Stein's method in concentration inequalities, spin glasses, random matrices, and strong approximations

Sourav Chatterjee (UC Berkeley)

This tutorial consists of four lectures:

1. A new method of normal approximation
2. A new approach to strong embeddings
3. Spin glasses and Stein's method
4. Stein's method for concentration inequalities

A new method of normal approximation

Sourav Chatterjee (UC Berkeley)

Central limit theorems

- Classical CLT: If $X_1, \ldots, X_n$ are independent random variables with zero mean and finite variance and (...), then

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

has a gaussian distribution in the limit.

- Often, the X_i's need to be only *approximately* independent.
- Many ways to prove such results, e.g.
  1. Characteristic functions
  2. Martingales & Skorohod embeddings
  3. Little blocks, big blocks
  4. Stein's method (and variants)
  5. Hájek Projections
  6. Specialized techniques for special problems

... but they all require luck, and often, very hard work.

An example from statistics

- Let $X_1, \ldots, X_n$ be i.i.d. random d-vectors.
- Let
$$f(X_1, \ldots, X_n) = \frac{1}{\sqrt{n}} \sum_i f_i, \tag{1}$$
where f_i is a function of X_i and its k nearest neighbors in the set $\{X_1, \ldots, X_n\}$.
- Data might lie on a complicated lower-dimensional manifold.
- Specific example (Levina & Bickel '05): Unbiased estimate of the dimension of the manifold.
- **Routine question:** How to find normal approximation bounds for $f(X_1, \ldots, X_n)$?

An example from statistics

- **Routine answer:** Intuitively plausible; low dependence at distances, etc. But...
- ... technical nightmare, specially for general manifolds.
- **Deeper issue:** Why can't we prove such a plausible result in our sleep?

Another example: Linear statistics of random eigenvalues

- Let $A = (a_{ij})$ be a real symmetric random matrix of order n, with i.i.d. entries on and above the diagonal.
- Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $n^{-1/2}A$.
- Let f be any function on $\mathbb{R}$ and let $W = \sum_{i=1}^n f(\lambda_i)$. This is called a "linear statistic of the eigenvalues of A".
- Mysterious fact (Sinai & Soshnikov): Under some smoothness assumptions on f, $\mathrm{Var}(W)$ converges to a positive limit σ^2 as $n \to \infty$. Moreover, $W - \mathbb{E}(W)$ converges in law to $N(0, \sigma^2)$.

Linear statistics contd.

- Similar results hold for sample covariance matrices (Bai & Silverstein), random unitary matrices (Diaconis & Evans), and other ensembles (e.g. Rider & Virág).
- These are examples where the usual methods of normal approximation do not work. May be we don't understand everything about normal approximation?

Stein's method

- If $Z \sim N(0,1)$, then $\mathbb{E}(\varphi(Z)Z) = \mathbb{E}(\varphi'(Z))$ for all φ.
- Stein's idea: If $\mathbb{E}(\varphi(W)W) \approx \mathbb{E}(\varphi'(W))$ for many φ's, then W is approximately $N(0,1)$.
- Many variants, e.g.
  - Exchangeable pairs
  - Zero bias couplings
  - Size bias couplings
  - Generator approach
  - Dependency graphs
- Common complaint: Hard to apply to arbitrary problems.

A conceptual insight

- Define

$$S_p(W) := \sup\{|\mathbb{E}(\varphi(W)W - \varphi'(W))| : \|\varphi'(W)\|_p \le 1\}.$$

- From Stein's lemma: $d_{TV}(W, Z) \le 2S_p(W)$ for every $p > 1$. (Recall: $d_{TV}(X, Y) = \sup_A |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$.)

Theorem

If W has mean zero, unit variance, and density ρ, then

$$S_p(W) = \|h(W) - \mathbb{E}h(W)\|_q,$$

where

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

and $\frac{1}{p} + \frac{1}{q} = 1.$

- This shows that approximate normality of W = concentration of $h(W)$.
- Proof is based on L^p-L^q duality from functional analysis.

Idea

- Concentration problems, unlike normal approximation problems, are *transferable* via conditional expectation. That is, if we can write

$$h(W) = \mathbb{E}(T \mid W),$$

where T is a an explicit object arising from the given problem, then

$$\|h(W) - \mathbb{E}h(W)\|_q \le \|T - \mathbb{E}T\|_q.$$

This might wake you up...

Theorem

Suppose $W = f(X_1, \ldots, X_n)$, where X_i's are i.i.d. $N(0,1)$, and f is smooth. Assume $\mathbb{E}(W) = 0$ and $\mathbb{E}(W^2) = 1$. Let Z be a vector of n i.i.d. standard gaussians and define $T : \mathbb{R}^n \to \mathbb{R}$ as

$$T(x) := \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) \int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E}\left(\frac{\partial f}{\partial x_i}(\sqrt{t}x + \sqrt{1-t}Z) \right) dt.$$

Then $h(W) = \mathbb{E}(T(X_1, \ldots, X_n) \mid W)$.

(Recall: This implies that $d_{TV}(W, N(0,1)) \leq 2\sqrt{\mathrm{Var}(h(W))}$.)

Simple examples

$f(X_1, \ldots, X_n)$	$T(X_1, \ldots, X_n)$
$\frac{\sum_{i=1}^n X_i}{\sqrt{n}}$	1
$\frac{\sum_{i=1}^n X_i^2 - n}{\sqrt{2n}}$	$\frac{\sum_{i=1}^n X_i^2}{n}$
$\frac{\sum_{i=1}^n X_i X_{i+1}}{\sqrt{n}}$	$\frac{1}{2n} \sum_{i=2}^{n} (X_{i-1} + X_{i+1})^2 + \frac{X_1^2 + X_{n+1}^2}{2n}$

Sketch of proof

- h is characterized by $\mathbb{E}(\varphi(W)W) = \mathbb{E}(\varphi'(W)h(W))$.
- Suffices to show $\mathbb{E}(\varphi(W)W) = \mathbb{E}(\varphi'(W)T(X))$.
- Let $X^t = \sqrt{t}X + \sqrt{1-t}Z$. Then

$$\begin{aligned}\mathbb{E}(\varphi(W)W) &= \mathbb{E}(\varphi(W)(f(X^1) - f(X^0))) \\ &= \mathbb{E}\Big(\varphi(W)\int_0^1 \frac{d}{dt}f(X^t)dt\Big).\end{aligned}$$

- Note that

$$\frac{d}{dt}f(X^t) = \sum_{i=1}^n \Big(\frac{X_i}{2\sqrt{t}} - \frac{Z_i}{2\sqrt{1-t}}\Big)\frac{\partial}{\partial x_i}f(X^t).$$

- Proof is completed by a sequence of tricky integration-by-parts steps.

How does one bound $\mathrm{Var}(T)$ in general?

Answer: By the gaussian Poincaré inequality. If $X_1, \ldots, X_n$ are i.i.d. $N(0,1)$, and $T : \mathbb{R}^n \to \mathbb{R}$ is absolutely continuous, then

$$\mathrm{Var}(T(X_1, \ldots, X_n)) \le \mathbb{E}\|\nabla T(X_1, \ldots, X_n)\|^2.$$

(Recall: $\nabla T = (\partial T/\partial x_1, \ldots, \partial T/\partial x_n)$ is the gradient of T.)

A general result

- Let $\mathcal{L}(c_1, c_2)$ be the class of probability measures on $\mathbb{R}$ that arise as laws random variables like $h(Z)$, where $Z \sim N(0,1)$ and $h \in C^2(\mathbb{R})$ satisfies
$$|h'(x)| \le c_1 \text{ and } |h''(x)| \le c_2.$$
- Let $\mathcal{L}$ be the class of all distributions that belong to $\mathcal{L}(c_1, c_2)$ for some finite c_1, c_2.
- In the next slide, we have a general normal approximation theorem for smooth functions of independent r.v.'s with laws in $\mathcal{L}$.

Only for the professionals...

Theorem

Let $X = (X_1, \ldots, X_n)$ be a vector of independent random variables in $\mathcal{L}(c_1, c_2)$ for some finite c_1, c_2. Take any $g \in C^2(\mathbb{R}^n)$ and let ∇g and $\nabla^2 g$ denote the gradient and Hessian of g. Let

$$\kappa_0 = \left(\mathbb{E}\sum_{i=1}^{n}\left|\frac{\partial g}{\partial x_i}(X)\right|^4\right)^{1/2},$$
$$\kappa_1 = (\mathbb{E}\|\nabla g(X)\|^4)^{1/4}, \text{ and}$$
$$\kappa_2 = (\mathbb{E}\|\nabla^2 g(X)\|^4)^{1/4}.$$

Suppose $W = g(X)$ has a finite fourth moment and let $\sigma^2 = \mathrm{Var}(W)$. Let Z be a normal random variable having the same mean and variance as W. Then

$$d_{TV}(W, Z) \leq \frac{2\sqrt{5}(c_1 c_2 \kappa_0 + c_1^3 \kappa_1 \kappa_2)}{\sigma^2}.$$

Applications to linear statistics of eigenvalues

- If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of a Wigner matrix with entries in $\mathcal{L}$ and $W = \sum_{i=1}^n f(\lambda_i)$, and let Z be a gaussian random variable with the same mean and variance as W. Then

$$d_{TV}(W, Z) \leq \frac{C(f)}{\sigma^2 \sqrt{n}},$$

  where $C(f)$ is an explicit constant depending on f.

