Introduction

Goal: Illustrate how Stein’s method can be applied to a variety of distributions

General approaches

• Generator method (Torkel Erhardsson’s lectures)
• Coupling equations
• Densities

We shall certainly cover the first two approaches

Main examples to bear in mind

1. Normal \( \mathcal{N}(0,1) \)
   \[ Ef'(X) - EXf(X) = 0 \]

2. Poisson(\( \lambda \))
   \[ \lambda Ef(X + 1) - EXf(X) = 0 \]

General situation

Target distribution \( \mu \)

1. Find characterization: operator \( A \) such that \( X \sim \mu \) if and only if for all smooth functions \( f \), \( E Af(X) = 0 \)

2. For each smooth function \( h \) find solution \( f = f_h \) of the Stein equation
   \[ h(x) - \int hd\mu = Af(x) \]

3. Then for any variable \( W \),
   \[ Eh(W) - \int hd\mu = E Af(W) \]

Usually need to bound \( f, f' \), or \( \Delta f \)

Here: \( h \) smooth test function; for nonsmooth functions: see techniques used by Shao, Chen, Rinott and Rotar, Götze

The generator approach

Barbour 1989, 1990; Götze 1993

Choose \( A \) as generator of a Markov process with stationary distribution \( \mu \)

That is:

Let \( (X_t)_{t \geq 0} \) be a homogeneous Markov process

Put \( T_t f(x) = E(f(X_t)|X(0) = x) \)

Generator \( Af(x) = \lim_{t \to 0} \frac{1}{t} (T_t f(x) - f(x)) \)

Facts (see Ethier and Kurtz (1986), for example)
1. $\mu$ stationary distribution then $X \sim \mu$ if and only if $E Af(X) = 0$ for $f$ for which $Af$ is defined

2. $T_t h - h = A \left( \int_0^t T_u h du \right)$ and formally

$$\int h d\mu - h = A \left( \int_0^\infty T_u h du \right)$$

if the r.h.s. exists

**Examples**

1. $Ah(x) = h''(x) - xh'(x)$ generator of Ornstein-Uhlenbeck process, stationary distribution $N(0, 1)$

2. $Ah(x) = \lambda(h(x + 1) - h(x)) + x(h(x - 1) - h(x))$ or

$$Af(x) = \lambda f(x + 1) - xf(x)$$

Immigration-death process, immigration rate $\lambda$, unit per capita death rate; stationary distribution Poisson($\lambda$) (see Torkel Erhardsson’s lectures)

Advantage: generalisations to multivariate, diffusions, measure space...

Careful: does not always work, see compound Poisson distribution

**Heuristic to find generator**

**Assume**: distribution based on the limit of $\Phi_n(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ i.i.d.; assume $EX = 0$, $EX^2 = 1$

Construct reversible Markov chain (exchangeable pairs):

1. Start with $Z_n(0) = (X_1, \ldots, X_n)$

2. Pick index $I \in \{1, \ldots, n\}$ independently uniformly at random; if $I = i$, replace $X_i$ by independent copy $X_i^*$

3. Put $Z_n(1) = (X_1, \ldots, X_{i-1}, X_i^*, X_{i+1}, \ldots, X_n)$

4. Draw another index uniformly at random, throw out corresponding random variable and replace by independent copy

5. Repeat

Make time continuous: Put $N(t)$ Poisson process, rate 1, and

$$W_n(t) = Z_n(N(t))$$

Then generator $A_n$, with $x = (x_1, \ldots, x_n)$, $f$ smooth,

$$A_n f(\Phi_n(x)) = \frac{1}{n} \sum_{i=1}^n E f(\Phi_n(x_1, \ldots, x_{i-1}, X_i^*, x_{i+1}, \ldots, x_n)) - f(\Phi_n(x))$$

