On the contraction method for convolution equations

Erwin Bolthausen, University of Zürich

Paris, December 9, 2009

The (weakly) SAW is described by the propagator on $\mathbb{Z}^d$:

$$C_n(x) \stackrel{\text{def}}{=} (2d)^{-n} \sum_{\substack{\omega:0\to x\\ |\omega|=n}} \prod_{0\le i<j\le n} \left(1-\lambda 1_{\omega_i=\omega_j}\right),\ 0<\lambda\le 1.$$

The lace expansion leads to an equation

$$C_n = C_{n-1} * S + \lambda \sum_{k=1}^{n} \Pi_k * C_{n-k},\ C_0 = \delta_0,$$

$$S(x) = \frac{1}{2d},\ |x| = 1.$$

The (signed) kernels Π_k are complicated, but there is a further diagrammatic splitting

$$\Pi_k \stackrel{\text{def}}{=} \sum_{m\ge 1} \Pi_k^{(m)}.$$

The $\Pi_k^{(m)}$ can be *estimated* in terms of the C's, e.g.

$$\Pi_k^{(2)}(x) \leq \lambda \sum_{j_1+j_2+j_3=k} C_{j_1}(x) C_{j_2}(x) C_{j_3}(x) .$$

By the well known result by **Brydges-Spencer**, **Hara-Slade**: C_n/c_n satisfies a CLT for $d \geq 5$, where $c_n \stackrel{\text{def}}{=} \sum_x C_n(x)$.

Key idea in the manuscript with Christine Ritzmann: Π_k have the same leading order decay as the C's. So put

$$B_k \stackrel{\text{def}}{=} \Pi_k / c_k$$

$$C_n = C_{n-1} * S + \lambda \sum_{k=1}^{n} c_k B_k * C_{n-k}.$$

Consider that as an equation for $\mathbf{C} = \{C_n\}$, $C_0 = \delta_0$, **with input** $\{B_k\}$, and **not** the original equation **with input** $\{\Pi_k\}$.

Main work: solve **this** problem properly. This is independent of the SAW.

The result we have (with Ch. Ritzmann)

Theorem Assume $d \geq 5$, and $\lambda > 0$ small enough. If x and n have the same parity, then with the proper chosen variance $\delta = \delta(\lambda)$, and some $\nu > 0$

$$\left| \frac{C_n(x)}{c_n} - 2\varphi_{n\delta}(x) \right| \leq \text{const} \times \left(n^{-(d+1)/2} \mathrm{e}^{-\nu |x|^2/n} + n^{-d/2} \gamma_n(x) \right),$$

$$\varphi_t(x) = (2\pi t)^{-d/2} \exp\left[-\frac{|x|^2}{2t} \right].$$

where γ_n satisfies

$$\lim_{K\to\infty} \sup_n \sup_{|x|\geq K} \gamma_n(x) = 0,$$

$$\sum_x \gamma_n(x) \leq \text{const} \times n^2.$$

γ_n takes care of the fact that there cannot be a local CLT. This is as close as possible to a local CLT.

First task: Extract the exponential decay. With $b_k \stackrel{\text{def}}{=} \sum_x B_k(x)$.

$$c_n = c_{n-1} + \lambda \sum_{k=1}^{n} c_k b_k c_{n-k},\; c_0 = 1.$$

Although, it is not strictly a recursion, it is if λ is small enough, and $b_n \to 0$.

Ansatz: $c_n = \rho^n a_n,\; a_\infty \stackrel{\text{def}}{=} \lim_{n\to\infty} a_n$ exists, and is $\neq 0$.

$$\begin{aligned} a_n\rho^n &= a_{n-1}\rho^{n-1} + \lambda\rho^n \sum_{k=1}^{n} a_k b_k a_{n-k} \\ a_n &= a_{n-1}\rho^{-1} + \lambda \sum_{k=1}^{n} a_k b_k a_{n-k} \\ a_\infty &= a_\infty\rho^{-1} + \lambda \sum_{k=1}^{\infty} a_k b_k a_\infty, \\ \rho^{-1} &= 1 - \lambda \sum_{k=1}^{\infty} a_k b_k. \end{aligned}$$

Plugging that into the original equation:

$$\begin{aligned} a_n &= a_{n-1}\left(1-\lambda\sum\nolimits_{k=1}^{\infty} a_k b_k\right)+\lambda\sum_{k=1}^{n} a_k b_k a_{n-k} \\ &= a_{n-1}+\lambda\sum_{k=1}^{n} a_k b_k\left(a_{n-k}-a_{n-1}\right)-\lambda a_{n-1}\sum_{k=n+1}^{\infty} a_k b_k. \end{aligned}$$

