115}$$

where $\mathcal{L}(W)$ is the law of W. In particular, if $\sigma^2 = \text{Var}(W)$, it is easy to see that $\mathbf{E}[T] = \sigma^2$, and hence

$$d_{TV}\left(\mathcal{L}(W), \mathcal{N}(0, \sigma^2)\right) \leq \frac{2\sqrt{\text{Var}(T)}}{\sigma^2} . \tag{116}$$

So how do we plan to use this? We have the following canonical construction: let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a vector of iid $\mathcal{N}(0, 1)$ random variables and $f : \mathbb{R}^n \to \mathbb{R}$ be an absolutely continuous function. Then for $W = f(\mathbf{X})$ with $\mathbf{E}[W] = 0$, $\mathbf{E}[W^2] < \infty$ we have

$$\mathbf{E}[W\varphi(W)] = \mathbf{E}[T\varphi'(W)] , \tag{117}$$

for all Lipschitz φ, where

$$T = \int_0^1 \frac{1}{2\sqrt{t}} \sum_{i=1}^n \frac{\partial f}{\partial X_i}(\mathbf{X}) \frac{\partial f}{\partial X_i}(\mathbf{X}_t) dt \tag{118}$$

where $\mathbf{X}_t = \sqrt{t}\ \mathbf{X} + \sqrt{1-t}\ \mathbf{Y}$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$ are also iid $\mathcal{N}(0, 1)$.

With this tool we proved the Gaussian Poincaré inequality and the Gaussian concentration inequality. Today we will start a method for obtaining normal approximations for quite complicated functions. For example, we will look at linear statistics of the eigenvalues of random matrices.

35 A CLT for functions of Gaussians

To get a CLT we first need to prove the concentration of T given by (118). Clearly, we can replace T by the conditional expectation $T(\mathbf{x}) = \mathbf{E}[T|\mathbf{X} = \mathbf{x}]$. This requires some ugly but straightforward calculation[4]. We begin by writing $T(\mathbf{x})$:

$$T(x_1, x_2, \ldots, x_n) = \int_0^1 \frac{1}{2\sqrt{t}} \mathbf{E}\left[\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}) \frac{\partial f}{\partial x_i}(\sqrt{t}\ \mathbf{x} + \sqrt{1-t}\ \mathbf{Y})\right] dt \tag{119}$$

Letting $\sigma^2 = \mathrm{Var}(W)$, we have

$$d_{TV}\left(\mathcal{L}(W), \mathcal{N}(0, \sigma^2)\right) \leq \frac{2}{\sigma^2}\sqrt{\mathrm{Var}(T(\mathbf{x}))} \tag{120}$$

By the Poincaré inequality,

$$\mathrm{Var}(T(\mathbf{X})) \leq \mathbf{E}\left\|\nabla T(\mathbf{X})\right\|^2 \tag{121}$$

These give what one might call the "2nd order Poincaré inequalities."

Continuing with the computation:

$$\frac{\partial T}{\partial x_i}(\mathbf{x}) = \int_0^1 \frac{1}{2\sqrt{t}} \mathbf{E}\left[\underbrace{\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \frac{\partial f}{\partial x_j}(\sqrt{t}\ \mathbf{x} + \sqrt{1-t}\ \mathbf{Y})}_{A_i(t)} \right. \tag{122}$$

$$\left. + \sqrt{t}\underbrace{\sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\mathbf{x}) \frac{\partial^2 f}{\partial x_i \partial x_j}(\sqrt{t}\ \mathbf{x} + \sqrt{1-t}\ \mathbf{Y})}_{B_i(t)}\right] dt\ . \tag{123}$$

What we really want to bound is the sum of squares of this expression. Using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ and Jensen's inequality, we get

$$\sum_{i=1}^{n}\left(\frac{\partial T}{\partial x_i}(\mathbf{x})\right)^2 \leq 2\sum_{i=1}^{n}\left(\int_0^1 \frac{1}{2\sqrt{t}}\mathbf{E}[A_i(t)]dt\right)^2 + 2\sum_{i=1}^{n}\left(\int_0^1 \frac{1}{2}\mathbf{E}[B_i(t)]dt\right)^2 \tag{124}$$

$$\leq 2\mathbf{E}\int_0^1 \frac{1}{2\sqrt{t}}\sum_{i=1}^{n} A_i(t)^2 dt + 2\mathbf{E}\int_0^1 \frac{1}{4}\sum_{i=1}^{n} B_i(t)^2 dt\ . \tag{125}$$

[4] You should be used to this by now!