Taylor expansion:

$$A_n f(\Phi_n(x)) \approx \frac{1}{n} \sum_{i=1}^n E(X_i^* - x_i) f'(\Phi_n(x)) \frac{\partial}{\partial x_i} \Phi_n(x)$$

$$+ \frac{1}{2n} \sum_{i=1}^n E(X_i^* - x_i)^2 \left\{ f''(\Phi_n(x)) \left( \frac{\partial}{\partial x_i} \Phi_n(x) \right)^2 + f'(\Phi_n(x)) \frac{\partial^2}{\partial^2 x_i} \Phi_n(x) \right\}$$

$$= -\frac{1}{n} \sum_{i=1}^n x_i f'(\Phi_n(x)) \frac{\partial}{\partial x_i} \Phi_n(x) + \frac{1}{2n} \sum_{i=1}^n (1 + x_i^2) \left\{ f''(\Phi_n(x)) \left( \frac{\partial}{\partial x_i} \Phi_n(x) \right)^2$$

$$+ f'(\Phi_n(x)) \frac{\partial^2}{\partial^2 x_i} \Phi_n(x) \right\}.$$
Example: Put $\Phi_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i$
Then $\frac{\partial}{\partial x_i} \Phi_n(x) = \frac{1}{\sqrt{n}}$ and
$$\frac{\partial^2}{\partial x_i^2} \Phi_n(x) = 0$$
and
$$\mathcal{A}_n f(\Phi_n(x)) \approx -\frac{1}{n^{3/2}} \sum_{i=1}^n x_i f'(\Phi_n(x)) + \frac{1}{2n^2} \sum_{i=1}^n \left(1 + x_i^2\right) f''(\Phi_n(x))$$
$$= -\frac{1}{n} \Phi_n(x) f'(\Phi_n(x)) + \frac{1}{2n^2} f''(\Phi_n(x)) \left(1 + \frac{1}{n} \sum_{i=1}^n x_i^2\right)$$
$$\approx \frac{1}{n} \left(f''(\Phi_n(x)) - \Phi_n(x) f'(\Phi_n(x))\right)$$
by the law of large numbers
If Poisson process with rate $n$ instead of 1: factor $\frac{1}{n}$ vanishes
Suggests
$$\mathcal{A} f(x) = f''(x) - x f'(x)$$

1. Chi-square distributions

Find generator

$X_1, \ldots, X_p$ i.i.d. random vectors, $X_i = (X_{i,1}, \ldots, X_{i,n})$ i.i.d., mean zero, $EX_{i,j}^2 = 1$, finite fourth moment

$$\Phi_n(x) = \sum_{i=1}^p \left(\frac{1}{\sqrt{n}} \sum_{j=1}^n x_{i,j}\right)^2$$
Choose index uniformly from $\{1, \ldots, p\} \times \{1, \ldots, n\}$
We have
$$\frac{\partial}{\partial x_{i,j}} \Phi_n(x) = \frac{2}{n} \sum_{k=1}^n x_{i,k}$$
and
$$\frac{\partial^2}{\partial x_{i,j}^2} \Phi_n(x) = \frac{2}{n}$$
and
$$\mathcal{A}_n f(\Phi_n(x)) \approx -\frac{2}{pn} f'(\Phi_n(x)) \sum_{i=1}^p \sum_{j=1}^n x_{i,j} \frac{1}{n} \sum_{k=1}^n x_{i,k}$$
$$+ \frac{1}{2dn} f''(\Phi_n(x)) \sum_{i=1}^p \sum_{j=1}^n \left(1 + x_{i,j}^2\right) \frac{4}{n^2} \left(\sum_{k=1}^n x_{i,k}\right)^2$$
$$+ \frac{1}{2dn} f'(\Phi_n(x)) \sum_{i=1}^p \sum_{j=1}^n \left(1 + x_{i,j}^2\right) \frac{2}{n}$$
$$\approx -\frac{2}{pn} f'(\Phi_n(x)) \Phi_n(x) + \frac{4}{pn} f''(\Phi_n(x)) \Phi_n(x) + \frac{2}{n} f'(\Phi_n(x))$$
by the law of large numbers
Suggests
$$\mathcal{A} f(x) = \frac{4}{p} x f''(x) + 2 \left(1 - \frac{x}{p}\right) f'(x)$$