If the ansatz is OK, then $\mathbf{a}=\{a_n\}$ solves this equation. $\{a_n\}$ is no longer a recursively defined. Regard it as a fixed point for an operation on sequences $\mathbf{x}=\{x_n\}$, $x_0=1$:

$$\phi(\mathbf{x})_n-\phi(\mathbf{x})_{n-1}=\lambda\left[\sum_{k=1}^{n} x_k b_k\left(x_{n-k}-x_{n-1}\right)-x_{n-1}\sum_{k=n+1}^{\infty} x_k b_k\right].$$

In an appropriate sequence space with norm

$$\|\mathbf{x}\| \stackrel{\text{def}}{=} \sum_{k\geq 1}\left|x_k-x_{k-1}\right|,$$

this is contractive provided

$$\sum_k k\,|b_k| < \infty,$$

and λ is small enough. (This condition reflects $d \geq 5$ for SAW).
This leads to a fixed point

$$\mathbf{a} = \lim_{n\to\infty} \phi^n(\mathbf{1}),\ \mathbf{1} = (1,1,1,\ldots).$$

It is *not* necessary to prove that the operation ϕ has a unique fixed point, as *any* fixed point defines, by going backwards, a sequence $\{c_n\}$ which satisfies the original equation, which has a unique solution. The conclusion is that if $\sum k\,|b_k| < \infty$, and λ is small then there exists ρ such that

$$\lim_{n\to\infty} c_n \rho^{-n} \in (0,\infty).$$

Side remark: A puzzling **open question** is the case of borderline sequences $\{b_k\}$, e.g. $b_k = k^{-2}$, or $b_k = k^{-2}L(k)$, L slowly varying with $\sum_k k|b_k| = \infty$.
Under which "natural"condition does one have that there is a slowly varying function $L'(n)$ such that

$$\lim_{n\to\infty} L'(n)\,\rho^{-n}c_n \in (0,\infty)?$$

ρ is well defined if only $\sum_k |b_k| < \infty$.
In general, this cannot be true(even for small λ).

Next step: Repeat that line of argument for the sequence of distributions $\{C_n\}$. Put

$$A_n \overset{\text{def}}{=} \rho^{-n} C_n$$

$$A_n = \rho^{-1} A_{n-1} * S + \lambda \sum_{k=1}^{n} a_k B_k * A_{n-k}.$$

The first idea we had was to define an operation Ψ on sequences $\boldsymbol{\xi} = \{\xi_n\}$ by

$$\Psi(\boldsymbol{\xi})_n = \rho^{-1} \Psi(\boldsymbol{\xi})_{n-1} * S + \lambda \sum_{k=1}^{n} a_k B_k * \xi_{n-k}.$$

A fixed point has to be $\boldsymbol{\xi} = \mathbf{A}$. Prove a CLT in the following way:

- Take a norm on sequences such that if $\left\|\boldsymbol{\xi} - \boldsymbol{\xi}'\right\| < \infty$, then ξ_n, ξ_n' have to be asymptotically close, e.g. $\sup_n n^\delta \left\|\xi_n - \xi_n'\right\|_{\text{var}}$.
- Take an appropriate sequence $\mathbf{G}$ of which you already know that it is asymptotically normal (with the proper variance), and prove that $\|\mathbf{G} - \Psi(\mathbf{G})\| < \infty$.
- Prove that $\|\Psi(\mathbf{G}) - \Psi(\mathbf{G}')\| \leq (1 - \varepsilon) \|\mathbf{G} - \mathbf{G}'\|$,

Then you are finished.

This however did not quite work. One first has to adjust the variance.
It is easy to find the correct variance by making the ansatz

$$\sum A_n(x)|x|^2 \approx a_n \delta n,$$

(assuming the B_k invariant under lattice isometries), entering the ansatz into the equation, and solve for δ. We then replaced S by a distribution E which has the correct variance, write the above equation as

$$\xi_n = \xi_{n-1} * E + \xi_{n-1} * \left(\rho^{-1}S - E\right) + \lambda \sum_{k=1}^{n} a_k B_k * \xi_{n-k},$$

and define the operation Ψ by

$$\Psi(\boldsymbol{\xi})_n = \Psi(\boldsymbol{\xi})_{n-1} * E + \xi_{n-1} * \left(\rho^{-1}S - E\right) + \lambda \sum_{k=1}^{n} a_k B_k * \xi_{n-k}.$$

This worked, but became technically quite heavy. One chooses an appropriate norm $\|\cdot\|$ on sequences of distributions, and proves the above scheme with $G_n = a_n E^{*n}$. This leads to a CLT (local or nearly local depending on the norm) under appropriate decay properties of the B's.