Turning to the first term,

$$\sum_{i=1}^{n} A_i(t)^2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \frac{\partial f}{\partial x_j}(\sqrt{t}\ \mathbf{x} + \sqrt{1-t}\ \mathbf{Y}) \right)^2 \tag{126}$$

$$= \left\| \text{Hess}\, f(\mathbf{x}) \cdot \nabla f(\sqrt{t}\ \mathbf{x} + \sqrt{1-t}\ \mathbf{Y}) \right\|^2 \tag{127}$$

$$\leq \|\text{Hess}\, f(\mathbf{x})\|^2 \cdot \left\| \nabla f(\sqrt{t}\ \mathbf{x} + \sqrt{1-t}\ \mathbf{Y}) \right\|^2 . \tag{128}$$

We can get a similar bound for $B_i(t)$.

Note that in computing $\mathbf{E}\, \|\nabla T(\mathbf{X})\|$ we will encounter terms that can be bounded using the Cauchy-Schwarz inequality and the fact that $\mathbf{X} \stackrel{d}{=} \mathbf{X}_t$.

$$\mathbf{E}\left[\|\text{Hess}\, f(\mathbf{X})\|^2 \cdot \|\nabla f(\mathbf{X}_t)\|^2 \right] \leq \left(\mathbf{E}\, \|\text{Hess}\, f(\mathbf{X})\|^4 \right)^{1/2} \left(\mathbf{E}\, \|\nabla f(\mathbf{X}_t)\|^4 \right)^{1/2} . \tag{129}$$

Thus

$$\text{Var}(T(\mathbf{X})) \leq \mathbf{E}\, \|\nabla T(\mathbf{X})\| \tag{130}$$

$$\leq \left(2 \int_0^1 \frac{1}{2\sqrt{t}} dt + 2 \int_0^1 \frac{1}{4} dt \right) \left(\mathbf{E}\, \|\text{Hess}\, f(\mathbf{X})\|^4\, \mathbf{E}\, \|\nabla f(\mathbf{X}_t)\|^4 \right)^{1/2} \tag{131}$$

$$= \frac{5}{2} \left(\mathbf{E}\, \|\text{Hess}\, f(\mathbf{X})\|^4\, \mathbf{E}\, \|\nabla f(\mathbf{X}_t)\|^4 \right)^{1/2} . \tag{132}$$

We then have

$$d_{TV}\left(\mathcal{L}, \mathcal{N}(0, \sigma^2) \right) \leq \frac{2}{\sigma^2} \sqrt{\frac{5}{2}} \left(\mathbf{E}\, \|\text{Hess}\, f(\mathbf{X})\|^4\, \mathbf{E}\, \|\nabla f(\mathbf{X}_t)\|^4 \right)^{1/4} \tag{133}$$

$$= \frac{\sqrt{10}}{\sigma^2} \left(\mathbf{E}\, \|\text{Hess}\, f(\mathbf{X})\|^4\, \mathbf{E}\, \|\nabla f(\mathbf{X}_t)\|^4 \right)^{1/4} . \tag{134}$$

We have proved the following

Theorem 85 *If $W = f(X_1, X_2, \ldots, X_n)$ where $X = (X_1, X_2, \ldots, X_n)$ is a vector of iid $\mathcal{N}(0,1)$ random variables with $\mathbf{E}[W] = 0$, $\mathbf{E}[W^2] = \sigma^2$, and $f \in C^2(\mathbb{R}^n)$, then*

$$d_{TV}\left(\mathcal{L}(W), \mathcal{N}(0, \sigma^2) \right) \leq \frac{\sqrt{10}}{\sigma^2} \left(\mathbf{E}\, \|\text{Hess}\, f(\mathbf{X})\|^4\, \mathbf{E}\, \|\nabla f(\mathbf{X}_t)\|^4 \right)^{1/4} . \tag{135}$$

Exercise 86 *Improve this theorem so that it doesn't have any 4th powers.*

36 Looking forward : eigenvalues of random matrices

What sort of problems can we tackle with this machinery? Suppose

$$(X_{ij})_{1\le i,j<\infty} \tag{136}$$

are iid $\mathcal{N}(0,1)$ random variables and let

$$A_n = \frac{1}{\sqrt{n}}\,(X_{ij})_{1\le i,j<\infty}\ . \tag{137}$$

This is sometimes called the real Ginibré ensemble. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A_n are approximately uniformly distributed on the unit disc, in the following sense:

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i} \xrightarrow{a.s.} \text{Uniform(unit disc in } \mathbb{C})\ . \tag{138}$$

We can look at sums of the form

$$\sum_{i=1}^{n} f(\lambda_i)\ , \tag{139}$$

for some function $f : \mathbb{C} \to \mathbb{C}$. It turns out that under very general conditions on f, this is asymptotically Gaussian, meaning

$$\sum_{i=1}^{n} f(\lambda_i) - \mathbf{E}\left[\sum_{i=1}^{n} f(\lambda_i)\right] \tag{140}$$

converges in law. For symmetric random matrices, Sinai and Soshnikov proved this in 1998.