3
More convenient: Generator for $\chi^2_p$

$$Af(x) = xf''(x) + \frac{1}{2}(p-x)f'(x)$$

Luk 1994: Stein operator for Gamma($r, \lambda$) is

$$Af(x) = xf''(x) + (r - \lambda x)f'(x)$$

and $\chi^2_p = \text{Gamma}(d/2, 1/2)$

Luk also showed that, for $\chi^2_p$, $A$ is the generator of a Markov process given by the solution of the stochastic differential equation

$$X_t = x + \frac{1}{2} \int_0^t (p - X_s) ds + \int_0^t \sqrt{2X_s} dB_s$$

where $B_s$ is standard Brownian motion.

Luk found the transition semigroup, which can be used to solve the Stein equation

$$(\chi^2_p) h(x) - \chi^2_p h = xf''(x) + \frac{1}{2}(p-x)f'(x)$$

where $\chi^2_p h$ is the expectation of $h$ under the $\chi^2_p$-distribution.

**Lemma 1 (Pickett 2002)**

Suppose $h : \mathbb{R} \to \mathbb{R}$ is absolutely bounded, $|h(x)| \leq ce^{ax}$ for some $c > 0$ and $a \in \mathbb{R}$, and the first $k$ derivatives of $h$ are bounded. Then the equation $(\chi^2_p)$ has a solution $f = f_h$ such that

$$\| f^{(j)} \| \leq \frac{\sqrt{2\pi}}{\sqrt{p}} \| h^{(j-1)} \|$$

with $h^{(0)} = h$.

(Improvement over Luk 1994 in $\frac{1}{\sqrt{p}}$)

**Example: squared sum (R. + Pickett)**

$X_i, i = 1, \ldots, n$, i.i.d. mean zero, variance one, existing $8^{th}$ moment

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

and

$$W = S^2$$

Want

$$2EWf''(W) + E(1-W)f'(W)$$

Put

$$g(s) = sf'(s^2)$$

then

$$g'(s) = f'(s^2) + 2s f''(s^2)$$

and

$$2EWf''(W) + E(1-W)f'(W) = Eg'(S) - Ef'(W) + E(1-W)f'(W)$$

$$= Eg'(S) - ESg(S)$$
Now proceed as in $\mathcal{N}(0, 1)$:

Put

$$S_i = \frac{1}{\sqrt{n}} \sum_{j \neq i} X_j$$

Then

$$ESg(S) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} EX_i g(S)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} EX_i g(S_i) + \frac{1}{n} \sum_{i=1}^{n} EX_i^2 g'(S_i) + R_1$$

where

$$R_1 = \frac{1}{n^{3/2}} \sum_i E X^3_i g''(S_i) + \frac{1}{2n^2} \sum_i E X^4_i g^{(3)} \left( S_i + \theta \frac{X_i}{\sqrt{n}} \right)$$

by Taylor expansion, some $0 < \theta < 1$

From independence

$$ESg(S) = \frac{1}{n} \sum_{i=1}^{n} Eg'(S_i) + R_1$$

$$= Eg'(S) + R_1 + R_2$$

where

$$R_2 = \frac{1}{n^{3/2}} \sum_i E X^3_i g''(S_i) + \frac{1}{2n^2} \sum_i E X^2_i g^{(3)} \left( S_i + \theta \frac{X_i}{\sqrt{n}} \right)$$

$$= \frac{1}{2n^2} \sum_i E X^2_i g^{(3)} \left( S_i + \theta \frac{X_i}{\sqrt{n}} \right)$$