For the application to SAW, one needs a *simple* bootstrapping argument by estimating the Π's in terms of the C's, leading to the "appropriate" decay properties of the B's in terms of CLT properties of C_n/c_n. If λ is small enough the circular argument "contracts".

We took a somewhat complicated norm

$$\|\boldsymbol{\xi}\| = \sup_{n,x} \frac{|\xi_n(x)|}{n^{-(d+1)/2} \mathrm{e}^{-\nu|x|^2/n} + n^{-d/2} \gamma_n(x)}$$

which was chosen to make the estimate of the Π's in terms of the C's easy.

The key technical difficulty comes from the use of quite heavy (and not totally standard) Edgeworth type estimates we needed on E^{*n}.

There was another awkward small point: For periodic nearest neighbor S, there is not possibility to contract the variance by an E. (This was irrelevant for SAW).

The improvement we now have is based on a very simple observation:
One can of course explicitly solve the equation for $\Psi(\boldsymbol{\xi})$ by iteration

$$\Psi(\boldsymbol{\xi})_n = \xi_n - \sum_{l=1}^{n} E^{*(n-l)} * \left[\xi_l - \rho^{-1} S * \xi_{l-1} - \lambda \sum_{m=1}^{l} a_m B_m * \xi_{l-m}\right], \; n \geq 1,$$
$$\Psi(\boldsymbol{\xi})_0 = \xi_0.$$

Observe now: Replace E^{*k} by *any* sequence $\mathbf{F} = \{F_k\}_{k \geq 0}$ of probability distributions satisfying $F_0 = \delta_0$. Call the operation $\Psi_{\mathbf{F}}$. This has $\{A_n\}$ as its only fixed point with $A_0 = \delta_0$. In fact,

$$\Psi_{\mathbf{F}}(\boldsymbol{\xi}) = \boldsymbol{\xi} \Longrightarrow \sum_{l=1}^{n} F_{n-l} * \left[\xi_l - \rho^{-1} S * \xi_{l-1} - \lambda \sum_{m=1}^{l} a_m B_m * \xi_{l-m}\right] = 0, \; \forall n$$

and by induction

$$\xi_n = \rho^{-1} S * \xi_{n-1} + \lambda \sum_{m=1}^{n} a_m B_m * \xi_{l-m}, \; \forall n \Longrightarrow \xi_n = A_n, \; \forall n.$$

The whole trick is make a clever choice of $\{F_n\}$.

Toy problem: Take $d = 1$, state space $\mathbb{R}$ instead of $\mathbb{Z}$. S a distribution with a density, $\int xS(dx) = 0$, $\int x^2 S(dx) = 1$, $\int |x|^3 S(dx) < \infty$. ϕ_n the normal distribution with variance n. Try to prove with the method that

$$\|S^{*n} - \phi_n\|_{\mathrm{var}} = O\left(n^{-1/2}\right).$$

(The total variation norm is just for warmup: It is powerless for the SAW).
Operator $\Psi : \mathbf{F}$ with $F_0 = \delta_0$.

$$\begin{aligned}\Psi_{\mathbf{F}}(\boldsymbol{\xi})_n &= \xi_n - \sum_{l=1}^{n} F_{n-l} * [\xi_l - \xi_{l-1} * S] \\ &= F_n * \xi_0 - \sum_{l=0}^{n-1} (F_{n-l} - S * F_{n-l-1}) * \xi_l.\end{aligned}$$

Then every fixed point is S^{*n}.
First trial: $F_n \stackrel{\text{def}}{=} \phi_n$, and take the norm $\|\boldsymbol{\xi}\| \stackrel{\text{def}}{=} \sup_n \sqrt{n}\, \|\xi_n\|_{\mathrm{var}}$. Then

$$\Psi_{\boldsymbol{\phi}}(\boldsymbol{\phi})_n = \phi_n - n[\phi_n - \phi_{n-1} * S],$$

so

$$\|\boldsymbol{\phi}-\Psi_{\boldsymbol{\phi}}(\boldsymbol{\phi})\| < \infty.$$

The problem arises when trying to prove contraction: If $\xi_0 = 0,\ \|\boldsymbol{\xi}\| < \infty$

$$\|\Psi_{\boldsymbol{\phi}}(\boldsymbol{\xi})_n\|_{\mathrm{var}} \leq \|\boldsymbol{\xi}\| \sum_{l=1}^{n-1} \|\phi_{n-l} - S * \phi_{n-l-1}\|_{\mathrm{var}}\, l^{-1/2}.$$

