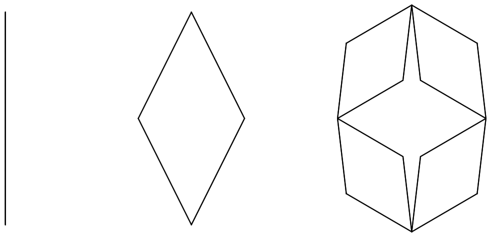
Zero Biasing and the Diamond Lattice

Larry Goldstein

http://math.usc.edu/∼larry

The Diamond Lattice: Three Scales



Properties (e.g. conductance) at level 2 depend on properties at level 1, ...

$$X_2 = F(\mathbf{X}_1), \qquad X_1 = F(\mathbf{X}_0)\ldots$$

Hierarchical Models

$$X_{n+1} = F(\mathbf{X}_n) \quad where \quad \mathbf{X}_n = (X_{n,1}, \ldots, X_{n,k})^\mathsf{T}$$

with $X_{n,i}$ independent, each with distribution X_n.

Conditions on F, due to Shneiberg, Li and Rogers, and Wehr, imply the weak law (here assumed)

$$X_n \to_p c,$$

and by Woo and Wehr which imply

$$W_n \to_d \mathcal{N}(0,1), \quad \text{for} \quad W_n = \frac{X_n - EX_n}{\sqrt{\mathsf{Var}(X_n)}}.$$

Classical Central Limit Theorem as Hierarchical Model

Taking F to give the *average*

$$F(x_1, x_2) = \frac{x_1 + x_2}{2}$$

gives in distribution

$$X_n = \frac{X_{0,1} + \cdots + X_{0,2^n}}{2^n}.$$

At stage n there are $N = 2^n$ variables, would expect a bound to the normal Z of the form

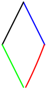
$$d(W_n, Z) \le C\gamma^n \quad \text{where} \quad \gamma^n = N^{-1/2} = (1/\sqrt{2})^n.$$

Averaging Functions

We say F is (strictly) averaging

1. $\min_i x_i \le F(\mathbf{x}) \le \max_i x_i$, and strictly when $\min_i x_i < \max_i x_i$.
2. $F(\mathbf{x}) \le F(\mathbf{y})$ whenever $x_i \le y_i$, and strictly when $x_j < y_j$ for some j.

Say F is scaled averaging when $F(\mathbf{x})/F(\mathbf{1}_k)$ is averaging, where $\mathbf{1}_k = (1, \ldots, 1)$.



Diamond Lattice Conductivity Function

Parallel and series resistor combination rules

$$L_1(x_1, x_2) = x_1 + x_2, \quad L_{-1}(x_1, x_2) = (x_1^{-1} + x_2^{-1})^{-1}$$

gives the weighted $w_i > 0$ diamond lattice conductivity function

$$F(\mathbf{x}) = \left(\frac{1}{w_1 x_1} + \frac{1}{w_2 x_2}\right)^{-1} + \left(\frac{1}{w_3 x_3} + \frac{1}{w_4 x_4}\right)^{-1},$$

a scaled strictly averaging function.

Approximate Linear Recursion

Write $X_{n+1} = F(\mathbf{X}_n)$ as a linear recursion with ('small') perturbation R_n,

$$X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n + R_n, \quad n \geq 0,$$

where $c_n = EX_n$, $\boldsymbol{\alpha}_n = F'(\mathbf{c}_n)$, $\mathbf{c}_n = (c_n, \ldots, c_n)^\mathsf{T} \in \mathbb{R}^k$, and F' the gradient of F.

Rule out trivial cases such as $F(x_1, x_2) = x_1$; when $\boldsymbol{\alpha} = F'(\mathbf{c})$ at limiting c is not a multiple of a standard basis vector, then $\lambda = ||\boldsymbol{\alpha}|| < 1$ when F is averaging.