by Taylor expansion, some $0 < \theta < 1$

**Bounds on $R_1, R_2$**

Calculate

$$g''(s) = 6sf''(s^2) + 4s^3 f^{(3)}(s^2)$$

and

$$g^{(3)}(s) = 24s^2 f^{(3)}(s^2) + 6f''(s^2) + 8s^4 f^{(4)}(s^2)$$

so with $\beta_i = EX_i^1$

$$\frac{1}{2n^2} \sum_i E X^2_i \left| g^{(3)}(S_i + \theta \frac{X_i}{\sqrt{n}}) \right|$$

$$\leq \frac{24}{n} \| f^{(3)} \| \left( 1 + \frac{\beta k}{n} \right) + \frac{6}{n} \| f'' \|
+ \frac{8}{n} \| f^{(4)} \| \left( 6 + \frac{\beta k}{n} + 4 \frac{\beta^2}{n} + 6 \frac{\beta^3}{n^2} + \frac{\beta^4}{n^2} \right)
= c(f) \frac{1}{n}$$

Similarly for $\frac{1}{2n^2} \sum_i E X^4_i \left| g^{(3)}(S_i + \theta \frac{X_i}{\sqrt{n}}) \right|$, employ $\beta_k$
For \( \frac{1}{n^{3/2}} \sum_i E X_i^3 g''(S_i) \) have, for some \( c(f) \)
\[
\frac{1}{n^{3/2}} \sum_i E X_i^3 g''(S_i) = \frac{1}{\sqrt{n}} \beta_3 Eg''(S) + c(f) \frac{1}{n}
\]
and
\[
Eg''(S) = 6ESf''(S^2) + 4ESf(3)(S^2)
\]
Note that \( g'' \) is antisymmetric, \( g''(-s) = -g''(s) \), so for \( Z \sim \mathcal{N}(0,1) \) we have
\[
Eg''(Z) = 0
\]
(A) routine now to show that \(|Eg''(S)| \leq c(f)/\sqrt{n} \) for some \( c(f) \).

Combining these bounds show: the bound on the distance to \( \text{Chisquare}(1) \) for smooth test functions is of order \( \frac{1}{n} \).

2. The weak law of large numbers
Using the generator method, we find for \( \delta_0 \), point mass at 0, 
\[
\mathcal{A}f(x) = -xf'(x)
\]
and the corresponding transition semigroup is given by 
\[
T_t h(x) = h(xe^{-t})
\]
Stein equation for point mass at 0 
\[
(\delta_0) \quad h(x) - h(0) = -xf'(x)
\]

Lemma 2 If \( h \in C^2_b(\mathbb{R}) \), then the Stein equation \( (\delta_0) \) has solution \( f = f_h \in C^2_b \) such that 
\[
\|f'\| \leq \|h'\|
\]
\[
\|f''\| \leq \|h''\|
\]
Proof
May assume \( h(0) = 0 \). Generator method gives
\[
f(x) = -\int_0^\infty h(xe^{-t}) \, dt = -\int_0^x \frac{h(t)}{t} \, dt
\]
so for \( x \neq 0 \)
\[
|f'(x)| = \left| \frac{h(x)}{x} \right| \leq \|h'\|
\]
and for \( x = 0 \) we have 
\[
f'(0) = -h'(0)
\]
giving the first assertion. For the second assertion, for \( x \neq 0 \)
\[
|f''(x)| = \left| \frac{h(x)}{x^2} - \frac{h'(x)}{x} \right| \leq \|h''\|
\]
and for \( x = 0 \) we have 
\[
f''(0) = -h''(0).
\]
Example:
$X_1, \ldots, X_n$ mean zero

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

Then, by Taylor, for some $0 < \theta < 1$,

$$E Af(W) = -EWf'(W)$$
$$= -EWf'(0) + EW^2 f''(\theta W)$$

and

$$|E Af(W)| \leq \|f''\| Var(W).$$

If $Var(W_n) \to 0$ as $n \to \infty$ then the weak law of large numbers holds.

