Some applications of Stein’s method

A.D. Barbour, Universität Zürich

Symposium in memory of Charles Stein
Singapore, 2019
A BOUND FOR THE ERROR IN THE NORMAL APPROXIMATION TO THE DISTRIBUTION OF A SUM OF DEPENDENT RANDOM VARIABLES

CHARLES STEIN
Stanford University
**Corollary 3.1.** If $X_1, X_2, \cdots$ is a stationary $m$-dependent sequence of random variables satisfying (3.1), (3.2), and (3.3), there exists a constant $A$ (depending on the distribution of the sequence $X_1, X_2, \cdots$ but not on $n$) such that for all $n$ and all real $a$

$$\left| P \left\{ \frac{\sum_{i=1}^{n} X_i}{\sqrt{nC}} \leq a \right\} - \Phi(a) \right| \leq An^{-1/2}.$$
Example. Let $Y_1, Y_2, \ldots$ be i.i.d. continuous r.v.'s.

Set $Z_{ij} := I \left[ Y_i = \max \{ Y_{i-j}, Y_{i-j+1}, \ldots, Y_{i-j+k-1} \} \right], \quad 0 \leq j \leq k-1$.

$Z_i := \max_{0 \leq j \leq k-1} Z_{ij} = I \left[ "Y_i \text{ is a } k\text{-maximal element}" \right].$

Then $X_i := Z_i - EZ_i$ has mean zero, finite moments, and $(X_i, i \geq k)$ form a stationary, $2k-1$ dependent sequence.

Write $S_n := \sum_{i=k}^{n+k-1} X_i$: What about its distribution?
It is easy to check that $n^{-1} \text{var} S_n \to \sigma^2 > 0$ as $n \to \infty$. Hence Corollary 3.1 of Stein's paper immediately gives a normal approximation, with error of order $n^{-1/2}$.

This problem was investigated using recurrence relations in Austin, Fegan, Lehrer and Penney, Ann. Math. Stat. 1957.
Chen, Louis H. Y.:
Poisson approximation for dependent trials.

The author applies an ingenious technique of Ch. Stein [Proc. 6th
Zbl. 278,60026)] to estimating the error involved in approximating the distribu-
tion of a sum of weakly dependent Bernoulli random variables by a Poisson dis-
tribution. The method, described below in the Poisson context, is applicable
to other distributions: Stein considered approximation to the normal law. The
first step is to find a difference operator for which the Poisson distribution
with given parameter is an integrating factor: an obvious choice is indicated
by the identity (for any $x(\cdot)$)

$$
\sum_{j=0}^{n} e^{-\lambda} \frac{\lambda^j}{j!} \{ \lambda x(j+1) - j x(j) \} = x(n+1) e^{-\lambda} \frac{\lambda^{n+1}}{n+1!}
$$

Thus if $W$ is distributed as Poisson with parameter $\lambda$,

$$
E[\lambda x(W+1) - W x(W)] = 0
$$

for any bounded $x(\cdot)$. Since (1) also gives a bounded solution $x(\cdot)$ if, for any
$r$, we choose $\lambda x(j+1) - j x(j) = I[j < r] - \sum_{k=0}^{r} e^{-\lambda} \frac{\lambda^k}{k!}$, it follows that (2) char-
acterises the Poisson distribution $P_\lambda$, and that, if $W$ has some other distribu-


Equilibria of Markov population processes.

Generator

$$A h(x) = x(1+\Theta x/N)(h(x-1)-h(x)) + \lambda x \sum_{r \geq 1} \phi_r (h(x+r)-h(x))$$

$$+ \varepsilon \sum_{x_0} (h(1)-h(0))$$

Drift: $$-x(1+\Theta x/N) + \lambda x \mu$$, where $$\mu = \sum_{r \geq 1} \phi_r$$
So drift is zero for $\bar{X} = N \bar{x}$, where $\bar{x} = (\mu - 1)/\sigma$.

- Now write $Y = (X - \bar{X})/\sqrt{N}$, or $X = \bar{X} + Y\sqrt{N}$.