- For the sample covariance matrix given by a $p \times N$ data matrix of i.i.d. entries, the corresponding bound is

$$\frac{C(f)(p \wedge N)}{\sigma^2 N^{3/2}}.$$

  (Bai and Silverstein have shown that when $p/N \to \alpha \in (0, 1)$, σ^2 converges to a positive limit.)

- Works for Double Wishart matrices and Gaussian Toeplitz matrices (new results).

Part II: Arbitrary functions of independent random variables

Notation

- Suppose $X = (X_1, \ldots, X_n)$ is a vector of independent random variables.
- Let $W = f(X)$ be an arbitrary function of X .
- Let X' be an independent copy of X.
- For each $A \subseteq \{1, \ldots, n\}$, define the vector X^A as

$$X_i^A = \begin{cases} X_i' & \text{if } i \in A, \\ X_i & \text{if } i \notin A. \end{cases}$$

Notation

- Let
$$\Delta_j f(X) = f(X) - f(X^j).$$
(Recall: $X^j = (X_1, \ldots, X_{j-1}, X_j', X_{j+1}, \ldots, X_n)$.)
- Define
$$T(X, X') = \frac{1}{2} \sum_A \frac{1}{\binom{n}{|A|}(n - |A|)} \sum_{j \notin A} \Delta_j f(X) \Delta_j f(X^A).$$

Our main result

Theorem

Suppose $\mathbb{E}(W) = 0$ *and let* $\sigma^2 = \mathbb{E}(W^2)$. *Then*

$$\mathcal{W}(\sigma^{-1}W, N(0,1)) \le \frac{\sqrt{\mathrm{Var}(\mathbb{E}(T|W))}}{\sigma^2} + \frac{1}{2\sigma^3}\sum_{j=1}^{n} \mathbb{E}|\Delta_j f(X)|^3,$$

where $\mathcal{W}$ *is the Wasserstein distance and* T *is defined as in the previous slide.*

(Recall: $\mathcal{W}(X,Y) = \sup\{|\mathbb{E}f(X) - \mathbb{E}f(Y)| : \|f\|_{\mathrm{Lip}} \le 1\}$.)

Simplest example

- Suppose $f(X) = n^{-1/2} \sum_{i=1}^{n} X_i$. Then
$$T(X, X') = \frac{1}{2n} \sum_{j=1}^{n} (X_j - X_j')^2.$$
Thus, $\mathrm{Var}(T(X, X')) = O(n^{-1})$.
- Also,
$$\sum_{j=1}^{n} \mathbb{E}|\Delta_j f(X)|^3 = n^{-3/2} \sum_{j=1}^{n} \mathbb{E}|X_j - X_j'|^3 = O(n^{-1/2}).$$
- Combining, we get an $O(n^{-1/2})$ error bound.

Quadratic forms

- Suppose $X_1, \ldots, X_n$ are i.i.d. symmetric ± 1-valued random variables.
- Let $\mathbf{A} = (a_{ij})$ be a real symmetric matrix of order n.
- Let $W = \sum_{i<j} a_{ij} X_i X_j$ and $\sigma^2 = \mathrm{Var}(W) = \frac{1}{2}\,\mathrm{Tr}(\mathbf{A}^2)$.
- Can compute: $\mathbb{E}(T|X) = \frac{1}{2} X^t \mathbf{A}^2 X$.

Putting $W' = (W - \mathbb{E}(W))/\sigma$, we have

$$\mathcal{W}(W', N(0,1)) \le \left(\frac{\mathrm{Tr}(\mathbf{A}^4)}{2\sigma^4}\right)^{1/2} + \frac{5}{2\sigma^3} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij}^2\right)^{3/2}.$$

This is slightly stronger than the best known results (Rotar '73, Hall '84, Götze & Tikhomirov '99, '02).

Proof of main theorem (brief sketch)

- For any g, f, A, and $j \notin A$,

$$\begin{aligned}
&\mathbb{E}(\Delta_j g(X)\Delta_j f(X^A)) \\
&= \mathbb{E}(g(X)\Delta_j f(X^A)) - \mathbb{E}(g(X^j)\Delta_j f(X^A)) \\
&= 2\mathbb{E}(g(X)\Delta_j f(X^A)) \text{ by exchangeability of } (X_j, X_j').
\end{aligned}$$

- With $g = \varphi \circ f$, we have $\Delta_j g(X) \approx \varphi'(f(X))\Delta_j f(X)$, and hence

$$\begin{aligned}
&\frac{1}{2}\mathbb{E}(\varphi'(f(X))\Delta_j f(X)\Delta_j f(X^A)) \\
&\approx \frac{1}{2}\mathbb{E}(\Delta_j g(X)\Delta_j f(X^A)) = \mathbb{E}(\varphi(f(X))\Delta_j f(X^A)).
\end{aligned}$$

- Thus,

$$\begin{aligned}
&\mathbb{E}(\varphi'(f(X))T(X,X')) \\
&\approx \mathbb{E}\biggl(\varphi(f(X))\sum_A \frac{1}{\binom{n}{|A|}(n-|A|)}\sum_{j\notin A}\Delta_j f(X^A)\biggr).
\end{aligned}$$

Proof of main theorem (brief sketch)

- Now note that

$$\sum_A \frac{1}{\binom{n}{|A|}(n-|A|)} \sum_{j \notin A} \Delta_j f(X^A) = f(X) - f(X'),$$

which is just an algebraic identity.

- Thus,

$$\begin{aligned}\mathbb{E}(\varphi'(f(X))T(X,X')) &\approx \mathbb{E}[\varphi(f(X))(f(X) - f(X'))] \\ &= \mathbb{E}(\varphi(f(X))f(X)).\end{aligned}$$

- Exact equality holds for $\varphi(u) = u$, which gives

$$\mathbb{E}(T(X,X')) = \mathrm{Var}(f(X)) = \sigma^2.$$

- Thus, if $\mathrm{Var}(T(X,X'))$ is tiny, then

$$\mathbb{E}(\varphi(f(X))f(X)) \approx \sigma^2 \mathbb{E}(\varphi'(f(X))),$$

which shows that $f(X) \dot{\sim} N(0,\sigma^2)$.

Again, how to bound $\mathrm{Var}(T)$ in general?

Very useful tool: Analog of Poincaré inequality, known as the *Efron-Stein* inequality in the statistical literature.

Theorem (Efron-Stein inequality)

Let $Z = f(Y_1, \ldots, Y_m)$ be a function of independent random objects $Y_1, \ldots, Y_m$. Let Y_i' be an independent copy of Y_i, $i = 1, \ldots, m$. Then

$$\begin{aligned}&\mathrm{Var}(Z)\\ &\leq \frac{1}{2}\sum_{i=1}^{m} \mathbb{E}\big[(f(Y_1, \ldots, Y_{i-1}, Y_i', Y_{i+1}, \ldots, Y_m) - f(Y_1, \ldots, Y_m))^2\big].\end{aligned}$$

A nearest neighbor problem

- Let $X_1, \ldots, X_n$ be i.i.d. random variables lying on a nice manifold of unknown dimension m.
- For a fixed $k \geq 2$, the Levina-Bickel estimate of m with tuning parameter k is given by

$$\hat{m}_k = \frac{1}{n} \sum_{\ell=1}^{n} \left(\frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{D_{\ell k}}{D_{\ell j}} \right)^{-1}, \tag{2}$$

  where $D_{\ell j}$ is the distance between X_ℓ and its j^{th} nearest neighbor.
- Under assumptions, $\hat{m}_k$ is consistent. How to prove a CLT?
- Existing results and techniques (Bickel-Breiman, Avram-Bertsimas, Penrose-Yukich, etc.) provide no immediate help.

A general nearest neighbor CLT

Theorem

Suppose $X_1, \ldots, X_n$ are i.i.d. $\mathbb{R}^d$-valued random vectors such that $\|X_1 - X_2\|$ is a continuous r.v. Fix $k \geq 1$, and suppose that for each i, f_i is a function of only X_i and its k nearest neighbors, and $W = n^{-1/2}\sum_i f_i$. Suppose for some $p \geq 8$, $\gamma_p := \max_i \mathbb{E}|f_i|^p$ is finite. Let $\sigma^2 = \mathrm{Var}(W)$ and $W' = (W - \mathbb{E}(W))/\sigma$. Then

$$\mathcal{W}(W', N(0,1)) \leq C\frac{\alpha(d)^3 k^4 \gamma_p^{2/p}}{\sigma^2 n^{(p-8)/2p}} + C\frac{\alpha(d)^3 k^3 \gamma_p^{3/p}}{\sigma^3 n^{(p-6)/2p}},$$

where $\alpha(d)$ is the minimum number of 60° cones at the origin required to cover $\mathbb{R}^d$, and C is a universal constant.