Remarks

- For point mass at $\mu$ obtain $Af(x) = (\mu - x)f'(x)$
- Explicit bound, no need for $n \to \infty$

Empirical measures

$E = \mathbb{R}, \mathbb{R}^d, \mathbb{R}_{+}, \ldots$ (locally compact Hausdorff space with countable basis)

can define a metric on $E$, Borel sets $B$

for $\mu$ signed measure on $E$ define

$$\| \mu \| = \sup_{A \in B} |\mu(A)|$$

Then

$$M_b(E) = \{ \mu : \| \mu \| \leq M < \infty \}$$

is a linear space

Put

$$C_c(E) = \{ f : E \to \mathbb{R} \text{ continuous with compact support} \}$$

vague convergence

$$\nu_n \overset{\nu}{\Rightarrow} \nu \iff \text{ for all } f \in C_c(E) : \int f d\nu_n \to \int f d\nu$$

Not equal to weak convergence: $\delta_n \overset{\nu}{\Rightarrow} 0$ but does not converge weakly

Class of test functions

$$(F) \quad F(\nu) = f \left( \int \phi_i d\nu, i = 1, \ldots, m \right)$$

for some $m, f \in C_b^\infty(\mathbb{R}^m)$ and $\phi_1, \ldots, \phi_m \in C_c(E)$

$$\mathcal{F} = \text{ class of these } F$$

Using Stone-Weierstrass we can show

Lemma 3 $\mathcal{F}$ is convergence-determining for vague convergence. So is the restricted class $\mathcal{F}_0$ that assumes that $\| f' \| \leq 1, \| f'' \| \leq 1, \| \phi_i \| \leq 1$ for $i = 1, \ldots, m$. Also, for $E = \mathbb{R}^d$ or connected open or closed subset of $\mathbb{R}^d$, could use $C_b^\infty$ instead of $C_c$. 

7
Here, $\| f' \| = \sum_{j=1}^{m} \| f(j) \|$.

Weak law of large numbers for empirical measures

$X_1, \ldots, X_n$ values in $E$

$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \mu_i$

$\xi_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$, empirical measure

Want to bound distance between $L(\xi_n)$ and $\delta_{\mu}$, say

Distance here

Generator for $F$ of form $(F)$

$AF(\nu) = F'(\nu)[\mu - \nu]$ Gateaux derivative

$= \sum_{j=1}^{m} f(j) \left( \int \phi_j d\nu, i = 1, \ldots, m \right) \left( \int \phi_j d\mu - \int \phi_j d\nu \right)$

Lemma 4 For every $H$ of the form $(F)$, with $h$ and $\phi_i, i = 1, \ldots, m$, the solution $F = F_H$ of the Stein equation is of the form $(F)$ with the same $\phi_i$’s. Furthermore, $\| f' \| \leq \| h' \|$, and $\| f'' \| \leq \| h'' \|$.

Proof like before.

Theorem 1 For all $H \in F$ we have

$|EH(\xi_n) - H(\mu)| \leq \sum_{j=1}^{m} \| h(j) \| \left( \int \phi_j d\bar{\mu} - \int \phi_j d\mu \right)$

$+ \sum_{j,k=1}^{m} \| h(j,k) \| \left\{ \max_{1 \leq j \leq m} \left[ \int \phi_j d\bar{\mu} - \int \phi_j d\mu \right] \right\}^2$

$+ \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i) \right) \leq \frac{1}{n} + 2 \sum_{i=1}^{n} |\Gamma_i|.$

Connection with mixing

Put $B_{i,j} = \{ A, B \in B : \mu_i(A) \neq 0, \mu_j(B) \neq 0 \}$

and

$\rho_n = \frac{1}{n^2} \sum_{i,j=1}^{m} \sup_{A,B \in B_{i,j}} |\text{Corr}(1(X_i \in A), 1(X_j \in B))|$

Then, if $\| \phi_j \| \leq 1$ for $i = 1, \ldots, m,$

$\text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i) \right) \leq 4\rho_n.$

Local approach

Assume that for all $i \in I = \{1, \ldots, n\}$ there is a set $\Gamma_i \subset I$ not containing $i$ such that $X_i$ is independent of $(X_j, j \notin \Gamma_i)$. Then, if $\| \phi_j \| \leq 1$ for $i = 1, \ldots, m,$

$\text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i) \right) \leq \frac{1}{n} + \frac{2}{n^2} \sum_{i=1}^{n} |\Gamma_i|.$
This could be extended to neighbourhoods of strong dependence.