Generator for $Y$ acting on $\hat{h}(Y) = h(\bar{X} + Y\sqrt{N})$:

$$(\bar{X} + Y\sqrt{N}) (1 + \frac{\Theta}{N}(\bar{X} + Y\sqrt{N})) (\hat{h}(Y - \frac{1}{\sqrt{N}}) - \hat{h}(Y))$$

$$+ \lambda (\bar{X} + Y\sqrt{N}) \sum_{n=1}^{\infty} \frac{\Theta}{n} (\hat{h}(Y + \frac{n}{\sqrt{N}}) - \hat{h}(Y))$$

- Expand in powers of $\sqrt{N}$, using

$$\hat{h}(Y + \frac{n}{\sqrt{N}}) - \hat{h}(Y) = \frac{n}{\sqrt{N}} \hat{h}'(Y) + \frac{1}{2} \frac{n^2}{N} \hat{h}''(Y) + \ldots.$$
The $N^{1/2}$ terms cancel.

Constant terms:

$$Y\sqrt{N} (1 + \Theta^2 + \Theta^2) (-\frac{1}{N} \hat{h}'(Y)) + \frac{1}{2} \frac{\Sigma}{N} (1 + \Theta^2) \hat{h}''(Y)$$

$$+ \frac{\Sigma}{\sqrt{N}} \sum_{i \geq 1} \frac{\hat{h}'(Y)}{\sqrt{N}} + \frac{1}{2} \frac{\Sigma}{N} \sum_{i \geq 1} r_i^2 \hat{h}''(Y)$$

$$= -\Theta \bar{Y} \hat{h}'(Y) + \frac{1}{2} 2 \bar{Y} (\mu + m_2) \hat{h}''(Y)$$

remainder of order $N^{-1/2}$.

and $E(\hat{h}(Y)) = 0$ in equilibrium.
Here, the operator $A$ is the $Q$-matrix of a Markov process, and its equilibrium distribution $\pi$ satisfies

$$\pi^T Q = 0.$$ 

If $\varepsilon = 0$, the equilibrium distribution is concentrated on $0$. Darroch and Seneta: the limiting conditional distribution $\pi$ describes the 'long-term' distribution on $N$, before reaching $0$: a 'Quasi-Stationary' distribution.

$Q^N$ has largest eigenvalue $-\lambda < 0$, and

$$\pi^T Q^N = -\lambda \pi^T$$
For $\varepsilon > 0$, we have $\Pi_{\varepsilon} (A_2 h) = 0$ for all $h$. We can solve the Stein equation

$$A_2 h = f - \Pi_{\varepsilon} (f)$$

to give $h_\varepsilon$, for any bounded $f$, and

$$\|h_\varepsilon\|_{L^2} \leq C \|f\|_{L^2} \left( \frac{1}{\varepsilon} + \log N \right).$$

So

$$\Pi_{\varepsilon} (f) - \Pi_{\varepsilon} (f) = \Pi_{\varepsilon} (A_2 h_\varepsilon)$$

and

$$\Pi_{\varepsilon} (A_2 h) = 0 \text{ for all } h.$$  

Hence

$$\Pi_{\varepsilon} (f) - \Pi_{\varepsilon} (f) = \Pi_{\varepsilon} ( (A_2 - A_2') h_\varepsilon).$$
Analytic arithmetic of algebraic function fields: a probabilistic approach.


"Analytic and algebraic topology of locally Euclidean metrizations of infinitely differentiable Riemannian manifolds." Lehner, 1953.
An additive arithmetic semigroup $G$ is a free commutative semigroup with identity, having countable free generating set $\mathcal{P}$ of primes, and a degree mapping $\mathcal{D} : G \to \mathbb{N}$ such that

1. $\mathcal{D}(gh) = \mathcal{D}(g) + \mathcal{D}(h)$ for all $g, h \in G$;

2. for $G_n := \{ g \in G : \mathcal{D}(g) = n \}$, $G(n) := |G_n| < \infty$.

$f : G \to \mathbb{R}$ is additive if $f(gh) = f(g) + f(h)$ when $g, h$ are coprime.

Strongly additive if $f(p^k) = f(p) \cdot p^k$ when $p$ is prime.