Stein's Method for Normal

$$Z \sim \mathcal{N}(0, \sigma^2) \quad \text{if and only if} \quad \sigma^2 E f'(Z) = E Z f(Z).$$

For $EW = 0, EW^2 = \sigma^2$, if $E[\sigma^2 f'(W) - W f(W)]$ is close to zero for enough f, then W should be close to Z in distribution. Given a test function h, let $Nh = Eh(Z/\sigma)$, and solve for f in the Stein equation

$$\sigma^2 f'(w) - w f(w) = h(w/\sigma) - Nh.$$

Now evaluate expectation of RHS by expectation of LHS.

Zero Bias Transformation

Goldstein and Reinert 1997: For W a mean zero variance σ^2 random variable, there exists W^* such that for all smooth f,

$$EWf(W) = \sigma^2 Ef'(W^*).$$

From Stein's characterization,

$$EZf(Z) = \sigma^2 Ef'(Z) \quad \text{if and only if} \quad Z \sim \mathcal{N}(0, \sigma^2).$$

Hence:

$$W^* =_d W \quad \text{if and only if} \quad W \sim \mathcal{N}(0, \sigma^2).$$

Zero Bias and Proximity to Normal

$W^* =_d W$ if and only $W \sim \mathcal{N}(0, \sigma^2)$, that is, the mean zero normal is the unique fixed point of the zero bias transformation.

$\therefore$ If W^* is close to W, W is close to being a fixed point, and therefore close to normal.

Size Bias Transformation

For $X \in \{0, 1, 2, \ldots\}$ with $EX = \mu < \infty$, consider the size biased distribution

$$P(X^s = k) = \frac{kP(X = k)}{\mu}.$$

Appears in sampling, generates the waiting time paradox.

The distribution is also characterized by

$$EXf(X) = \mu Ef(X^s) \quad \text{all } f,$$

and can be applied to any $X \geq 0$ with finite mean μ.

Zero and Size Bias

For $W \geq 0$ with $\mu = EW$, we say W^s has the W-size bias distribution if for all f,

$$EWf(W) = \mu Ef(W^s).$$

Zero biasing is the same, with variance replacing mean, and f' replacing f:

$$EWf(W) = \sigma^2 Ef'(W^*).$$

Size Bias Coupling

If $X_1, \ldots, X_n$ are non-negative independent variables with finite means $\mu_1, \ldots, \mu_n$, then with $W = X_1 + \cdots + X_n$,

$$W^s = W - X_I + X_I^s,$$

where

$$P(I = i) = \frac{\mu_i}{\sum_{j=1}^n \mu_j} = \frac{\mu_i}{\mu}.$$

The sum is size biased by replacing one summand, chosen with probability proportional to its expectation, by an independent variable having that summand's size biased distribution.

Coupling

$$\begin{aligned}
\mu E f(W^s) &= \mu E f(W - X_I + X_I^s) \\
&= \mu \sum_{i=1}^{n} E f(W - X_i + X_i^s) \frac{\mu_i}{\mu} \\
&= \sum_{i=1}^{n} \mu_i E f(\sum_{t \neq i} X_t + X_i^s) \\
&= \sum_{i=1}^{n} E X_i f(\sum_{t \neq i} X_t + X_i) \\
&= \sum_{i=1}^{n} E X_i f(W) \\
&= E W f(W)
\end{aligned}$$

Zero and Size Biasing an Independent Sum

To zero (size) bias a sum

$$W = \sum_{i=1}^{k} X_i$$

of mean zero (non-negative) independent variables, pick one proportional to its variance (mean) and replace with biased version.

Zero Bias Proof of CLT

If W is the sum of comparable, independent, mean zero variables then W^* differs from W by only one summand. Hence W^* is close to W, so W is nearly a fixed point of the zero bias transformation, and hence close to normal.

Wasserstein distance d

With

$$\mathcal{L} = \{g : \mathbb{R} \to \mathbb{R} : |g(y) - g(x)| \leq |y - x|\}$$

define

$$d(Y, X) = \sup_{g \in \mathcal{L}} |E[g(Y) - g(X)]|.$$

Dual form, minimal L_1 distance, achieved for $\mathbf{R}$ valued variables

$$d(Y, X) = \inf E|Y - X|,$$

where infimum is over all pairs with given marginals.