**Example: A dissociated family**

Let \((Y_i)_{i \in \mathbb{N}}\) be a family of i.i.d. random elements on a space \(\mathcal{X}\), let \(k \in \mathbb{N}\) be fixed, and set

\[
\Gamma = \{(j_1, \ldots, j_k) \in \mathbb{N}^k : j_r \neq j_s \text{ for } r \neq s\};
\]
\[
\Gamma^{(n)} = \{(j_1, \ldots, j_k) \in \Gamma : j_1, \ldots, j_k \in \{1, \ldots, n\}\}.
\]

Suppose \(\psi\) is a measurable function \(\mathcal{X}^k \to E\), and put, for \((j_1, \ldots, j_k) \in \Gamma\),

\[
X_{j_1,\ldots,j_k} = \psi(Y_{j_1}, \ldots, Y_{j_k}).
\]

Then, \((X_{j_1,\ldots,j_k})_{(j_1,\ldots,j_k) \in \Gamma}\) is a dissociated family of identically distributed elements; put \(\mu = \mathcal{L}(X_{j_1,\ldots,j_k})\).

That is, if \(J \in \Gamma^{(n)}\) and \(K \in \Gamma^{(n)}\) are disjoint multi-indices, then \(X_J\) and \(X_K\) are independent.

For \(n \in \mathbb{N}\) fixed, the set \(\Gamma^{(n)}\) has \(n(n-1) \cdots (n-k+1)\) elements.

Let \(r(n) = n(n-1) \cdots (n-k+1)\), then

\[
\xi_n = \frac{1}{r(n)} \sum_{i=1}^{r(n)} \delta x_{i,n}
\]

**Theorem 2** For the above dissociated family, we have for \(H \in \mathcal{F}_0\)

\[
|EH(\xi_n) - H(\mu)| \leq \frac{1}{n} \left(1 + 2 \frac{k}{n-k+1}\right).
\]

**Proof:**

For \(J \in \Gamma^{(n)}\) set

\[
\Gamma(J) = \{L \in \Gamma^{(n)} : J \neq L, L \cap J \neq \emptyset\}
\]

Then

\[
|\Gamma(J)| = k \left(\frac{(n-1)!}{(n-k+1)!} - 1\right)
\]

\[
\leq \frac{r(n)}{n} k^2
\]

and

\[
\frac{1}{r(n)^2} \sum_{J \in \Gamma^{(n)}} |\Gamma(J)| \leq k \frac{(n-k)!(n-1)!}{n!(n-k+1)!}
\]

\[
\leq \frac{k}{n(n-k+1)}.
\]

Note that the \(X_{j_1,\ldots,j_k}\)'s are identically distributed, and thus \(\bar{\mu} = \mu\).

Can be extended to family of functions \((\psi_{j_1,\ldots,j_k})_{(j_1,\ldots,j_k) \in \Gamma}\) (R. 1994)