Completely additive if $f(p^k) = k f(p)$ when $p$ is prime.
Knopfmacher's Axiom $A^*$: \( G(n) = K e^n \xi 1 + O(e^{-\alpha n}) \), for some \( \alpha > 0, \xi > 1 \).

Zhang allows \( K = k(n) = \sum_{\substack{i \geq 1 \\ \sigma(j) \leq n \\ \sigma(j) \leq \sum_{i=1}^{j} \sigma(i) \leq \sum_{i=1}^{j+1} \sigma(i) \leq n}} 1/j \), with \( \sigma(j) \leq j < j+1 \) and \( k_r > 0 \).

Under these global assumptions, they deduce information about \( m_n := \# \{ p \in G_n : p \text{ prime} \} \), a "prime number theorem." Typically, \( nm_n e^{-n} \rightarrow \infty \) as \( n \rightarrow \infty \). Monic polynomials: \( G(n) = e^n, nm_n e^{-n} \rightarrow 1 \).
Example.

monic polynomials over finite fields.

\[ g = x^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0, \]

with \( a_0, \ldots, a_{n-1} \in \text{GF}(q) \):

\[ \deg(g) = \text{degree of } g = n \]

\( g \) is prime if it is irreducible.
Now let $S_n'$ be uniformly distributed on $G_n$. What is the distribution of $f(S_n')$, where $f$ is additive?

For $g \in \mathfrak{S}_n$, let $c_j(g)$ denote the number of primes of degree $j$ in its representation. Then $f(g)$ is a sum of contributions from the primes of degree $j = 1, 2, \ldots, n$, and these depend on the choice of primes, and on the values $f(p^k)$, for $p \in \mathfrak{P}$ and $k \geq 1$. 

Define \( C^{(n)} := (c_1(S^n), \ldots, c_n(S^n)) \).

Then the distribution of \( f(S^n) \) is that of

\[
X_n := \sum_{j=1}^{n} I [C_j^{(n)} \geq 1] U_j(C_j^{(n)}),
\]

where \( \{U_j(l); j, l \geq 1\} \) are independent of each other and of \( C^{(n)} \). \( U_j(l) \) has the distribution of \( f \) applied to a random selection of \( l \) primes of degree \( j \), chosen with replacement.

So the distribution of the vector \( C^{(n)} \) is key.
Assume that $\gamma_m \equiv j^{-\beta} \rightarrow 0 > 0$ as $j \rightarrow \infty$.

For $w \in \mathbb{Z}_+^n$ and $0 \leq a < b$, define

$$T_{ab}(w) := \sum_{j=a+1}^{b} j^{m_j} w_j.$$ 

Fact: $L(C^n) = L(Z_1, \ldots, Z_n \mid T_{0n}(Z) = n)$, where $Z_1, Z_2, \ldots$ are independent, and

$$Z_j \sim \text{NB}(m_j, \phi) \text{ for any choice of } 0 < \phi < 1.$$ 

.$
Choose $p$ to make $P[T_{on}(z) = n]$ not too small.

$$T_{on}(z) = \sum_{j=1}^{n} j m_j \frac{p^j}{(1-p)^j}$$

$$\sim \sum_{j=1}^{n} \Theta \left( q \frac{p^j}{(1-p)^j} \right) \sim n \Theta \text{ as } n \to \infty,$$

if $p = \frac{1}{q}$.

With this choice of $p$, as $j \to \infty$,

$$P[Z_j = 1] = m_j q^{-j} (1-q^{-j})^{m_j} \sim \Theta \frac{q^{m_j}}{j} =: P[Z_j = 1]$$

$$P[Z_j = 0] \sim 1 - P[Z_j = 1].$$
We find two approximations of $L(C^{(n)})$ that are useful:

1. $L(C^{(n)} [1, b_n]) \sim L(Z [1, b_n])$ if $b_n = o(n)$ as $n \to \infty$;

2. $L(C^{(n)} [b_n, n]) \sim L(C^{(n)} [1, b_n])$ if $b_n \to \infty$ as $n \to \infty$.