Summary

- Aim of the work: To reduce normal approximation problems to technically manageable variance bounding exercises. Gives explicit bounds.
- Variance bounds can be handled effectively by Poincaré or martingale inequalities.
- Applications till now: random matrix eigenvalues, nearest neighbor statistics, Komlós-Major-Tusnády strong embedding, Sherrington-Kirkpatrick model of spin glasses, etc.
- Future plan: (1) Work on other examples that I have in mind. (2) Find more applications. (3) Convince others to use the method.

A new approach to strong embeddings

Sourav Chatterjee

UC Berkeley

Coupling of random walks with Brownian motion

- Suppose $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random variables with $\mathbb{E}(\varepsilon_1) = 0$ and $\mathbb{E}(\varepsilon_1^2) = 1$. For each k, let

$$S_k = \sum_{i=1}^{k} \varepsilon_i.$$

- Goal is to construct a standard Brownian motion $(B_t)_{t \geq 0}$ on the same probability space so as to minimize the growth rate of

$$\max_{1 \leq k \leq n} |S_k - B_k|.$$

Skorokhod embeddings

- Skorokhod (1961), Strassen (1966): Start with $(B_t)_{t \geq 0}$, and construct stopping times $T_1 \leq T_2 \leq \cdots$ such that
  - B_{T_1} has the same distribution as ε_1.
  - $B_{T_1}, B_{T_2} - B_{T_1}, B_{T_3} - B_{T_2}, \cdots$ are i.i.d. In other words, $(B_{T_k})_{k \geq 1}$ has the same law as $(S_k)_{k \geq 1}$.
  - $E(T_{i+1} - T_i \mid (B_t)_{t \leq T_i}) = 1$ for each i.
- Show that $T_k \simeq k$ in an average sense, to bound $\max_{1 \leq k \leq n} |B_{T_k} - B_k|$.
- Strassen (1966) proved that

$$\max_{1 \leq k \leq n} |B_{T_k} - B_k| = O((n \log \log n)^{1/4} (\log n)^{1/2}),$$

  and conjectured that this is the best possible rate under $\mathbb{E}(\varepsilon_1^2) < \infty$. Proved by Kiefer (1969).
- Is it possible to improve, assuming stronger moment conditions on the summands?

The KMT embedding theorem

Indeed, yes. If ε_1 has a finite moment generating function in a neighborhood of zero, then one can get

$$\max_{k \le n} |S_k - B_k| = O(\log n).$$

Moreover, this is the best possible.

Theorem (Komlós-Major-Tusnády, 1975)

Let $\varepsilon_1, \varepsilon_2, \ldots$ be i.i.d. random variables with $\mathbb{E} \exp \theta |\varepsilon_1| < \infty$ for some $\theta > 0$. Let $S_k := \sum_{i=1}^k \varepsilon_i$, $k = 0, 1, \ldots$ be the corresponding random walk. It is possible to construct a version of the sequence $(S_k)_{k \ge 0}$ and a standard Brownian motion $(B_t)_{t \ge 0}$ on the same probability space such that for every n and every $t \ge 0$,

$$\mathbb{P}\big(\max_{k \le n} |S_k - B_k| \ge C \log n + t\big) \le K e^{-\lambda t},$$

where C, K, and λ do not depend on n.

Further developments

- Numerous applications in both applied and theoretical problems.
- Invaluable for understanding fine properties of simple random walk, e.g. in the works of Dembo-Peres-Rosen-Zetouni, Lawler, etc.
- There is a different KMT theorem for empirical processes, in the same paper. Of great interest to statisticians.
- No version yet for the non-i.i.d. case.
- Original paper still considered to be very hard to read. We will give a new (simpler?) proof for Bernoulli summands.

The first step

Lemma

Suppose $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are i.i.d. symmetric ± 1-valued r.v. Let $S_n = \sum_{i=1}^n \varepsilon_i$. Then it is possible to construct a version of S_n and a gaussian r.v. Z_n with mean 0 and variance n on the same probability space such that for all $t \geq 0$,

$$\mathbb{P}(|S_n - Z_n| \geq t) \leq K e^{-\lambda t},$$

where K and λ do not depend on n.

Question: Can we have a version of this in more complex normal approximation problems?

An abstraction

Lemma

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

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

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

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

- Key idea, inspired by Stein's method: $T \simeq \sigma^2 \implies W$ is approximately $N(0, \sigma^2)$. I call T a Stein coefficient of W.
- However, classical Stein's method can only give bounds on quantities like $\sup_{f \in \mathcal{F}} |\mathbb{E}f(W) - \mathbb{E}f(Z)|$, for various classes $\mathcal{F}$. The above result seems to be of a fundamentally different nature.

Examples

- Suppose X is a random variable with $\mathbb{E}(X) = 0$, $\mathbb{E}(X^2) < \infty$, and density ρ. Let
$$h(x) = \frac{\int_x^\infty y\rho(y)dy}{\rho(x)}.$$
- Then, by integration-by-parts, we have
$$\mathbb{E}(X\varphi(X)) = \mathbb{E}(\varphi'(X)h(X)).$$
Thus, $h(X)$ is a Stein coefficient for X.
- Suppose $X_1, \ldots, X_n$ are i.i.d. copies of X, and let $W = n^{-1/2}\sum_{i=1}^n X_i$. Then by the above result,
$$\begin{aligned}\mathbb{E}(W\varphi(W)) &= n^{-1/2}\sum_{i=1}^n \mathbb{E}(X_i\varphi(W)) \\ &= n^{-1}\sum_{i=1}^n \mathbb{E}(h(X_i)\varphi'(W)).\end{aligned}$$
- Thus, $n^{-1}\sum h(X_i)$ is a Stein coefficient for W.

A discrete example

- Suppose $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. symmetric ± 1-valued r.v. Let $S_n = \sum_{i=1}^n \varepsilon_i$.
- Let $Y \sim \mathrm{Uniform}[-1, 1]$. Let $W = S_n + Y$.
- Let
$$T = n - S_n Y + \frac{1 - Y^2}{2}.$$
- It follows from a calculation involving integration-by-parts that for all φ,
$$\mathbb{E}(W\varphi(W)) = \mathbb{E}(\varphi'(W)T),$$
that is, T is a Stein coefficient for W.
- Letting $\sigma^2 = n$, the abstract lemma tells us that it is possible to construct $Z \sim N(0, n)$ such that
$$\mathbb{E}\exp(\theta|W - Z|) \le 2\,\mathbb{E}\exp\left(\frac{2\theta^2(T - n)^2}{n}\right).$$
Since $T = n + O(\sqrt{n})$, this proves the lemma.

A general class of examples

- Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ be a vector of i.i.d. standard gaussian r.v.
- Let $W = f(\mathbf{X})$, where f is absolutely continuous. Suppose $\mathbb{E}(W) = 0$.
- Let $\mathbf{X}' = (X_1', \ldots, X_n')$ be an independent copy of $\mathbf{X}$.
- Let
$$T = \int_0^1 \frac{1}{2\sqrt{t}} \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{X}) \frac{\partial f}{\partial x_i}(\sqrt{1-t}\mathbf{X} + \sqrt{t}\mathbf{X}')dt.$$
- Then, one can show that T is a Stein coefficient for W.
- This can be used, e.g., to prove CLTs for linear statistics of eigenvalues of random matrices. (Different talk.)

Proof of abstract lemma: Step I

- Recall setup: (W, T) is a pair of r.v. such that for all φ,

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

Given σ^2, we are trying to construct $Z \sim N(0, \sigma^2)$ such that

$$\mathbb{E}\exp(\theta|W - Z|) \le 2\,\mathbb{E}\exp\left(\frac{2\theta^2(T - \sigma^2)^2}{\sigma^2}\right).$$

- Let $h(W) = \mathbb{E}(T \mid W)$. Then for all φ,

$$\mathbb{E}(W\varphi(W)) = \mathbb{E}(\varphi'(W)h(W)).$$

The function h characterizes the distribution of W via this equation.

- One can show that it suffices to construct (W, Z) such that $Z \sim N(0, \sigma^2)$ and for all $\theta > 0$,

$$\mathbb{E}\exp(\theta|W - Z|) \le 2\,\mathbb{E}\exp(2\theta^2(\sqrt{h(W)} - \sigma)^2).$$

Step II: Construction of the coupling

- Fix a function $r : \mathbb{R}^2 \to \mathbb{R}$.
- For $f \in C^2(\mathbb{R}^2)$, let

$$\mathcal{L}f(x,y) := h(x)\frac{\partial^2 f}{\partial x^2} + 2r(x,y)\frac{\partial^2 f}{\partial x \partial y} + \sigma^2 \frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

- Suppose there exists a probability measure μ on $\mathbb{R}^2$ such that for all f,

$$\int_{\mathbb{R}^2} \mathcal{L}f \, d\mu = 0. \qquad (*)$$

- We will show that every choice of r that allows a μ satisfying $(*)$ gives rise to a coupling of W and Z.