We will conclude with a brief chronology of the relevant results.

1. **Jonsson, D. Some limit theorems for the eigenvalues of a sample covariance matrix J. Multivariate Anal. 12 1–38 (1982).**
   Discusses sample covariance or Wishart matrices, which are of the form $A^T A$, where A is a matrix whose rows are sample data points.

2. **Ya. Sinaĭ, A. Soshnikov, Central limit theorem for traces of large random symmetric matrices, Bol. Soc. Brasil. Mat., 29, No. 1, 1-24 (1998).**

3. **Ya. Sinaĭ, A. Soshnikov, A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices, Functional Anal. Appl. 32, No. 2, (1998).**
   These papers prove a refinement and CLT for Wigner matrices, which are symmetric random matrices.

4. **Johansson, K. On fluctuations of eigenvalues of random Hermitian matrices. Duke Math. J. 91 151–204 (1998).**
This paper studies matrices whose entries have joint density proportional to $\exp(-nTr(V(A)))$, where V is a polynomial.

5. **Diaconis, P. and Evans, S.N. Linear functionals of eigenvalues of random matrices. Trans. Amer. Math. Soc. 353 2615–2633 (2001).**
This paper studies random unitary matrices and uses connections to symmetric functions.

6. **Chatterjee, S. Fluctuations of eigenvalues and second order Poincaré inequalities. arXiv:0705.1224v2 [math.PR].**
This will be our plan for the next few lectures.

Lecture 32

Lecture date: Nov 9, 2007 *Scribe: Guy Bresler*

37 Matrix Norms

In this lecture we prove central limit theorems for functions of a random matrix with Gaussian entries. We begin by reviewing two matrix norms, and some basic properties and inequalities.

1. Suppose A is a $n \times n$ real matrix. The *operator norm* of A is defined as

$$\|A\| = \sup_{|x|=1} \|Ax\|, \quad x \in \mathbb{R}^n.$$

Alternatively,

$$\|A\| = \sqrt{\lambda_{\max}(A^T A)},$$

where $\lambda_{\max}(M)$ is the maximum eigenvalue of the matrix M.

Basic properties include:

$$\|A + B\| \leq \|A\| + \|B\|$$
$$\|\alpha A\| = |\alpha| \|A\|$$
$$\|AB\| \leq \|A\| \|B\|.$$

2. The *Hilbert Schmidt* (alternatively called the Schur, Euclidean, Frobenius) *norm* is defined as

$$\|A\|_{\text{HS}} = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\text{Tr}(A^T A)}.$$

Clearly,

$$\|A\|_{\text{HS}} = \sqrt{\text{sum of eigenvalues of } A^T A},$$

which implies that

$$\|A\| \leq \|A\|_{\text{HS}} \leq \sqrt{n} \|A\|.$$

Of course, $\|A\|_{\text{HS}}$ also satisfies the usual properties of a norm.

Proposition 87 *The following inequality holds:*

$$\|AB\|_{HS} \leq \|A\| \|B\|_{HS}.$$

Proof: Let $b_1, \ldots, b_n$ denote the columns of B. Then

$$\|AB\|_{\mathrm{HS}}^2 = \sum_{i=1}^n \|Ab_i\|^2 \le \sum_{i=1}^n \|A\|^2 \|b_i\|^2 = \|A\|^2 \|B\|_{\mathrm{HS}}^2.$$

□

3. A simple matrix inequality follows from the Cauchy-Schwarz inequality:

$$|\operatorname{Tr}(AB)| = \sum_{i,j} a_{ij} b_{ji} \le \|A\|_{\mathrm{HS}} \|B\|_{\mathrm{HS}}.$$

4. Combining the proposition above with observation 3 gives the inequality

$$|\operatorname{Tr}(ACBD)| \le \|AC\|_{\mathrm{HS}} \|BD\|_{\mathrm{HS}} \le \|A\| \|B\| \|C\|_{\mathrm{HS}} \|D\|_{\mathrm{HS}}.$$

More generally, it holds that

$$|\operatorname{Tr}(A_1 A_2 \ldots, A_k)| \le \|A_i\|_{\mathrm{HS}} \|A_j\|_{\mathrm{HS}} \prod_{l \ne i,j} \|A_l\|.$$

Next, recall the theorem from last lecture:

Theorem 88 *Let $X_1, \ldots, X_k$ be i.i.d. $\mathcal{N}(0,1)$ random variables. Let $f \in C^2(\mathbb{R}^n)$ and $W = f(X)$ with $\mathbf{E}\, W = 0$. Then*

$$d_{TV}(\mathcal{L}(W), \mathcal{N}(0,\sigma^2)) \le \frac{\sqrt{10}}{\sigma^2} (\mathbf{E} \| \operatorname{Hess} f(X)\|^4 \mathbf{E} \|\nabla f(X)\|^4)^{\frac{1}{4}}.$$

We will use this theorem to study the Gaussian random matrix.