**Coupling approach**

**Excursion: size biasing for real-valued random variables**

Let \(W \geq 0\) and assume \(EW > 0\). Then \(W^*\) is said to have the \(W\)-size biased distribution if

\[
EW g(W) = EW E_{g(W^*)}
\]

for all \(g\) for which the expectations exist.
Example: If $W \sim Be(p)$ then

$$EWg(W) = pg(1)$$

so $W^* = 1$

Example: If $W \sim Po(\lambda)$ then from the Stein-Chen equation

$$EWg(W) = \lambda Eg(W + 1)$$

so $W^* = W + 1$ in distribution

In weak law of large numbers:

$$E,Af(W) = E(EW - W)f'(W) = EWf'(W) - Ef'(W^*)$$

Construction

(Goldstein + Rinott 1996) Suppose $W = \sum_{i=1}^n X_i$ with $X_i \geq 0$, $EX_i > 0$, all $i$. Choose index $V$ according to

$$P(V = v) = \frac{EX_v}{EW}$$

If $V = v$: replace $X_v$ by $X_v^*$ having the $X_v$-size biased distribution, independent if $X_v^* = x$: choose $\hat{X}_u, u \neq v$, such that

$$\mathcal{L}(\hat{X}_u, u \neq v) = \mathcal{L}(X_u, u \neq v|X_v = x)$$

Put $W^* = \sum_{u \neq V} \hat{X}_u + X_V$

Example: $X_i \sim Be(p_i)$ for $i = 1, \ldots, n$

If $V = v$: choose $\hat{X}_u, u \neq v$, such that

$$\mathcal{L}(\hat{X}_u, u \neq v) = \mathcal{L}(X_u, u \neq v|X_v = 1)$$

Then $W^* = \sum_{u \neq V} \hat{X}_u + 1$

See Poisson approximation, Barbour, Holst, Janson 1992

Size biasing for random measures

Let $\xi$ be a random measure on $E$, $E[\xi] = \mu$, let $\phi \in C_e$ be nonnegative with $\int \phi d\mu > 0$. We say that $\xi^\phi$ has the $\xi$ size biased distribution in direction $\phi$ if

$$EG(\xi) \int \phi d\xi = \int \phi d\mu EG(\xi^\phi)$$

for all $G$ for which the expectations exist.

Example: Suppose $\xi = \delta_X$, and $\mathcal{L}(X) = \mu$, and $\phi$ is one-to-one. Then

$$EG(\xi) \int \phi d\xi = E\phi(X)G(\delta_{\phi^{-1}(\phi(X))}) = \int \phi d\mu EG(\delta_{\phi^{-1}(\phi(X)^*)})$$

Construction

Let $\xi_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, and $E[\xi_n] = \bar{\mu}_n$
Pick $V \in \{1, \ldots, n\}$ according to

$$P(V = v) = \frac{E\phi(X_v)}{n\int \phi d\mu_n}$$

If $V = v$ take $\delta^X_v$ to have the $\delta_X$-size bias distribution in direction $\phi$.

If $\delta^X_v = \eta$ then choose $\delta^X_{u, u \neq v}$ according to

$$L(\delta^X_{u, u \neq v}) = L(\delta^X_v | \delta^X_v = \eta)$$

In generator with $F$ of form ($F$)

$$EA_F[\xi] = EF^\prime(\xi)\mu - \xi$$

$$= \sum_{j=1}^{m} \phi_j d\mu \left\{ f_j \left( \int \phi_i d\xi_{n, i}^\text{adj}, i = 1, \ldots, n \right) - f_j \left( \int \phi_i d\xi_{n, i} = 1, \ldots, n \right) \right\}$$

Construction depends on $\phi$, but when independent: need to adjust only one.

Remarks

1. Only gives vague/weak convergence; need additional argument for a.s. convergence
2. Could be viewed as a shorthand for multivariate law of large numbers
3. Will see a more involved example (epidemic) later

3. Discrete distributions from a Gibbs view point

joint work with Peter Eichelsbacher

Examples for Stein operators for discrete distributions

univariate case only

Poisson($\lambda$): $A_f(k) = \lambda f(k+1) - kf(k)$

Chen 1975

Binomial($n, p$): $A_f(k) = (n-k)pf(k+1) - k(1-p)f(k)$

Ehm 1991

Hypergeometric ($n, a, b$):

$$p_k = \binom{a}{k}\binom{b}{n-k}\binom{n+k}{n}^{-1}$$

$$A_f(k) = (n-k)(a-k)f(k+1) - k(b-n+k)f(k)$$

Künsch; R. + Schoutens (1998) preprint

Geometric($p$) with start at 0: for $f(0) = 0$

$A_f(k) = (1-p)f(k+1) - f(k)$

Peköz 1996

General pattern?