Here, $L(C^{(n)}):= L(\mathbb{Z}_n^*, \ldots, \mathbb{Z}_n^* | T_{\text{on}}(\mathbb{Z}_n^*) = n)$, and $Z_j^* \sim \text{Po}(\theta/1_j)$ are independent.

Both approximations need to be quantified in terms of (total variation) error.
For (1).

Note that, by independence of $\Xi_1, \Xi_2, \ldots$,

$$L\left(C^n \mid T_{ob}(C^n) = k\right) = L\left(Z \mid T_{ob}(Z) = k\right),$$

for any $k \geq 0$. Hence

$$D_{wb} := d_{TV}(L(C^n | C), L(Z | C))$$

$$= d_{TV}(L(T_{ob}(C^n)), L(T_{ob}(Z))).$$

NB

As before, $E T_{ob}(Z) \sim b \Theta$, as $b \to \infty$, whereas

$$\text{var } T_{ob}(Z) = \sum_{j=1}^{b} j^2 \text{var } (Z_j) \sim \frac{b}{\sum_{j=1}^{b} j^2} \frac{\Theta}{\Theta} e^{-\Theta j} \frac{1}{2} \Theta b^2,$$

so $T_{ob}(Z) \geq 0$ and $E T_{ob}(Z) = SD(T_{ob}(Z))$ as $b \to \infty$. 
Now

\[ \Delta_n = \sum_{k>0} \left| \frac{P[T_{ob} = k \text{ and } T_{on} = n]}{P[T_{on} = n]} - P[T_{ob} = k] \right| \]

\[ = \sum_{k>0} P[T_{ob} = k] \left| \frac{P[T_{bn} = n-k]}{P[T_{on} = n]} - 1 \right| . \]

**Strategy:**

(a) Show that

\[ \left| \frac{P[T_{bn}(Z) = n-k]}{P[T_{bn}(Z^*) = n-k]} - 1 \right| = O\left(\frac{k}{n}\right) . \]

(b) Show that

\[ \left| \frac{P[T_{bn}(Z^*) = n-k]}{P[T_{on}(Z^*) = n]} - 1 \right| = O\left(\frac{k+b}{n}\right) . \]
For (2):

\[ P \left[ C^{(n)} [b+1, n] = x [b+1, n] \right] \]

\[ = \frac{P \left[ Z [b+1, n] = x [b+1, n] \right] \cdot P \left[ T_{ab} (Z) = n - T_{bn} (x) \right]}{P \left[ T_{on} (Z) = n \right]} \]

Now \[ \left| \frac{P \left[ Z [b+1, n] = x [b+1, n] \right]}{P \left[ Z^* [b+1, n] = x [b+1, n] \right]} - 1 \right| = O(b^{-1} n \theta) \],

directly. Hence

\[ \left| \frac{P \left[ C^{(n)} [b+1, n] = x [b+1, n] \right]}{P \left[ C^{(n)*} [b+1, n] = x [b+1, n] \right]} - 1 \right| \]

is small, if we have solved (17, part 6).
Part (a) of (1) is a local estimate. Try Stein’s method for
\[
L(T_{\mathbb{N}}(Z^*)) = L(\sum_{j=1}^{n} \xi_j Z_j) = CP(\Theta \phi(n+1) \mu_{n+1}),
\]
where \( \mu_{n+1} := \frac{1}{j \phi(n+1)} \) and \( \phi(n+1) := \sum_{j=1}^{n} \frac{1}{j} \sim \log n. \)

Stein operator:
\[
\sum_{j=1}^{n} j \cdot \Theta \phi(n+1) \mu_{n+1} g(w+j) - wg(w)
= \Theta \sum_{j=1}^{n} g(w+j) - wg(w), \quad w \in \mathbb{Z}_+.\]
So, for $W^*_n := T_{on}(\mathbb{Z}^*)$, taking $g := \sum_{k \geq 1} k^2$, we have

$$\Theta \mathbb{P}\big[ (k-n)_+ \leq W^*_n \leq k-1 \big] = k \mathbb{P}\big[ W^*_n = k \big];$$

hence point probabilities are implied by interval probabilities. Similarly, if $W_n := T_{on}(\mathbb{Z})$ is such that