38 CLT for $\operatorname{Tr}(A^k)$

Suppose

$$A = \frac{1}{\sqrt{N}} (X_{ij})_{1 \le i,j \le N},$$

where $X_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. Fix a positive integer k. We would like a CLT for $\operatorname{Tr}(A^k)$.

To begin, note that

$$\operatorname{Tr}(A^k) = \frac{1}{N^{k/2}} \sum_{1 \le i_1, i_2, \ldots, i_k \le N} X_{i_1,i_2} X_{i_2,i_3} \ldots X_{i_{k-1},i_k} X_{i_k,i_1}\,. \tag{141}$$

It turns out that the usual dependency graph theorem fails for $k \ge 3$, so a more powerful method must be used.

Exercise 89 *Find a dependency graph theorem that works for all k.*

In order to apply Theorem 88, we identify

$$X = (X_{11}, X_{12}, \ldots, X_{1k}, X_{21}, X_{22}, \ldots, X_{NN}),$$

and $f(X) = \mathrm{Tr}(A^k)$. Now

$$\frac{\partial f}{\partial x_{ij}} = \mathrm{Tr}(\frac{\partial}{\partial x_{ij}} A^k) \stackrel{(a)}{=} Tr\left(\sum_{r=0}^{k-1} A^r \frac{\partial A}{\partial x_{ij}} A^{k-1-r}\right) \stackrel{(b)}{=} kTr\left(\frac{\partial A}{\partial x_{ij}} A^{k-1}\right), \tag{142}$$

where (a) follows from the fact that for two matrices A and B, $\frac{\partial}{\partial x}AB = \frac{\partial A}{\partial x}B + A\frac{\partial B}{\partial x}$, and (b) from moving the trace inside the sum and using $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. But

$$\frac{\partial A}{\partial x_{ij}} = \frac{1}{\sqrt{N}} e_i e_j^T,$$

where e_i is the ith standard basis vector, i.e. the vector of all zeros with a 1 in the ith position.

Thus

$$\begin{aligned}
\frac{\partial f}{\partial x_{ij}} &= \frac{k}{\sqrt{N}} \mathrm{Tr}(e_i e_j^T A^{k-1}) \\
&= \frac{k}{\sqrt{N}} \mathrm{Tr}(e_j^T A^{k-1} e_i) \\
&= \frac{k}{\sqrt{N}} (A^{k-1})_{ji}
\end{aligned}$$

This allows us to calculate

$$\begin{aligned}
\|\nabla f(X)\|^2 = \sum \left(\frac{\partial f}{\partial x_{ij}}\right)^2 &= \frac{k^2}{N} \sum_{i,j} (A^{k-1})_{ji}^2 \\
&= \frac{k^2}{N} \|A^{k-1}\|_{\mathrm{HS}}^2 \\
&\leq \frac{k^2}{N} N \|A^{k-1}\|^2 \\
&\leq k^2 \|A\|^{2(k-1)}.
\end{aligned} \tag{143}$$

Lemma 90

$$\mathbf{E}\|A\|^p \leq C(p) \quad \forall p \in \mathbb{Z}_+,$$

where $C(p)$ is a constant independent of N.

Proof: The proof is essentially as follows. For a positive definite random matrix B, $\|B\| = \lambda_{\max}(B)$. Thus

$$\begin{aligned}\mathbf{E}\|B\|^p = \mathbf{E}\lambda_{\max}^p &\le (\mathbf{E}\lambda_{\max}^{pm})^{1/m} \quad \text{for any } m\\ &\le (\mathbf{E}\,\mathrm{Tr}(B^{pm}))^{1/m}.\end{aligned}$$

Now let $m \to \infty$ suitably with N.□

This shows that $\|\nabla f(X)\|^2 = O(1)$, and hence the Poincaré inequality implies that $\mathrm{Var}(f(X)) = O(1)$.