Step II contd.

- Recap: we have μ such that for all f

$$(*) \quad \int_{\mathbb{R}^2} \mathcal{L}f \; d\mu = 0, \quad \text{where}$$

$$\mathcal{L}f(x,y) := h(x)\frac{\partial^2 f}{\partial x^2} + 2r(x,y)\frac{\partial^2 f}{\partial x \partial y} + \sigma^2 \frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

- Let $(X, Y) \sim \mu$.
- Take any $\Phi \in C^2(\mathbb{R})$, and let $\varphi = \Phi'$. Putting $f(x,y) = \Phi(x)$ in $(*)$, we get

$$\mathbb{E}(h(X)\varphi'(X) - X\varphi(X)) = 0.$$

  Thus, X has the same law as W.
- Similarly, putting $f(x,y) = \Phi(y)$, we get $\mathbb{E}(Y\varphi(Y)) = \sigma^2\mathbb{E}(\varphi'(Y))$, and thus, $Y \sim N(0, \sigma^2)$.
- Note that these deductions are independent of the choice of $r(x,y)$, as long as μ exists. Each valid choice of $r(x,y)$ gives a coupling of W and Z.

Step III: The key lemma

Lemma

Suppose that the matrix valued function

$A(x, y) := \begin{pmatrix} h(x) & r(x, y) \\ r(x, y) & \sigma^2 \end{pmatrix}$ *is bounded, positive semidefinite, and continous everywhere. Then there exists a probability measure* μ *such that for all* $f \in C^2(\mathbb{R}^2)$ *satisfying certain mild conditions, we have*

$$(*) \quad \int_{\mathbb{R}^2} \mathcal{L}f \; d\mu = 0, \quad \textit{where}$$

$$\mathcal{L}f(x, y) := h(x)\frac{\partial^2 f}{\partial x^2} + 2r(x, y)\frac{\partial^2 f}{\partial x \partial y} + \sigma^2 \frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

Note that

- The operator $\mathcal{L}$ is not uniformly elliptic.
- The domain of f is unbounded.
- The functions h and r need not be Lipschitz.
- Can be solved by stochastic process techniques. Will come back to give a different, self-contained proof.

Step IV: Positive definite completion

- Thus, finding $r(x, y)$ is a problem of positive definite completion for the incomplete matrix

$$\begin{pmatrix} h(x) & ? \\ ? & \sigma^2 \end{pmatrix}.$$

- Intuition: the more nearly singular the completion, the tighter the coupling between the two coordinates.
- The 'most singular choice' is given by the geometric mean

$$r(x, y) = \sigma\sqrt{h(x)}.$$

- Another choice: $r(x, y) = h(x) \wedge \sigma^2$. This is not good: It brings us back to Skorokhod embeddings. (Will not elaborate now.)

Step V: Getting bounds

- With $r(x, y) = \sigma\sqrt{h(x)}$, we have

$$\mathcal{L}f(x, y) = h(x)\frac{\partial^2 f}{\partial x^2} + 2\sigma\sqrt{h(x)}\frac{\partial^2 f}{\partial x \partial y} + \sigma^2\frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$

- With $f(x, y) = \frac{1}{2k}(x - y)^{2k}$, we get

$$\mathcal{L}f(x, y) = (2k - 1)(x - y)^{2k-2}(\sqrt{h(x)} - \sigma)^2 - (x - y)^{2k}.$$

- So, if $\mathbb{E}(\mathcal{L}f(X, Y)) = 0$ for all f, then

$$\begin{aligned}\mathbb{E}(X - Y)^{2k} &= (2k - 1)\mathbb{E}((X - Y)^{2k-2}(\sqrt{h(X)} - \sigma)^2) \\ &\le (2k - 1)(\mathbb{E}(X - Y)^{2k})^{\frac{k-1}{k}}(\mathbb{E}(\sqrt{h(X)} - \sigma)^{2k})^{1/k}.\end{aligned}$$

- This gives

$$\mathbb{E}(X - Y)^{2k} \le (2k - 1)^k\mathbb{E}(\sqrt{h(X)} - \sigma)^{2k}.$$

The proof of the lemma is completed by summing over $k \ge 1$.

Proof of the key lemma

- Let A be a continous bounded map from $\mathbb{R}^2$ into the set of all 2×2 positive semidefinite matrices.
- Take any probability measure μ on $\mathbb{R}^2$. Suppose $\mathbf{X} \sim \mu$, and let $\mathbf{Z}$ be a standard gaussian random vector independent of $\mathbf{X}$.
- For each $\varepsilon > 0$, let $T_\varepsilon \mu$ be the law of the random vector

$$(1-\varepsilon)\mathbf{X} + \sqrt{2\varepsilon A(\mathbf{X})}\mathbf{Z},$$

  where $\sqrt{A}$ denotes the positive definite square root of A. Then T_ε is a continuous map.
- Suppose $\|A(\mathbf{x})\| \le b$ for all $\mathbf{x} \in \mathbb{R}^2$. Let K be the set of all probability measures on $\mathbb{R}^2$ satisfying

$$\int \mathbf{x} d\mu(\mathbf{x}) = 0 \quad \text{and} \quad \int \exp\langle \mathbf{u}, \mathbf{x}\rangle d\mu(\mathbf{x}) \le \exp(b\|\mathbf{u}\|^2) \text{ for all } \mathbf{u} \in \mathbb{R}^2.$$

  Easy to verify that K is non-empty, compact, and convex.
- Main observation: For any $0 < \varepsilon < 1$, $T_\varepsilon(K) \subseteq K$.
- So, by the Schauder-Tychonoff fixed point theorem for loally convex spaces, for every $\varepsilon \in (0,1)$, $\exists \mu_\varepsilon \in K$ such that $T_\varepsilon \mu_\varepsilon = \mu_\varepsilon$.

Proof of key lemma contd.

- Since K is compact, $\exists \mu \in K$ such that for some sequence $\varepsilon_n \downarrow 0$, $\mu_{\varepsilon_n} \to \mu$.
- Suppose $\mathbf{X}_\varepsilon \sim \mu_\varepsilon$. Using the identity $T_\varepsilon \mu_\varepsilon = \mu_\varepsilon$, we get that for any f,
$$\mathbb{E} f((1-\varepsilon)\mathbf{X}_\varepsilon + \sqrt{2\varepsilon A(\mathbf{X}_\varepsilon)}\mathbf{Z}) = \mathbb{E} f(\mathbf{X}_\varepsilon).$$
- Taylor expanding the LHS around $\mathbf{X}_\varepsilon$, dividing both sides by ε, and letting $\varepsilon \downarrow 0$ along $\{\varepsilon_n\}$, we end up with
$$(*) \quad \int_{\mathbb{R}^2} \mathcal{L} f \, d\mu = 0, \quad \text{where}$$
$$\mathcal{L} f(x,y) := A_{11}(x,y)\frac{\partial^2 f}{\partial x^2} + 2A_{12}(x,y)\frac{\partial^2 f}{\partial x \partial y} + A_{22}(x,y)\frac{\partial^2 f}{\partial y^2} - x\frac{\partial f}{\partial x} - y\frac{\partial f}{\partial y}.$$
- Thus, μ satisfies our quest and completes the proof of the lemma.

Completing the proof of the KMT theorem

We prove the following by induction on n.

Theorem

There exist positive universal constants C, K and λ_0 such that the following is true. Take any integer $n \geq 2$. Suppose $\varepsilon_1, \ldots, \varepsilon_n$ are exchangeable ± 1 random variables. For $k = 0, 1, \ldots, n$, let $S_k = \sum_{i=1}^{k} \varepsilon_i$ and let

$$W_k = S_k - \frac{k}{n} S_n.$$

It is possible to construct a version of $W_0, \ldots, W_n$ and a standard Brownian bridge $(\widetilde{B}_t)_{0 \leq t \leq 1}$ on the same probability space such that for any $0 < \lambda < \lambda_0$,

$$\mathbb{E} \exp(\lambda \max_{k \leq n} |W_k - \sqrt{n}\widetilde{B}_{k/n}|) \leq \exp(C \log n) \mathbb{E} \exp\left(\frac{K\lambda^2 S_n^2}{n}\right).$$

How to carry out the induction

The main ingredient is the following result, which is proved using the abstract lemma. (I won't go into more details.)