The factor is a $\mathrm{const}\times n^{-1/2}$, but in general, one cannot have the constant being < 1. The way out is to slightly modify the sequence $\mathbf{F}$:

$$F_n = \left[\left(1 - \frac{n \wedge N}{N}\right) S^{*n} + \frac{n \wedge N}{N}\phi_n\right].$$

Then if N is large enough, *and* $\|S^{*n} - \phi_n\|_{\mathrm{var}} < 1$ for large enough n, then one gets the desired contraction and therefore

$$\|S^{*n} - \phi_n\|_{\mathrm{var}} \leq \mathrm{const}\times n^{-1/2}.$$

This generalizes easily to our type of equations, say in $\mathbb{R}$ or $\mathbb{R}^d$

$$A_n = \rho^{-1} A_{n-1} * S + \lambda \sum_{k=1}^{n} a_k B_k * A_{n-k},$$

provided one has suitable decay properties of $\{B_k\}$. The crucial thing is to simply expand

$$(\varphi_n * B)(x) = \int \varphi_n (x-y) B(dy) = b\phi_n (x) - \sum\nolimits_i \partial_i \varphi_n (x) \left[\int y_i B(dy) \right]$$
$$+ \frac{1}{2} \sum\nolimits_{i,j} \partial_{ij}^2 \varphi_n (x) \left[\int y_i y_j B(dy) \right] + \mathrm{error},$$

φ_n the density of ϕ_n, and then one relates the derivatives of φ to the time derivative via the heat equation. This all works also in the the case of asymmetric S, B (on $\mathbb{R}$ or $\mathbb{R}^d$).

There are some technical difficulties on $\mathbb{Z}$ or $\mathbb{Z}^d$. The natural choice for the "guiding" sequence $\{F_n\}$ would be a discretization of the normal distribution, like

$$\hat{\varphi}_n(x) = \int_{x-1/2}^{x+1/2} \varphi_n(x+y)\, dy,\ x \in \mathbb{Z},$$

which leads to the problem that $\hat{\varphi}_n * \hat{\varphi}_m$ is not exactly $\hat{\varphi}_{n+m}$. What we need in the end is

$$\|\hat{\varphi}_{n-1} * \hat{\varphi}_m - \hat{\varphi}_n * \hat{\varphi}_{m-1}\|_{\text{var}} \leq \text{const} \times \max(n,m)^{-3/2},$$

then of course also in more sophisticated norms, which (at least for $\|\cdot\|_{\text{var}}$) is fine in the symmetric case, but turned out to becomes again messy for the case for the asymmetric case.
The cheap way on the lattice is to take simply the transition kernel of a suitable continuous time random walks on $\mathbb{Z}^d$. For instance on $\mathbb{Z}$, we take a nearest neighbor random walk $\{X_t\}$ with mean μt and variance t whose transition probabilities $p(t,x)$ satisfy

$$\frac{\partial p(t,x)}{\partial t} = \frac{1}{2}\left[p(t,x+1) + p(t,x-1) - 2p(t,x)\right] - \frac{\mu}{2}\left[p(t,x+1) - p(t,x-1)\right].$$

Then we take $F_n(x) = p(\delta n, \mu n)$ with appropriately adapted δ, μ. This has two advantages:

- $F_n * F_m = F_{n+m}$
- Discrete space derivatives can be translated into time derivatives.

We have checked that for $d = 1$, total variation norm, but not yet for the norms which are good enough for the SAW, but I cannot see a serious obstacle to handle that.

The problem we are presently working on are SAW with not necessary symmetric one-step distributions S where also $\sum_x xS(x)$ may be $\neq 0$. Then, depending on λ, the SAW may still have zero drift, or if $\sum_x xS(x) = 0$ the SAW drift may be non-zero. I expect that there is a d-dimensional manifold in the parameter space (S, λ) for which the SAW has zero drift.

Summarizing:

- The method is all based on direct x-space estimates. No Fourier- or Laplace-transforms with sometimes cumbersome inversion problems are used.
- There is a lot of flexibility in choosing the “guiding” sequence $\{F_n\}$ and the norm on sequences of distributions which can be adapted to specific problems.