Exercise 91 *Show that any two terms in the sum of equation* (141) *have non-negative covariance.*

The exercise implies that

$$\mathrm{Var}(f(X)) \ge \frac{1}{N^k}\sum \mathrm{Var}(X_{i_1,i_2}\dots X_{i_k,i_1}) \ge C(k) > 0\,.$$

Recalling the result of Theorem 88,

$$d_{TV}(\mathcal{L}(W), \mathcal{N}(0,\sigma^2)) \le \frac{\sqrt{10}}{\sigma^2}(\mathbf{E}\|\,\mathrm{Hess}\,f(X)\|^4\mathbf{E}\|\nabla f(X)\|^4)^{\frac{1}{4}}\,,$$

we see that $\sigma^2 = \mathrm{Var}(f(X)) \ge C(k)$ and from equation (143) and the fact noted above, $\mathbf{E}\|\nabla f(X)\|^2 \le C(k)$. Therefore it remains only to show that $\mathbf{E}\|\,\mathrm{Hess}\,f(X)\|^4 \to 0$ in order to prove the desired central limit theorem.

We have

$$\frac{\partial A}{\partial x_{ij}} = kTr\left(\frac{\partial A}{\partial x_{ij}}A^{k-1}\right),$$

and

$$\frac{\partial^2 A}{\partial x_{pq}x_{ij}} = kTr\left(\sum_{r=0}^{k-2}\frac{\partial A}{\partial x_{ij}}A^r\frac{\partial A}{\partial x_{pq}}A^{k-r-2}\right).$$

Fact about matrix norms: If A is a symmetric, real matrix then

$$\|A\| = \sup_{\|x\|=\|y\|=1}|x^TAy|\,.$$

Now, $\mathrm{Hess}\,f(X)$ is an $N^2 \times N^2$ symmetric, real matrix:

$$\|\,\mathrm{Hess}\,f(X)\| = \sup\left\{\sum_{ijpq}c_{ij}d_{pq}\frac{\partial^2 f}{\partial x_{ij}\partial x_{pq}} : \sum c_{ij}^2 = 1, \sum d_{pq} = 1\right\}.$$

Let $C = (c_{ij})$ and $D = (d_{pq})$ be two matrices with $\|C\|_{\text{HS}} = \|D\|_{\text{HS}} = 1$. Fix $0 \le r \le k-2$. Then

$$\sum_{ijpq} c_{ij} d_{pq} Tr\left(\frac{\partial A}{\partial x_{ij}} A^r \frac{\partial A}{\partial x_{pq}} A^{k-2-r}\right) = \frac{1}{N} \sum c_{ij} d_{pq} \operatorname{Tr}(e_i e_j^T A^r e_p e_q^T A^{k-2-r}) = \frac{1}{N} \operatorname{Tr}(CA^r DA^{k-2-r}),$$

where we used the fact that $\sum c_{ij} e_i e_j^T = C$ and similarly for D.

Now

$$|\operatorname{Tr}(CA^r DA^{k-2-r})| \le \|A\|^{k-2} \|C\|_{\text{HS}} \|D\|_{\text{HS}} = \|A\|^{k-2}.$$

Thus

$$\|\operatorname{Hess} f(X)\| \le \frac{k(k-1)\|A\|^{k-2}}{N}.$$

Combining, we get the desired result:

$$d_{TV}(\operatorname{Tr}(A^k), \mathcal{N}(0, \sigma^2)) \le \frac{C(k)}{N}.$$

Lecture 33

Lecture date: Nov 14, 2007 *Scribe: Maximilian Kasy*

39 Zero Bias Coupling

Consider a random variable W with $EW = 0$ and $EW^2 = \sigma^2 < \infty$. Define

$$\rho(x) = \frac{1}{\sigma^2} E(W 1_{(W \geq x)}).$$

Lemma 92 *ρ is a probability density*

Proof:

- $\rho(x) \geq 0$ for all x. For $x \geq 0$ this is obvious, for $x < 0$ note $E(W1_{(W \geq x)}) = -E(W1_{(W<x)}) \geq 0$
- $\int \rho(x)dx = 1$:

$$\int \rho(x)dx = \int_{x<0} -E(W1_{(W<x)})dx + \int_{x \geq 0} E(W1_{(W \geq x)})dx$$

Let μ denote the law of W. Then

$$\int_{x \geq 0} E(W1_{(W \geq x)})dx = \int_{x \geq 0} \int_{y \geq x} y d\mu(y) dx = \int_{y \geq 0} \int_{x \leq y} y \, dx \, d\mu(y) = \int_{y \geq 0} y^2 d\mu(y)$$

and similarly for the second term.□

The distribution corresponding to this density is called the "zero bias transform". If W^* is a random variable following the zero-bias transform of the law of W, then for all absolutely continuous φ we have

$$EW\varphi(W) = \sigma^2 E\varphi'(W^*)$$

(this is immediate from integration by parts). If $W \sim N(0,1)$ then

$$\rho(x) = \int_{y \geq x} \frac{y e^{-y^2/2}}{\sqrt{2\pi}} dy = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

so that W W^*, i.e. the standard normal distribution is a fixed point of the zero bias transform.