Zero Bias and distance d

Lemma 1 *Let W be a mean zero, finite variance random variable, and let W^* have the W-zero bias distribution. Then with d the Wasserstein distance, and Z a normal variable with the same variance as W,*

$$d(W,Z) \leq 2d(W,W^*).$$

Take $\sigma^2 = 1$. For $||h'|| \leq 1$, $||f''|| \leq 2$,

$$\begin{aligned} |Eh(W) - Nh| &= |E[f'(W) - Wf(W)]| \\ &= |[Ef'(W) - Ef'(W^*)]| \\ &\leq ||f''||E|W - W^*| \\ &\leq 2d(W,W^*). \end{aligned}$$

Contraction Mapping in d

Lemma 2 *For* $\boldsymbol{\alpha} \in \mathbb{R}^k$ *with* $\lambda = ||\boldsymbol{\alpha}|| \neq 0$, *let*

$$Y = \sum_{i=1}^{k} \frac{\alpha_i}{\lambda} W_i,$$

where W_i *are mean zero, variance one, independent random variables distributed as* W. *Then*

$$d(Y, Y^*) \leq \varphi \, d(W, W^*),$$

and $\varphi = \sum_i |\alpha_i|^3 / (\sum_i \alpha_i^2)^{3/2} < 1$ *if and only if* $\boldsymbol{\alpha}$ *is not a scalar multiple of a standard basis vector.*

Contraction by Coupling

With $\quad P(I=i)=\frac{\alpha_i^2}{\lambda^2}, \quad |Y-Y^*|=\frac{|\alpha_I|}{\lambda}|W_I-W_I^*|.$

Since $W_i =_d W$, we may take $(W_i, W_i^*) =_d (W, W^*)$

$$E|Y-Y^*|=\sum_{i=1}^{k}\frac{|\alpha_i|^3}{\lambda^3}E|W_i-W_i^*|=\varphi E|W-W^*|.$$

Choosing the pair W, W^* to achieve the infimum, we obtain

$$d(Y,Y^*)\le E|Y-Y^*|=\varphi E|W-W^*|=\varphi\, d(W,W^*).$$

The Classical CLT and d

Take W_i iid mean zero variance σ^2 and

$$Y = n^{-1/2} \sum_{i=1}^{n} W_i.$$

Setting $\alpha_i = n^{-1/2}$ gives $\varphi = n^{-1/2}$, and

$$d(Y, Z) \le 2d(Y, Y^*) \le 2n^{-1/2} d(W, W^*) \to 0$$

as $n \to \infty$, proof of the CLT with a bound in d and constant depending on $E|W^* - W| = ||W^* - W||_1$.

Linear Iteration

Normalizing $X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n$, with $\lambda_n = ||\boldsymbol{\alpha}_n||$ and $\sigma_n^2 = \text{Var}(X_n)$ we have

$$W_{n+1} = \sum_{i=1}^{k} \frac{\alpha_{n,i}}{\lambda_n} W_{n,i} \quad \text{with} \quad W_n = \frac{X_n - c_n}{\sigma_n}.$$

Iterated contraction gives

$$d(W_n, Z) \le 2d(W_n, W_n^*) \le 2 \left(\prod_{i=0}^{n-1} \varphi_i \right) d(W_0, W_0^*).$$

Non-linear Iteration

Let $X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n + R_n$, where $\mathbf{X}_n$ is a vector of iid variables distributed as X_n, $EX_n = c_n$, $\mathrm{Var}(X_n) = \sigma_n^2$, and $\lambda_n = ||\boldsymbol{\alpha}_n|| \neq 0$. Set

$$Y_n = \sum_{i=1}^{k} \frac{\alpha_{n,i}}{\lambda_n} W_{n,i} \quad \text{where} \quad W_n = \frac{X_n - c_n}{\sigma_n}$$

and, measuring the discrepancy from linearity,

$$\beta_n = E|W_{n+1} - Y_n| + \frac{1}{2} E|W_{n+1}^3 - Y_n^3|.$$

Theorem 1 *For* $X_{n+1} = \boldsymbol{\alpha}_n \cdot \mathbf{X}_n + R_n$*, if there exist* $(\beta, \varphi) \in (0,1)^2$ *such that*

$$\limsup_{n\to\infty} \frac{\beta_n}{\beta^n} < \infty \quad \textit{and} \quad \limsup_{n\to\infty} \varphi_n = \varphi,$$

then with $\gamma = \beta$ *when* $\varphi < \beta$*, and for any* $\gamma \in (\varphi, 1)$ *when* $\beta \leq \varphi$*, there exists* C *such that*

$$d(W_n, Z) \leq C\gamma^n.$$

Now apply Theorem 1 to sequences generated using averaging functions F.

