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, ... 

\[ X_2 = F(X_1), \quad X_1 = F(X_0) \ldots \]
Hierarchical Models

\[ X_{n+1} = F(X_n) \quad where \quad X_n = (X_{n,1}, \ldots, X_{n,k})^T \]

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)

\[ X_n \rightarrow_p c, \]

and by Woo and Wehr which imply

\[ W_n \rightarrow_d \mathcal{N}(0, 1), \quad for \quad W_n = \frac{X_n - EX_n}{\sqrt{\text{Var}(X_n)}}. \]
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

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

We say $F$ is (strictly) averaging

1. $\min_i x_i \leq F(x) \leq \max_i x_i$, and strictly when $\min_i x_i < \max_i x_i$.

2. $F(x) \leq F(y)$ whenever $x_i \leq y_i$, and strictly when $x_j < y_j$ for some $j$.

Say $F$ is scaled averaging when $F(x)/F(1_k)$ is averaging, where $1_k = (1, \ldots, 1)$.
Diamond Lattice Conductivity Function

Parallel and series resistor combination rules

\[ L_1(x_1, x_2) = x_1 + x_2, \quad L^{-1}_1(x_1, x_2) = (x_1^{-1} + x_2^{-1})^{-1} \]

gives the weighted \( w_i > 0 \) diamond lattice conductivity function

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

a scaled strictly averaging function.
Approximate Linear Recursion

Write $X_{n+1} = F(X_n)$ as a linear recursion with (‘small’) perturbation $R_n$,

$$X_{n+1} = \alpha_n \cdot X_n + R_n, \quad n \geq 0,$$

where $c_n = E X_n$, $\alpha_n = F'(c_n)$, $c_n = (c_n, \ldots, c_n)^T \in \mathbb{R}^k$,

and $F'$ the gradient of $F$.

Rule out trivial cases such as $F(x_1, x_2) = x_1$; when $\alpha = F'(c)$ at limiting $c$ is not a multiple of a standard basis vector, then $\lambda = \|\alpha\| < 1$ when $F$ is averaging.
Stein’s Method for Normal

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

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

\[ \sigma^2 f'(w) - wf(w) = h(w/\sigma) - Nh. \]

Now evaluate expectation of RHS by expectation of LHS.
Goldstein and Reinert 1997: For $W$ a mean zero variance $\sigma^2$ random variable, there exists $W^*$ such that for all smooth $f$,

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

From Stein’s characterization,

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

Hence:

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

∴ 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 $\mu$. 
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 $f$:

$$EWf(W) = \sigma^2Ef'(W^*).$$
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.
\[ \mu Ef(W^s) = \mu Ef(W - X_I + X_I^s) \]
\[ = \mu \sum_{i=1}^{n} Ef(W - X_i + X_i^s) \frac{\mu_i}{\mu} \]
\[ = \sum_{i=1}^{n} \mu_i Ef(\sum_{t \neq i} X_t + X_i^s) \]
\[ = \sum_{i=1}^{n} EX_i Ef(\sum_{t \neq i} X_t + X_i) \]
\[ = \sum_{i=1}^{n} EX_i Ef(W) \]
\[ = EW f(W) \]
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

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

define

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

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

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

where infimum is over all pairs with given marginals.
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 $\|h'\| \leq 1$, $\|f''\| \leq 2$,

$$|Eh(W) - Nh| = |E[f'(W) - Wf(W)]|$$
$$= \|[Ef'(W) - Ef'(W^*)]\|$$
$$\leq \|f''\| |E[W - W^*]|$$
$$\leq 2d(W, W^*).$$
Lemma 2  For $\alpha \in \mathbb{R}^k$ with $\lambda = \|\alpha\| \neq 0$, 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 $\alpha$ is not a scalar multiple of a standard basis vector.
Contraction by Coupling

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

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

\[
d(Y, Y^*) \leq E|Y - Y^*| = \varphi E|W - W^*| = \varphi d(W, W^*).\]
The Classical CLT and \( d \)

Take \( W_i \) iid mean zero variance \( \sigma^2 \) and

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

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

\[
d(Y, Z) \leq 2d(Y, Y^*) \leq 2n^{-1/2}d(W, W^*) \to 0
\]
as \( n \to \infty \), proof of the CLT with a bound in \( d \) and constant depending on \( E|W^* - W| = \|W^* - W\|_1 \).
Normalizing $X_{n+1} = \alpha_n \cdot X_n$, with $\lambda_n = ||\alpha_n||$ and $
abla^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

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

Let \( X_{n+1} = \alpha_n \cdot X_n + R_n \), where \( X_n \) is a vector of iid variables distributed as \( X_n \), \( EX_n = c_n \), \( \text{Var}(X_n) = \sigma_n^2 \), and \( \lambda_n = ||\alpha_n|| \neq 0 \). 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 $X_{n+1} = \alpha_n \cdot X_n + R_n$, if there exist $(\beta, \varphi) \in (0, 1)^2$ such that

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

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 $X_{n+1} = F(X_n)$ with $F : [a, b]^k \rightarrow [a, b]$, twice continuously differentiable.
Suppose $F$ is averaging and that $X_n \rightarrow_p c$, with $\alpha = F'(c)$ 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

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

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

\[
F_1(x) = F_0(F_1(x_1), \ldots, F_k(x_k))
\]

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

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

For $1 \leq w < 2$ we have $1/2 \leq \varphi < 1/\sqrt{2}$, 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 $F(w) = 1$ consider for $t > 0$,

$$w = (1 + 1/t, s, t, 1/t)^T$$

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

Since $11\sqrt{2}/27 < 1/\sqrt{2}$, $\gamma$ 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