$$\mathbf{(*)} \quad \mathbb{E}\left[ \Theta \sum_{j=1}^{n} g(W_n+j) - W_n g(W_n) \right] = \varepsilon_n(g),$$

with $|\varepsilon_n(\mathbb{Z}^*)|$ small, then we can approximate

$$k \mathbb{P}\big[ T_{on}(\mathbb{Z}) = k \big] \mathbf{(*)} \mathbb{P}\big[ (k-n)_+ \leq T_{on}(\mathbb{Z}) \leq k-1 \big].$$
Finally, we can use (\ref{eq:xxx}) to show that
\[
d_{TV}(L(Torn(Z^k)), L(Torn(Z^k)))
\]
is small. We just need to control the solution $g_A$ to
\[
\sum_{j=1}^{n} g(w+j) - wg(w) = A_A^i(w) - P[W^eA].
\]

- general CP bound for $g$ has factor $e^{O(n^{1+})} \sim n^\theta$. 
- $\mu_{ij}^{(n)}$ is decreasing in $j$, but $\mu_1^{(n)} - 2\mu_2^{(n)} = 0$. 
Hence we need to derive bounds particular to the example.

Writing \( h(w) := g(w) - g(w-1) \) gives the generator version of the Stein operator

\[
\Theta \left\{ h(w+n) - h(w) \right\} + w \left\{ h(w-1) - h(w) \right\},
\]

an immigration - death process with unit per capita death rate and immigration in batches of \( n \).

Analogously, \( L_n \) \( \lambda \) \( \text{Ton}(Z^n) \) can be approximated by \( P_0 \) on \( R_+ \), with operator \( \Theta \left\{ h(x+1) - h(x) \right\} - \lambda h'(x), x > 0 \).
Small worlds models

Circle $C$ of circumference $N \sim Po \left( \frac{1}{2} P \right)$ randomly chosen chords as short cuts;

$P$ a tuning parameter.

Distribution of distance between two randomly chosen points

Interval branching process.

Intervals grow at unit rate at each end. If number of intervals is \( M \), branching rate is \( 2\lambda M \). Centres of new intervals are \( Y_0, Y_1, Y_2, \ldots \) i.i.d. uniform on \( C \); birth times \( 0 = T_0 < T_1 < \ldots \) such that \( T_j - T_{j-1} \sim \text{Exp}(2\lambda j) \).

Covered set at time \( t \) is

\[
\mathcal{C}(t) = \bigcup_{j: T_j \leq t} \left[ Y_j - (t - T_j), Y_j + (t - T_j) \right]
\]
$M$ is a Yule process with birth rate $2p$, so

$E M(t) = e^{2pt}$ and $M(t) e^{-2pt} \rightarrow W \sim \text{Exp}(1)$ a.s.

Also 'total length' $s(t) = \int_0^t 2M(u) du$, so

$E s(t) = \beta^t (e^{2pt} - 1)$ and $s(t) e^{-2pt} \rightarrow \beta^t W$ a.s.

If, at time $t$, the intervals have lengths $l_1, ..., l_{M(t)}$, then the expected number of overlaps is

$$\sum_{1 \leq i < j \leq M(t)} (l_i + l_j)/L = \frac{(M(t) - 1)s(t)}{L} \sim \frac{W e^{2pt}}{4p}$$
Now consider two independent processes starting at $y_0$ and $y'_0$, each running for time $t_v := \frac{1}{4p} \log (4p) + \frac{v}{p}$.

$$d(y_0, y'_0) > \frac{1}{2p} \log (4p) + \frac{2v}{p}$$ if they have no intersection. The expected number of intersections is

$$\frac{M'(t_v) s(t_v) + M(t_v) s'(t_v)}{L} \sim \frac{2W^4 e^{4pt_v}}{4p} = 2W^4 e^{4v}.$$
But now the Stein–Chen method easily gives a distributional approximation:

$$P \left[ d(Y_0, Y_1) > \frac{1}{2p} \log(2p) + \frac{2r}{p} \right]$$

$$\sim \mathbb{E} \left\{ e^{-2WW'Ve^{4r}} \right\} .$$

[All the errors in the approximations can be explicitly controlled.]