Example: $W \sim \pm 1$. Then $W^* \sim Uni[-1, 1]$.

Theorem 93 *If $EW = 1$, $EW^2 = 1$, then $Wass(W, Z) \leq 2Wass(W, W^*)$*

Proof: Take any 1-Lipschitz h. Find φ such that $\varphi'(x) - x\varphi(x) = h(x) - Eh(Z)$. By earlier results we know $\|\varphi''\|_\infty \leq 2\|h'\|_\infty \leq 2$. Now suppose W and W^* live on the same space. Then

$$|E(h(W)) - E(h(Z))| = |E(\varphi'(W) - W\varphi(W))| = |E(\varphi'(W) - \varphi'(W^*))| \leq 2E|W^* - W|$$

SInce this is true for any coupling, any h Lipschitz we have

$$Wass(W, Z) \leq 2 \inf_{W,W^*} E|W^* - W| = 2Wass(W, W^*)$$

Example: suppose $X_1, \ldots, X_n$ i.i.d. mean 0 variance 1, $W = \frac{\sum X_i}{\sqrt{n}}$. For each i let X_i be independent of everything else. Let $I \sim Unif\{1, \ldots, n\}$ independent of everything else. Define

$$W^* := \frac{1}{\sqrt{n}} \left[\sum_{i \neq I} X_i + X_I^* \right]$$

We claim this is a zero bias transform:

$$EW\varphi(W) = \frac{1}{\sqrt{n}} \sum EX_i\varphi(W) =$$

$$\frac{1}{\sqrt{n}} \sum E \left(X_i \varphi \left(\frac{1}{\sqrt{n}} \left[\sum_{j \neq i} X_j + X_i \right] \right) \right)$$

$$\frac{1}{\sqrt{n}} \sum E \left(\varphi' \left(\frac{1}{\sqrt{n}} \left[\sum_{j \neq i} X_j + X_i^* \right] \right) \frac{1}{\sqrt{n}} \right)$$

$$= E(\varphi'(W^*))$$

Thus

$$Wass(W, Z) \leq 2E|W^* - W| = 2E|\frac{1}{\sqrt{n}}(X_I - X_I^*)| = 2E|\frac{1}{\sqrt{n}}(X_1 - X_1^*)| \leq \ldots$$

Theorem 94 *(Goldstein-Reinert) Suppose Y, Y' are an exchangeable pair, $EY = 0$, $EY^2 = \sigma^2$ and $E(Y'|Y) = (1 - \lambda)Y$. Let ν denote the joint distribution of (Y, Y'). Let*

$$d\mu(y, y') = \frac{(y - y')^2}{E(Y - Y')^2} d\nu(y, y').$$

Suppose $(\hat{Y}, \hat{Y}') \sim \mu$. Let $U \sim Uni[0, 1]$ independent of all else. $Y^* = U\hat{Y} + (1 - U)\hat{Y}'$. Then Y^* is a zero bias transform of Y.

Note that $Ef'(Ua + (1-U)b) = \frac{f(b)-f(a)}{b-a}$, hence

$$\sigma^2 E(f'(Y^*)) = \sigma^2 E\left(\frac{f(\hat{Y}) - f(\hat{Y}')}{\hat{Y} - \hat{Y}'}\right)$$

$$= \frac{\sigma^2}{E(Y-Y')^2} E[(Y-Y')(f(Y) - f(Y'))] = \frac{2\lambda\sigma^2 E[Yf(Y)]}{E(Y-Y')^2} = EYf(Y)$$

Exercise: Try to get Hoeffding CLT using this method.

Exercise: Suppose $EW\varphi(W) = ET\varphi'(W)$.Then $E(T|W)$ is the density of $Law(W^*)$ w.r.t. $Law(W)$ evaluated at W. We have Tusnády's lemma based on concentration of T. On the other hand, if W, W^* can be constructed to be close to each other then we can construct W, Z such that $E|W-Z|$ is small. Question: If we know tail bounds on $|W - W^*|$ can we construct (W, Z) with fast decaying tails for $W - Z$?

Lecture 34

Lecture date: November 16, 2007 *Scribe: Joel Mefford*

Size Bias Transformations

Size bias transformations are a classical topic in probability, with applications to sampling theory. Their first connection with Stein's method was in a paper by Goldstein and Rinott.