Theorem

Let $\varepsilon_1, \ldots, \varepsilon_n$ be n arbitrary elements of $\{-1, 1\}$. Let π be a uniform random permutation of $\{1, \ldots, n\}$. For each $1 \le k \le n$, let $S_k = \sum_{\ell=1}^{k} \varepsilon_{\pi(\ell)}$, and let

$$W_k = S_k - \frac{kS_n}{n}.$$

There exist universal constants $c > 1$ and $\theta_0 > 0$ satisfying the following. Take any $n \ge 3$, any possible value of S_n, and any $n/3 \le k \le 2n/3$. It is possible to construct a version of W_k and a gaussian random variable Z_k with mean 0 and variance $k(n-k)/n$ on the same probability space such that for any $\theta \le \theta_0$,

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

A few remarks about the intuition

- If $\mathbb{E}(X\varphi(X)) = \mathbb{E}(\varphi'(X)h(X))$ for all φ, then the law of X is an invariant measure for the diffusion

$$dX_t = -X_t dt + \sqrt{2h(X_t)}dB_t.$$

- Suppose we have independent diffusions

$$dX_t^i = -X_t^i dt + \sqrt{2h(X_t^i)}dB_t^i, \quad i = 1, \ldots, n.$$

- Let $W_t = n^{-1/2} \sum_{i=1}^n X_t^i$. Then

$$dW_t = -W_t dt + n^{-1/2} \sum_{i=1}^n \sqrt{2h(X_t^i)}dB_t^i.$$

Intuition contd.

- If we define another process $(\widetilde{B}_t)_{t\geq 0}$ by

$$d\widetilde{B}_t = \frac{\sum_{i=1}^n \sqrt{2h(X_t^i)}dB_t^i}{\sqrt{\sum_{i=1}^n 2h(X_t^i)}},$$

then $\widetilde{B}_t$ is again a standard Brownian motion, and

$$dW_t = -W_t dt + \sqrt{\frac{\sum_{i=1}^n 2h(X_t^i)}{n}}d\widetilde{B}_t.$$

- Since $\mathbb{E}(h(X)) = \mathbb{E}(X^2) =: \sigma^2$, therefore $n^{-1}\sum_{i=1}^n 2h(X_t^i) \simeq 2\sigma^2$ with high probability. Thus, W_t is 'approximately' an Ornstein-Uhlenbeck process.
- If we define $dZ_t = -Z_t dt + \sqrt{2}\sigma d\widetilde{B}_t$, then Z_t is an actual O-U process.
- Moreover, Z_t and W_t 'come close' at infinity. Our abstract lemma gets its hands on (W_∞, Z_∞).

An important question

Recall:

Lemma

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

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

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

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

Does there exist a multidimensional version of this lemma?

Immediate consequence: Direct proof of KMT treating the problem as a coupling of random vectors; possibly many other implications.

Spin glasses and Stein's method

Sourav Chatterjee

UC Berkeley

Spin glasses

- Magnetic materials with strange behavior.
- Not explained by ferromagnetic models like the Ising model.
- Theoretically studied since the 70's. An important example: Sherrington-Kirkpatrick model.
- High temperature phase: Thouless-Anderson-Palmer. Low temperature phase: Mézard and Parisi.
- Mathematically almost intractable until the late 90's.
  - Early results due to Aizenman, Lebowitz, Ruelle ('87), Fröhlich and Zegarliński ('87).
  - Notable papers due to Comets and Neveu ('95), Shcherbina ('99) etc.
  - Series of breakthroughs from Talagrand (1998 - present).
  - Groundbreaking contributions from Guerra & Toninelli ('02) and Guerra ('03), taken to completion by Talagrand in 2006.

    The Parisi formula. *Ann. Math. (2)* **163** no 1, 221–263.
  - Still, a lot of mysteries.

The Sherrington-Kirkpatrick model

- N spins. State space: $\Sigma_N = \{-1, 1\}^N$.
- Gibbs measure on Σ_N for the SK model:

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

  where
  - $(g_{ij})_{i<j\leq N}$ is a fixed realization of independent standard gaussian random variables (the disorder),
  - β and h are parameters,
  - Z_N is the normalizing constant (partition function).
- Notation:

$$\langle f(\sigma)\rangle := \sum_{\sigma\in\Sigma_N} f(\sigma)G_N(\sigma).$$

High temperature phase

- Let z be a standard gaussian r.v. Then

$$\lim_{N\to\infty} \frac{\log Z_N}{N} = \log 2 + \mathbb{E}\log\cosh(\beta z\sqrt{q} + h) + \frac{\beta^2}{4}(1-q)^2,$$

where q is determined by

$$q = \mathbb{E}\tanh^2(\beta z\sqrt{q} + h).$$

Rigorously proven to hold for $\beta < 1/2$, all h (Talagrand).

- Conjectured description of the full high temperature regime: All (β, h) such that

$$\beta^2 \mathbb{E}\frac{1}{\cosh^4(\beta z\sqrt{q} + h)} < 1.$$

The line where equality holds is called the Almeida-Thouless line.

The Gibbs measure at high temperature

- Spins are approximately independent, i.e. with high probability,

$$\langle \sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_k}\rangle \approx \langle \sigma_{i_1}\rangle\langle \sigma_{i_2}\rangle\cdots\langle \sigma_{i_k}\rangle.$$

  This is a deep fact and surprisingly hard to prove. Both sides are nondegenerate random variables in the limit.
- Moreover, $\langle \sigma_{i_1}\rangle, \ldots, \langle \sigma_{i_k}\rangle$ are approximately i.i.d.
- First proved by Talagrand using his cavity argument.
- If $h = 0$, then $\langle \sigma_i\rangle \equiv 0$ for all i.
- What if $h \neq 0$? No simple formulas for $\langle \sigma_1\rangle, \ldots, \langle \sigma_N\rangle$.

The TAP equations

When $(\beta, h) \in$ the high temperature regime, the random quantities $\langle \sigma_1 \rangle, \ldots, \langle \sigma_N \rangle$ satisfy

$$\langle \sigma_i \rangle \approx \tanh\biggl(\frac{\beta}{\sqrt{N}} \sum_{j \neq i} g_{ij} \langle \sigma_j \rangle + h - \beta^2 (1-q) \langle \sigma_i \rangle \biggr),$$

where q solves $q = \mathbb{E} \tanh^2(\beta z \sqrt{q} + h)$, z being a standard gaussian r.v.

- Discovered by Thouless, Anderson, and Palmer ('77).
- Rigorous proof by Talagrand ('03) using the cavity method.
- Unique solution if β is smaller than a constant.
- Talagrand also shows that $\langle \sigma_i \rangle$ converges in law to $\tanh(\beta z \sqrt{q} + h)$.

The cavity method

- Basically, a very complex induction over N.
- Coefficient of σ_N in the Hamiltonian is

$$\frac{\beta}{\sqrt{N}} \sum_{j=1}^{N-1} g_{Nj}\sigma_j + h.$$

- Replace this by an independent gaussian r.v. to get a new Hamiltonian.
- Determine the mean and variance of this r.v. such that the annealed measures remain 'approximately the same'.
- Using gaussian interpolation and a certain recursive argument, Talagrand shows that it is possible to do this if β is sufficiently small. This is the foundation of the cavity induction.
- We will follow a different route, starting from the next slide.

Explaining the TAP equations: (I) Local fields

- Local field at site i:
$$\ell_i = \frac{1}{\sqrt{N}} \sum_{j \neq i} g_{ij}\sigma_j.$$
- Easy to show: The conditional expectation of σ_i given $(\sigma_j)_{j\neq i}$ is exactly $\tanh(\beta\ell_i + h)$.
- This gives the equations
$$\langle \sigma_i \rangle = \langle \tanh(\beta\ell_i + h) \rangle.$$
- Thus, if we knew the limiting distribution of ℓ_i, $\langle \sigma_i \rangle$ could be computed directly. But this was not known previously.

Explaining the TAP equations: (II) The Onsager correction

- If X is a random variable with small variance, then

$$\mathbb{E}\tanh(aX + b) \approx \tanh(a\mathbb{E}(X) + b).$$

This is the usual mean-field approximation.

- Applying the naïve mean-field logic (although $\mathrm{Var}(\ell_i) \not\to 0$), we may wonder whether

$$\begin{aligned}\langle \sigma_i \rangle = \langle \tanh(\beta\ell_i + h)\rangle &\overset{?}{\approx} \tanh\big(\beta\langle \ell_i\rangle + h\big)\\ &= \tanh\bigg(\frac{\beta}{\sqrt{N}}\sum_{j\neq i} g_{ij}\langle \sigma_j\rangle + h\bigg).\end{aligned}$$

Not surprisingly, this is incorrect.

- In the TAP equations, we have

$$\langle \sigma_i \rangle \approx \tanh\bigg(\frac{\beta}{\sqrt{N}}\sum_{j\neq i} g_{ij}\langle \sigma_j\rangle + h - \beta^2(1-q)\langle \sigma_i\rangle\bigg).$$

The extra term is called the Onsager correction term in physics.

Onsager correction and mixture gaussians

- Let ϕ_{μ,σ^2} be the gaussian density with mean μ and variance σ^2. Let $\psi_{p,\mu_1,\mu_2,\sigma^2}$ be the mixture gaussian density

$$\psi_{p,\mu_1,\mu_2,\sigma^2} = p\phi_{\mu_1,\sigma^2} + (1-p)\phi_{\mu_2,\sigma^2}.$$

- Suppose $\mu_2 > \mu_1$, and we define

$$a = \frac{\mu_2 - \mu_1}{2\sigma^2},\ \ b = \frac{1}{2}\log\frac{p}{1-p} - \frac{\mu_2^2 - \mu_1^2}{4\sigma^2}.$$

  Then, if $X \sim \psi_{p,\mu_1,\mu_2,\sigma^2}$, then

$$\mathbb{E}\tanh(aX + b) = \tanh(a\mathbb{E}(X) + b - (2p-1)a^2\sigma^2).$$

- This is what happens! The local fields have mixture gaussian laws and the highlighted term is the Onsager correction term.