Theorem 2 *Let* X_0 *be a non constant random variable with* $P(X_0 \in [a,b]) = 1$ *and* $X_{n+1} = F(\mathbf{X}_n)$ *with* $F : [a,b]^k \to [a,b]$*, twice continuously differentiable. Suppose* F *is averaging and that* $X_n \to_p c$*, with* $\boldsymbol{\alpha} = F'(\mathbf{c})$ *not a scalar multiple of a standard basis vector. Then with* Z *a standard normal variable, for all* $\gamma \in (\varphi, 1)$ *there exists* C *such that*

$$d(W_n, Z) \le C\gamma^n \quad \textit{where} \quad \varphi = \frac{\sum_{i=1}^k |\alpha_i|^3}{(\sum_{i=1}^k |\alpha_i|^2)^{3/2}},$$

is a positive number strictly less than 1. The value φ *achieves a minimum of* $1/\sqrt{k}$ *if and only if the components of* $\boldsymbol{\alpha}$ *are equal.*

Averaging by Composition

Under simple non-triviality conditions, if $F_0, F_1, \ldots, F_k$ are scaled, strictly averaging and F_0 is (positively) homogeneous, then

$$F_{\mathbf{1}}(\mathbf{x}) = F_0(F_1(\mathbf{x}_1), \ldots, F_k(\mathbf{x}_k))$$

is a scaled strictly averaging function.

Hence, the diamond lattice conductivity function is again scaled strictly averaging when replacing the L_1 and L_{-1} in the parallel and series combination rules

$$L_1(x_1, x_2) = x_1 + x_2 \quad \text{and} \quad L_{-1}(x_1, x_2) = (x_1^{-1} + x_2^{-1})^{-1}$$

by, say L_2 and L_{-2}, respectively.

Fast Rates for the Diamond Lattice

Define the 'side equally weighted network' to be the one with $\mathbf{w} = (w, w, 2-w, 2-w)^\mathsf{T}$ for $w \in (1,2)$; such weights are positive and satisfy $F(\mathbf{w}) = 1$.

For $w = 1$ all weights are equal, and we have $\boldsymbol{\alpha} = 4^{-1}\mathbf{1}_4$, and hence φ achieves its minimum value $1/2 = 1/\sqrt{k}$ corresponding to the rate $N^{-1/2+\epsilon}$.

For $1 \le w < 2$ we have $1/2 \le \varphi < 1/\sqrt{2}$, the case $w \uparrow 2$ corresponding to the least favorable rate for the side equally weighted network of $N^{-1/4+\epsilon}$.

Slow Rates for the Diamond Lattice

With only the restriction that the weights are positive and satisfy $F(\mathbf{w}) = 1$ consider for $t > 0$,

$$\mathbf{w} = (1 + 1/t, s, t, 1/t)^{\mathsf{T}} \quad \text{where}$$

$$s = [(1 - (1/t + t)^{-1})^{-1} - (1 + 1/t)^{-1}]^{-1}.$$

When $t = 1$ we have $s = 1$ and $\varphi = 11\sqrt{2}/27$.

As $t \to \infty$, $\boldsymbol{\alpha}$ tends to the vector $(1, 0, 0, 0)$, so $\varphi \to 1$.

Since $11\sqrt{2}/27 < 1/\sqrt{2}$, γ takes on all values in the range $(1/2, 1)$, corresponding to $N^{-\theta}$ for any $\theta \in (0, 1/2)$.

Some Further Directions

1. Dependent Variables
2. Kolmogorov Distance
3. Random Networks