The definition and idea behind size bias transformations

Suppose W is a non-negative random variable with mean λ. A random value W^* is called a size bias transformation of W if

$$\forall g. \, \mathbf{E}\, W g(W) = \lambda \, \mathbf{E}\, g(W^*).$$

Now, if μ is the law of W, then

$$\mathrm{d}\,\mu^*(x) = \frac{x}{\lambda}\,\mathrm{d}\,\mu(x)$$

is again a probability measure, and if $W^* \sim \mu^*$, then W^* is a size bias transform of W.

For example:
Let $X_1, \ldots, X_n$ be independent, non-negative random variables and $W = \sum_{i=1}^{n} X_i$.
For each i, let X_i^* be a size bias transform of X_i.
Let

$$W^* = W_i + X_i^* = \sum_{j \neq i} X_j + X_i^*,$$

with probability

$$\frac{\mathbf{E}(X_i)}{\sum_{j=1}^{n} \mathbf{E}(X_j)}.$$

Then,

$$\begin{aligned}
\mathbf{E}Wg(W) &= \sum_{i=1}^{n} \mathbf{E}\left(X_i g(W)\right) \\
&= \sum_{i=1}^{n} \mathbf{E}\left(X_i g(W_i + X_i)\right) \\
&= \sum_{i=1}^{n} \mathbf{E}(X_i)\mathbf{E}g\left(W_i + X_i^*\right) \\
&= \left(\sum_{i} \mathbf{E}(X_i)\right) \sum_{i=1}^{n} \frac{\mathbf{E}(X_i)}{\sum_j \mathbf{E}(X_j)} \mathbf{E}g(W_i + X_i^*) \\
&= \mathbf{E}(W)\mathbf{E}g(W^*)
\end{aligned}$$

So, now the question is:
If we have a size bias transformation, how can we use it to get a Central Limit Theorem?

The idea is that if $W^* - W$ is "small", say relative to W, and $\mathbf{E}(W^* - W \mid W)$ is concentrated in an appropriate scale, then W is approximately normal.

Why is that?

Take any bounded Lipschitz function h, and let $z \sim N(0,1)$.
Let f solve the usual Stein's Equation:

$$f'(x) - xf(x) = h(x) - \mathbf{E}h(Z)$$

Let $g(x) = f\left(\frac{x-\lambda}{\sigma}\right)$, where $\lambda = \mathbf{E}(W)$, and $\sigma^2 = \mathrm{Var}(W)$.

We would like to bound

$$\begin{aligned}
& \mathbf{E}h\left(\frac{W-\lambda}{\sigma}\right) - \mathbf{E}h(Z) \\
= \; & \mathbf{E}\left[f'\left(\frac{W-\lambda}{\sigma}\right) - \left(\frac{W-\lambda}{\sigma}\right) f\left(\frac{W-\lambda}{\sigma}\right)\right] \\
= \; & \mathbf{E}\left[\sigma g'(W) - \left(\frac{W-\lambda}{\sigma}\right) g(W)\right].
\end{aligned}$$

Now $\mathbf{E}Wg(W) = \lambda \mathbf{E}g(W^*)$, and so,

$$\begin{aligned} & \mathbf{E}\left[\sigma g'(W) - \left(\frac{W-\lambda}{\sigma}\right) g(W)\right] \\ = \; & \mathbf{E}\left[\sigma g'(W) - \frac{\lambda}{\sigma}\left(g(W^*) - g(W)\right)\right]. \end{aligned}$$

Using the size bias property, one can show that,

$$\mathbf{E}(W^* - W) = \frac{\sigma^2}{\lambda}$$

(take $g(x) = x - \lambda$).

We can approximate,

$$g(W^*) - g(W) \approx (W^* - W)g'(W),$$

and

$$\mathbf{E}\left((W^* - W)g'(W)\right) = \mathbf{E}\left(\mathbf{E}(W^* - W \mid W)g'(W)\right).$$

Thus, if $\mathbf{E}(W^* - W \mid W)$ is concentrated, then

$$\mathbf{E}\left(\mathbf{E}(W^* - W \mid W)g'(W)\right) \approx \mathbf{E}(W^* - W)\mathbf{E}g'(W) = \frac{\sigma^2}{\lambda}\mathbf{E}g'(W).$$

Thus,

$$\begin{aligned} \mathbf{E}\left[\frac{\lambda}{\sigma}\left(g(W^*) - g(W)\right)\right] & \approx \mathbf{E}\left[\frac{\lambda}{\sigma}\frac{\sigma^2}{\lambda}g'(W)\right] \\ & = \sigma \mathbf{E}g'(W) \end{aligned}$$

So, if the assumptions that $W^* - W$ is small and $\mathbf{E}(W^* - W \mid W)$ is concentrated hold, we get a Normal approximation theorem.