Our main result: Limit law of local fields

- Recall: The local field at site i is defined as

$$\ell_i := \frac{1}{\sqrt{N}} \sum_{j \neq i} g_{ij} \sigma_j.$$

- For each i, let r_i be the random variable

$$r_i := \frac{1}{\sqrt{N}} \sum_{j \neq i} g_{ij} \langle \sigma_j \rangle - \beta(1-q)\langle \sigma_i \rangle.$$

- Let ν_i be the (random) mixture gaussian probability measure with density function

$$p_i \phi_{r_i + \beta(1-q), 1-q} + (1-p_i)\phi_{r_i - \beta(1-q), 1-q},$$

where

$$p_i = \frac{e^{\beta r_i + h}}{e^{\beta r_i + h} + e^{-\beta r_i - h}}.$$

- Under the Gibbs measure, ν_i approximates the law of ℓ_i for large N.

Our main result: Limit law of local fields

Theorem

Suppose (β, h) *is in the high temperature regime and* q *satisfies*

$$q = \mathbb{E}\tanh^2(\beta z\sqrt{q} + h),$$

where z *is a standard gaussian r.v. Let* $\ell_1, \ldots, \ell_N$ *be the local fields and let* $\nu_1, \ldots, \nu_N$ *be defined as in the previous slide. Then for any bounded measurable* $f : \mathbb{R} \to \mathbb{R}$ *and any* $1 \le i \le N$*, we have*

$$\mathbb{E}\biggl(\bigl\langle f(\ell_i)\bigr\rangle - \int_{\mathbb{R}} f(x)\nu_i(dx)\biggr)^2 \le \frac{C(\beta, h)\|f\|_\infty^2}{\sqrt{N}},$$

where $C(\beta, h)$ *is a constant depending only on* β *and* h*.*

- Recall: $\langle \sigma_i \rangle = \bigl\langle \tanh(\beta\ell_i + h)\bigr\rangle$. Taking $f(x) = \tanh(\beta x + h)$, the TAP equations follow easily.
- Proof is by Stein's method.

Stein's method

- Let X and Z be two random variables. Suppose we want to show that they approximately have the same distribution.
- Basic steps in Stein's method:
  1. Identify an operator T such that for all functions h,

     $$\mathbb{E}\big(Th(Z)\big) = 0.$$

     (T is called a Stein characterizing operator.) For example, if Z is standard gaussian, then $Th(x) = h'(x) - xh(x)$ is a characterizing operator.
  2. Given a function f, find h such that

     $$Th(x) = f(x) - \mathbb{E}(f(Z)).$$

     Relate the properties of h to those of f.
  3. By the definition of h it follows that

     $$|\mathbb{E}f(X) - \mathbb{E}f(Z)| = |\mathbb{E}(Th(X))|.$$

     Compute a bound on $|\mathbb{E}(Th(X))|$ by whatever means possible.

Stein's method for gaussian approximation

- Let Z be a standard gaussian r.v. Then for all h,

$$\mathbb{E}(h'(Z) - Zh(Z)) = 0.$$

  Thus, $Th(x) := h'(x) - xh(x)$ is a characterizing operator for the standard gaussian distribution.
- Thus, to show that a r.v. W is approximately standard gaussian, one has to show that for all h,

$$\mathbb{E}(h'(W) - Wh(W)) \approx 0.$$

Stein's method for mixture gaussians

- For the mixture gaussian density

$$p\phi_{\mu_1,\sigma^2} + (1-p)\phi_{\mu_2,\sigma^2}$$

the characterizing operator is

$$Th(x) = h'(x) - \left(\frac{x-\mu}{\sigma^2} - a \cdot \tanh(ax+b)\right)h(x),$$

where μ, a, b are defined as

$$\mu = \frac{\mu_1+\mu_2}{2}, \quad a = \frac{\mu_2-\mu_1}{2\sigma^2},$$
$$b = \frac{1}{2}\log\frac{p}{1-p} - \frac{\mu_2^2-\mu_1^2}{4\sigma^2}.$$

- Appears to be naturally connected to physical models. Does not occur in the literature on Stein's method.

The Approximation Lemma

Lemma

Suppose $g = (g_1, \ldots, g_n)$ is a collection of independent standard gaussian random variables, and $h_1, \ldots, h_n$ are absolutely continuous functions of g. Then

$$\mathbb{E}\biggl(\sum_{i=1}^n g_i h_i - \sum_{i=1}^n \frac{\partial h_i}{\partial g_i}\biggr)^2 = \sum_{i=1}^n \mathbb{E}(h_i^2) + \sum_{i,j=1}^n \mathbb{E}\biggl(\frac{\partial h_i}{\partial g_j}\frac{\partial h_j}{\partial g_i}\biggr).$$

Idea: If the right hand side is small, then the lemma 'generates' the equation

$$\sum_{i=1}^n g_i h_i \approx \sum_{i=1}^n \frac{\partial h_i}{\partial g_i}.$$

Can be used to obtain Stein characterizing equations for highly complex functions of gaussian random variables.

Example

- Consider the SK model with zero external field (the simple case):

$$G_N(\sigma) = Z_N^{-1} \exp\left(\frac{\beta}{\sqrt{N}} \sum_{i<j\le N} g_{ij}\sigma_i\sigma_j \right).$$

- We wish to generate a characterizing equation for the local field at site 1:

$$\ell_1 = \frac{1}{\sqrt{N}} \sum_{j=2}^{N} g_{1j}\sigma_j.$$

- Fix a smooth function f. For each $j = 2, \ldots, N$, let

$$h_j = \frac{1}{\sqrt{N}} \langle \sigma_j f(\ell_1) \rangle.$$

- Then

$$\sum_{j=2}^{N} g_{1j} h_j = \langle \ell_1 f(\ell_1) \rangle.$$

Example contd.

- On the other hand, an easy computation gives

$$\frac{\partial h_j}{\partial g_{1j}} = \frac{\langle f'(\ell_1)\rangle + \beta\langle \sigma_1 f(\ell_1)\rangle - \beta\langle \sigma_j f(\ell_1)\rangle\langle \sigma_1\sigma_j\rangle}{N}.$$

- The 2nd term. Note that ℓ_1 does not depend on σ_1, and the conditional expectation of σ_1 given $\sigma_2, \ldots, \sigma_N$ is $\tanh(\beta\ell_1)$. Thus,

$$\langle \sigma_1 f(\ell_1)\rangle = \langle \tanh(\beta\ell_1) f(\ell_1)\rangle.$$

- The 3rd term. It follows from the high temperature theory for $\beta < 1$ that for $2 \le j \le N$,

$$\langle \sigma_1\sigma_j\rangle \approx \langle \sigma_1\rangle\langle \sigma_j\rangle = 0.$$

Example contd.

- Thus, if we can apply the Approximation Lemma, we get the approximation

$$\sum_{j=2}^{N} g_{1j} h_j \approx \sum_{j=2}^{N} \frac{\partial h_j}{\partial g_{1j}}$$

which is equivalent to

$$\big\langle \ell_1 f(\ell_1) - f'(\ell_1) - \beta \tanh(\beta \ell_1) f(\ell_1) \big\rangle \approx 0.$$

- Now, the operator

$$Tf(x) = xf(x) - f'(x) - \beta \tanh(\beta x) f(x)$$

is a Stein characterizing operator for the mixture gaussian density

$$\tfrac{1}{2}\phi_{\beta,1} + \tfrac{1}{2}\phi_{-\beta,1}.$$

- This procedure 'discovers' the limiting distribution of ℓ_1.

When $h \neq 0$

- When $h \neq 0$, ℓ_1 does not have a nonrandom limiting distribution. The situation becomes more complex.
- Given a function u, we start with a solution $f(x,y)$ of the p.d.e.

$$\frac{\partial f}{\partial x}(x,y) - \left(\frac{x-y}{\sigma^2} - \beta \tanh(\beta x + h)\right) f(x,y)$$
$$= u(x) - \int_{\mathbb{R}} u(t) \frac{\cosh(\beta t + h) e^{-\frac{(t-y)^2}{2(1-q)}}}{\sqrt{2\pi(1-q)}\cosh(\beta\mu + h) e^{\frac{1}{2}\beta^2(1-q)}} dt.$$

- Then, defining

$$r_1 = \frac{1}{\sqrt{N}} \sum_{j=2}^{N} g_{1j}\langle \sigma_j \rangle - \beta(1-q)\langle \sigma_1 \rangle,$$

we let

$$h_j = \frac{1}{\sqrt{N}} \big\langle \big(\sigma_j - \langle \sigma_j \rangle\big) f(\ell_1, r_1) \big\rangle.$$

- The proof is completed by an application of the Approximation Lemma with these h_j's.