Goldstein and Rinott show that

$$\left|\mathbf{E}h\left(\frac{W-\lambda}{\sigma}\right) - \mathbf{E}h(Z)\right| \leq \frac{2\lambda\|h\|_\infty}{\sigma^2}\sqrt{\mathrm{Var}\left(\mathbf{E}(W^* - W \mid W)\right)} + \|h\|_\infty \frac{\lambda}{\sigma^3}\mathbf{E}(W^* - W)^2.$$

Their paper has details and examples.

Note that a “zero bias transform” gets its name by being a limiting case of a size bias transform.

Interaction Graphs

Suppose you have n points chosen uniformly at random from the unit square, $[0,1]^2$.
Let d_i be the distance from the i'th point to its nearest neighbor.

Can we get a Central Limit Theorem for $\sum_{i=1}^{n} d_i$ *?*

The challenge with such a limit is that the dependence between the points, or the values of d_i, are only apparent after all of the points are present.

We can approach the problem by using Interaction Graphs.

Definitions

Consider the Polish Space $\mathscr{X}$, e.g. $\mathbb{R}^d$.
A map G that takes a point $x \in \mathscr{X}^n$ and outputs an undirected graph on $[n] = \{1, \ldots, n\}$ will be called a "graphical rule".

A graphical rule is called "symmetric" if when you permute the coordinates, the graph permutes:

$$\forall \pi \in \mathrm{S}_n, \forall x \in \mathscr{X}^n, G\left(x_{\pi(1)}, \ldots, x_{\pi(n)}\right) = \{(\pi(i), \pi(j)) \mid (i,j) \in G(x)\}$$

Given a function $f : \mathscr{X}^n \longrightarrow \mathbb{R}$, we want to associate a graph G with it. First, some more definitions:

$\forall i$ let $x^i = (x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)$

$\forall i, j, x^{ij} = \left(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_{j-1}, x_j', x_{j+1}, \ldots, x_j\right)$

Say that two indices i and j are "non-interacting" under the triplet (f, x, x') if,

$$f(x) - f(x^j) = f(x^i) - f(x^{ij}).$$

Note that the definition is symmetric in i and j. This is a discrete analog of $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = 0$.

Given a function f, we say that a symmetric graphical rule or model G is an interction graph for f if,

$\forall x, x' \in \mathscr{X}^n, \forall i, j,$
$(i,j) \notin G(x), G(x^i), G(x^j), \text{or } G(x^{ij}) \implies (i,j)$ non-interacting under (f, x, x').

This definition of an interaction graph states that x^i and x^j are far apart in all four graphical rules. Thus i and j are "non-interacting" in (f, x, x').

This is a type of functional independence.

Given a symmetric graphical rule G on $\mathscr{X}^n$, and another rule G' on $\mathscr{X}^m, m > n$, we say G' is an extension of G if,

whenever $x = (x_1, \ldots, x_n) \subseteq y = (y_1, \ldots, y_m)$, in the sense that
$\exists i_1, i_2, \ldots, i_n$(distinct) $\in [m]$ such that $\forall k, y_{i_k} = x_k$,
we have $G(x)$ as the induced subgraph of $G'(y)$.

Theorem 95 *Let $\mathscr{X}$ be a Polish space and $X_1, X_2, \ldots$ be iid $\mathscr{X}$-valued random variables.*

Suppose $f : \mathscr{X} \to \mathbb{R}$ is measurable.

Let $W = f(X_1, \ldots, X_n)$ such that $\mathbf{E}(W) = 0$ and $E(W^2) = \sigma^2 < \infty$.

Let G be an interaction rule for f with extension G' to $\mathscr{X}^{n+4}$.

Let $X_1', X_2', \ldots$ be independent copies of $X_1, X_2, \ldots$.

Let there be the discrete derivative Δ_j:

$$\Delta_j f(X) := f(X_1, \ldots, X_n) - f(X_1, \ldots, X_{j-1}, X_j', X_{j+1}, \ldots, X_n).$$

Let $M = \max |\Delta_j f(X)|$, and let $\delta = 1+$ the degree of vertex 1 in $G'(X_1, \ldots, X_{n+4})$.

Then,

$$\text{Wass}\left(\frac{W - \mathbf{E}(W)}{\sqrt{\text{Var}(W)}}, N(0,1)\right) \leq \frac{Cn^{\frac{1}{2}}\left[\mathbf{E}(M^8)^{\frac{1}{4}}\mathbf{E}(\delta^4)^{\frac{1}{4}}\right]}{\sigma^2} + \frac{1}{2\sigma^3}\sum_j \mathbf{E}\left|\Delta_j f(X)\right|^3$$

Next time we will see examples of the uses of interaction graphs, such as in calculating the number of empty boxes when n balls are dropped into αn boxes.