Distribution of $\langle \sigma_1 \rangle$

- By the TAP equations,

$$\langle \sigma_1 \rangle \approx \tanh(\beta r_1 + h),$$

where

$$r_1 = \frac{1}{\sqrt{N}} \sum_{j=2}^{N} g_{1j} \langle \sigma_j \rangle - \beta(1-q)\langle \sigma_1 \rangle.$$

- Thus, it suffices to find the limiting distribution of r_1.
- Repeatedly integrating by parts, we can get

$$\mathbb{E}(r_1 f(r_1)) \approx q\mathbb{E}\big(f'(r_1)\langle \eta_1 \rangle\big)$$

where

$$\eta_1 = 1 + \frac{\beta\sigma_1}{\sqrt{N}} \sum_{j=2}^{N} g_{1j}\big(\sigma_j - \langle \sigma_j \rangle\big) - \beta^2(1-q)\big(1 - \langle \sigma_1 \rangle^2\big).$$

Distribution of $\langle \sigma_1 \rangle$ contd.

- Applying the Approximation Lemma with

$$h_j = \frac{\beta}{\sqrt{N}}\big(\langle \sigma_1 \sigma_j \rangle - \langle \sigma_1 \rangle \langle \sigma_j \rangle\big),$$

  we can show that $\langle \eta_1 \rangle \approx 1$.

- Combined with the earlier approximation

$$\mathbb{E}(r_1 f(r_1)) \approx q\mathbb{E}\big(f'(r_1)\langle \eta_1 \rangle\big),$$

  this shows that $\mathbb{E}(r_1 f(r_1)) \approx q\mathbb{E}(f'(r_1))$.

- Thus, r_1 is a gaussian r.v. with mean zero and variance q in the large N limit.

Summary and future directions

- Often in spin glasses and other models we have the situation that for some random quantity X,

$$\mathbb{E}\tanh(aX+b) = \tanh(a\mathbb{E}(X) + b + \text{a correction term}).$$

- We show that this happens if the distribution of X is a mixture of two gaussian densities.
- Following this line, we derive the TAP equations for the SK model by showing that the local field is indeed a mixture of two gaussians in the limit.
- Gives total variation error bounds.
- The key tool is the Approximation Lemma that generates characterizing equations for Stein's method.
- The Approximation Lemma can possibly be used to derive/discover limiting distributions of other objects.
- Results apply only to the high temperature phase. Possible to extend to low temperature???

Stein's method for concentration inequalities

Sourav Chatterjee

Concentration inequalities

- Concentration inequalities give bounds on

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \geq x\}$$

  where X is some random variable, usually high dimensional, and f is a well behaved (usually Lipschitz) function.
- Useful in a variety of fields. Important tool in combinatorics, machine learning and theoretical computer science.
- Huge literature. Powerful results available for functions of independent random variables.
- Not so much known in dependent settings.

Stein's method for concentration inequalities?

- **Question:** Can we use exchangeable pairs to get concentration inequalities?
- Attempted by Stein in his book without success.
- Suppose (X, X') is an exchangeable pair of random variables on some space, $F(X, X')$ is an antisymmetric function and $f(X) = \mathbb{E}(F(X, X')|X)$.
- Then for any g,

$$\mathbb{E}(g(X)f(X)) = \mathbb{E}(g(X)F(X, X')).$$

- Exchangeability $\Rightarrow$

$$\mathbb{E}(g(X)F(X, X')) = \mathbb{E}(g(X')F(X', X)).$$

- Antisymmetry $\Rightarrow$

$$\mathbb{E}(g(X')F(X', X)) = -\mathbb{E}(g(X')F(X, X')).$$

- Combining, we get

$$\mathbb{E}(g(X)f(X)) = \frac{1}{2}\mathbb{E}((g(X) - g(X'))F(X, X')).$$

- Taking $g = f$, we have the variance formula:

$$\mathrm{Var}(f(X)) = \frac{1}{2}\mathbb{E}((f(X) - f(X'))F(X, X')).$$

- Example: Magnetization in Curie-Weiss model.

Curie-Weiss model

- Configuration space of n spins: $\{-1, 1\}^n$.
- Probability mass on this space:

$$p_\beta(\sigma) = Z(\beta)^{-1} \exp\left(\frac{\beta}{n} \sum_{i<j} \sigma_i \sigma_j\right).$$

  Here β is a parameter and $Z(\beta)$ is the normalizing constant.
- Magnetization: $m(\sigma) = \frac{1}{n} \sum_{i=1}^n \sigma_i$.
- Well-known: $m(\sigma) \approx \tanh(\beta m(\sigma))$ with high probability.

- Say two configurations are “neighbors” if they differ at exactly one site.
- Get σ' from σ by taking one step in the Gibbs sampler chain along a random neighbor.
- Set $F(\sigma, \sigma') = n(m(\sigma) - m(\sigma'))$. Then

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

  where $m_i(\sigma) = \frac{1}{n}\sum_{j\neq i} \sigma_j$.
- Then $|F(\sigma, \sigma')| \leq 2$ and $|f(\sigma) - f(\sigma')| \leq 2(1+\beta)/n$.

- Thus,

$$\begin{aligned}
&\mathrm{Var}\left(m(\sigma) - \frac{1}{n}\sum_{i=1}^{n} \tanh(\beta m_i(\sigma)) \right) \\
&= \frac{1}{2}\mathbb{E}((f(\sigma) - f(\sigma'))F(\sigma, \sigma')) \\
&\le \frac{2(1+\beta)}{n}.
\end{aligned}$$

- Finally, note that $|m_i(\sigma) - m(\sigma)| \le 1/n$. Combining, we get

$$\mathbb{E}(m(\sigma) - \tanh(\beta m(\sigma)))^2 \le \frac{2(1+\beta)}{n} + \frac{\beta^2}{n^2}.$$

Tail bounds

We have the following moment and tail inequality version of the earlier variance formula:

Theorem (Chatterjee '05)

Define

$$\Delta(X) := \frac{1}{2}\mathbb{E}\big(|(f(X) - f(X'))F(X,X')|\big|X\big).$$

Then for any positive integer p, we have

$$\|f(X) - \mathbb{E}f(X)\|_{2p}^2 \le (2p-1)\|\Delta(X)\|_p.$$

Moreover, if $|\Delta(X)| \le C$ almost surely, then

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \ge x\} \le 2e^{-x^2/2C}$$

for each $x \ge 0$.

(Recall that: (X, X') is an exchangeable pair, F is antisymmetric, and $\mathbb{E}(F(X,X')|X) = f(X) - \mathbb{E}f(X)$.)

Proof

- Let $\varphi(\theta) = \mathbb{E}(e^{\theta f(X)})$.
- Using the same trick as before, we have

$$\begin{aligned}\varphi'(\theta) &= \mathbb{E}(e^{\theta f(X)} f(X)) \\ &= \frac{1}{2}\mathbb{E}((e^{\theta f(X)} - e^{\theta f(X')})F(X, X')).\end{aligned}$$

- Using $|e^x - e^y| \leq \frac{1}{2}(e^x + e^y)|x - y|$, we get

$$\varphi'(\theta) \leq |\theta|\mathbb{E}(e^{\theta f(X)}\Delta(X)),$$

where

$$\Delta(X) = \frac{1}{2}\mathbb{E}\big(|(f(X) - f(X'))F(X, X')|\big|X\big).$$

- Similarly,

$$\mathbb{E}(f(X)^{2p}) \leq (2p-1)\mathbb{E}(f(X)^{2p-2}\Delta(X))$$

and apply Hölder's inequality.

'Self-bounded functions'

Before we give an example, let us state a refinement of the previous tail bound:

Theorem (Chatterjee '05)

Suppose $\Delta(X) \leq Bf(X) + C$ *a.s. Then*

$$\mathbb{P}\{|f(X) - \mathbb{E}f(X)| \geq x\} \leq 2e^{-x^2/(2C+2Bx)}$$

for each $x \geq 0$.

While the moment bounds are generalizations of the Burkholder-Gundy-Davis inequalities, and the first tail bound generalizes the Hoeffding inequality, the above can be seen as an exchangeable pair version of Bernstein's inequality.

Example: Random permutations

Proposition

Let $\{a_{ij}\}$ *be an* n *by* n *array of elements of* $[0,1]$. *Let* π *be a random (uniform) permutation of* $\{1,\ldots,n\}$, *and let* $W = \sum_{i=1}^n a_{i\pi(i)}$. *Then for any* $x \geq 0$,

$$\mathbb{P}\{|W - \mathbb{E}(W)| \geq x\} \leq 2e^{-x^2/(4\mathbb{E}(W)+2x)}.$$

For instance, if $a_{ij} = \mathbb{I}\{i = j\}$, then W is the number of fixed points of π and $\mathbb{E}(W) = 1$.

Sketch of Proof:

- Obtain π' by applying a random transposition to π. Let $W' = W(\pi')$.
- Let $F(\pi,\pi') = \frac{1}{2}n(W - W')$.
- Easy: $\mathbb{E}(F(\pi,\pi')|\pi) = W - \mathbb{E}(W) =: f(\pi)$.

Proof (contd.)

- Using $0 \leq a_{ij} \leq 1$, we can show

$$\begin{aligned}\Delta(\pi) &= \frac{1}{2}\mathbb{E}\big(|(f(\pi) - f(\pi'))F(\pi,\pi')|\big|\pi\big)\\ &= \frac{n}{4}\mathbb{E}((W - W')^2|\pi)\\ &\leq f(\pi) + 2\mathbb{E}(W).\end{aligned}$$

- Use Theorem.

A further refinement

Lemma (Dey '08)

Suppose (X, X') *be an exchangeable pair of random variables. Let* $F(X, X'), f(X)$ *and* $\Delta(X)$ *be as before. Suppose that for some real number* $\alpha \in [0, 1]$ *we have*

$$\Delta(X) \leq B\,|f(X)|^{\alpha} + C$$

a.s. where $B > 0, C \geq 0$ *are constants. Assume that*

$$\mathbb{E}(e^{\theta f(X)}|F(X, X')|) < \infty \text{ for all } \theta.$$

Then for any $t \geq 0$*, we have*

$$\mathbb{P}(|f(X)| > t) \leq 2^{2/\alpha} \exp\left(-\frac{t^{2-\alpha}}{16 \max\{B, C\}}\right).$$

Application: Curie Weiss model at Criticality ($\beta = 1$)

- The Curie-Weiss model at inverse temperature β and zero external field is given by the following Gibbs measure on $\{+1, -1\}^n$.
- For a typical spin configuration $\boldsymbol{\sigma}$ of the complete graph on n vertices, the probability of $\boldsymbol{\sigma}$ is

$$\mathbb{P}(\{\boldsymbol{\sigma}\}) := Z_\beta^{-1} \exp\left(\frac{\beta}{n}\sum_{i<j}\sigma_i\sigma_j\right).$$

- There is a phase transition at $\beta_c = 1$ w.r.t. the magnetization $m(\boldsymbol{\sigma}) := \frac{1}{n}\sum_{i=1}^n \sigma_i$.

Constructing the exchangeable pair

- Consider the Glauber dynamics, *i.e.*, Given a configuration $\boldsymbol{\sigma}$ get a new configuration $\boldsymbol{\sigma}'$ as follows: choose a coordinate I u.a.r. from $[n]$ and replace σ_I by a random spin drawn from the conditional distribution given the spin of all the neighbors.
- We have $\mathbb{E}[\sigma_i \mid \sigma_j, j \neq i] = \tanh(m_i)$ where

$$m_i = m_i(\boldsymbol{\sigma}) = n^{-1} \sum_{j \neq i} \sigma_j = m(\boldsymbol{\sigma}) - \sigma_i / n.$$

- Define the antisymmetric function

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

Bounding the Δ function

- Clearly

$$f(\boldsymbol{\sigma}) = \mathbb{E}[F(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \mid \boldsymbol{\sigma}] = \frac{1}{n}\sum_{i=1}^{n}(\sigma_i - \tanh m_i(\boldsymbol{\sigma}))$$
$$= m - \tanh(m) + O(n^{-1}).$$

- Using Taylor expansion we have

$$n|f(\boldsymbol{\sigma}) - f(\boldsymbol{\sigma}')| \leq 2(1 - |\text{sech} m(\boldsymbol{\sigma})|^2) + 6n^{-1}.$$

- Now $\tanh|m| \leq |m|$ and $|m|^3 \leq 5|m - tanh(m)|$ implies that

$$|\Delta(\boldsymbol{\sigma})| \leq 6n^{-1}|f(\boldsymbol{\sigma})|^{2/3} + 12n^{-5/3}.$$

Concentration

- Now applying the lemma with $\alpha = 2/3, B = 6n^{-1}$ we have

$$\mathbb{P}(|m| \geq t) \leq \mathbb{P}(|m - \tanh(m)| \geq t^3/5) \leq 8e^{-bnt^{3\cdot(2-\alpha)}} = 8e^{-bnt^4}$$

for some constant b.

Theorem (Chatterjee & Dey '08)

Suppose $\boldsymbol{\sigma}$ is drawn from the Curie-Weiss model at the critical temperature $\beta = 1$. Then, for any $n \geq 1$ and $t \geq 0$ the magnetization satisfies

$$\mathbb{P}(|n^{1/4} m(\boldsymbol{\sigma})| \geq t) \leq 8 \exp(-bt^4)$$

for some absolute constant $b > 0$.

An application to large deviations

For each $r, s \in (0,1)$, let

$$I(r,s) := r \log \frac{r}{s} + (1-r) \log \frac{1-r}{1-s}.$$

Theorem (Chatterjee & Dey '08)

Let T_n be the number of triangles in $G(n,p)$, where $p > p_0$, with $p_0 = 2/(2+e^{3/2}) \approx 0.31$. Then for any $r \in (p,1]$,

$$\mathbb{P}\left(T_n \geq \binom{n}{3} r^3\right) = \exp\left(-\frac{n^2 I(r,p)}{2}(1 + O(n^{-1/2}))\right).$$

Moreover, even if $p \leq p_0$, there exist p', p'' such that $p < p' \leq p'' < 1$ and the same result holds for all $r \in (p,p') \cup (p'',1]$.

Remarks

- Long history of concentration inequalities for subgraph counts. The problem of the upper tail is one of the well known hard problems in random graph theory.
- Best available results for triangles are due to Kim and Vu (2004).
- Our theorem gives the first exact computation of the rate function. Only upper and lower bounds were available till now.
- We do not say anything about sparse graphs.

Approach

- We have to consider a Gibbs measure on the space of graphs on n vertices, with Hamiltonian
$$\exp\left(\beta \frac{\#\text{Triangles}}{n} + h\#\text{Edges} - \psi_n(\beta, h)\right).$$
- If $\beta = 0$, this is just $G(n,p)$ with $p = e^h/(1+e^h)$.
- If $\beta \neq 0$, $\psi_n(\beta, h) = ?$

Solution in a "high temperature regime"

- Define $g : [0,1] \to \mathbb{R}$ as

$$g(x) = \frac{e^{\beta x^2 + h}}{1 + e^{\beta x^2 + h}}.$$

- **Theorem.** *If* $(\beta, h) \in$ *a certain 'high temperature region', there is a unique* $p_* = p_*(\beta, h) \in [0,1]$ *such that* $p_* = g(p_*)$*, and*

$$\lim_{n\to\infty} \frac{\psi_n(\beta, h)}{n^2} = \frac{\beta p_*^3}{6} + \frac{h p_*}{2} - \frac{p_* \log p_*}{2} - \frac{(1-p_*)\log(1-p_*)}{2}.$$

- Physical solution by Juyong and Newman ('05). Predicts phase transition.

Sketch of proof

- Suppose a random graph is drawn from the Gibbs measure. Let

$$X_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ is an edge} \\ 0 & \text{otherwise.} \end{cases}$$

- Let $L_{ij} = \frac{1}{n}\sum_{k \notin \{i,j\}} X_{ik}X_{jk}$.
- Main step: A temperature free result. Let

$$\varphi(x) = \frac{e^{\beta x + h}}{1 + e^{\beta x + h}}.$$

Then with high probability,

$$L_{ij} \approx \frac{1}{n}\sum_{k \notin \{i,j\}} \varphi(L_{ik})\varphi(L_{jk}) \quad \text{for all } i,j,$$

irrespective of β *and* h.
- Proof of the main step: By defining an appropriate exchangeable pair and an antisymmetric function and applying the theorem.

Sketch of proof contd.

- Using similar techniques, can also show other temperature free results:

$$\#\text{Edges} \sim \sum_{i<j} \varphi(L_{ij})$$

and

$$\#\text{Triangles} \sim \sum_{i<j<k} \varphi(L_{ij})\varphi(L_{jk})\varphi(L_{ki}).$$

- Finally, when $(\beta, h) \in$ a certain 'high temperature region', the system of equations

$$x_{ij} = \frac{1}{n} \sum_{k \notin \{i,j\}} \varphi(x_{ik})\varphi(x_{jk}),\ 1 \leq i < j \leq n$$

has a unique solution at $x_{ij} \approx \varphi^{-1}(p_*)$ for all i, j.

Sketch of proof contd.

- Using monotonicity of φ, it is possible to conclude that with high probability, $L_{ij} \approx \varphi^{-1}(p_*)$ for all i, j.
- Since the Hamiltonian is

$$\frac{\beta}{6} \sum X_{ij} L_{ij} + \frac{h}{2} \sum X_{ij},$$

  this allows us to 'replace' it by a linear function of the X_{ij}'s (i.e. independence).
- Resulting model is quite different, but certain characteristics are approximately the same, e.g. normalizing constant, correlations, central limit theorems, etc.