Connection with birth-death processes

See also Brown and Xia, Holmes, Weinberg

Discrete Gibbs measure $\mu$:

Assume supp($\mu$) = $\{0, \ldots, N\}$, where $N \in \mathbb{N}_0 \cup \{\infty\}$,
\[ \mu(k) = \frac{1}{\mathcal{Z}} \exp(V(k)) \frac{\omega^k}{k!}, \quad k = 0, 1, \ldots, N, \]

with \( \mathcal{Z} = \sum_{k=0}^{N} \exp(V(k)) \frac{\omega^k}{k!} \), where \( \omega > 0 \) is fixed.

Assume \( \mathcal{Z} \) exists.

**Example: Poissons-distribution**

\( \omega = \lambda, \quad V(k) = -\lambda, \quad k \geq 0, \quad \mathcal{Z} = 1 \)

or \( V(k) = 0, \quad \omega = \lambda, \quad \mathcal{Z} = e^\lambda \)

For a given probability distribution \( \mu(k) \),

\[ V(k) = \log \mu(k) + \log k! + \log \mathcal{Z} - k \log \omega, \quad k = 0, 1, \ldots, N, \]

with \( V(0) = \log \mu(0) + \log \mathcal{Z} \).

To each such Gibbs measure associate a birth-death process:

- **unit per-capita death rate** \( d_k = k \)
- **birth rate** \( b_k = \omega \exp\{V(k+1) - V(k)\} = (k+1) \frac{\mu(k+1)}{\mu(k)} \),

for \( k, k+1 \in \text{supp}(\mu) \)

then invariant measure \( \mu \)

**generator**

\( (Ah)(k) = (h(k+1) - h(k)) \exp\{V(k+1) - V(k)\} \omega + k(h(k-1) - h(k)) \)

or

\[ (Af)(k) = f(k+1) \exp\{V(k+1) - V(k)\} \omega - kf(k) \]


**Examples**

1. **Poisson-distribution** with parameter \( \lambda > 0 \): We use \( \omega = \lambda, V(k) = -\lambda, Z = 1 \). The Stein-operator is

\[ (Af)(k) = f(k+1) \lambda - kf(k) \]

2. **Binomial-distribution** with parameters \( n \) and \( 0 < p < 1 \): We use \( \omega = \frac{p}{1-p}, V(k) = -\log((n-k)!), \) and \( Z = (n!(1-p)^n)^{-1} \). The Stein-operator is

\[ (Af)(k) = f(k+1) \frac{p(n-k)}{(1-p)} - kf(k). \]

3. **Hypergeometric distribution**: The Stein-operator is

\[ (Af)(k) = f(k+1) (a-k)(n-k) - (b-n-x) kf(k). \]

4. **Pascal distribution** with parameter \( \gamma \in \{1, 2, \ldots\} \) and \( 0 < p < 1 \): \( \mu(k) = \binom{k+\gamma-1}{k} p^\gamma (1-p)^k \) for \( k = 0, 1, \ldots \)

We obtain the Stein-operator

\[ (Af)(k) = f(k+1) (1-p)(k + \gamma) - kf(k). \]
5. **Geometric distribution** with parameter $p$, shifted by one: $\gamma = n = 1$ in Pascal; $\mu(k) = p(1 - p)^k$ for $k = 0, 1, \ldots$

The Stein-operator is

$$(Af)(k) = f(k+1) (1 - p)(k+1) - kf(k).$$

**Bounds**

Solution of Stein equation $f$ for $h$: $f(0) = 0$, $f(k) = 0$ for $k \not\in \text{supp}(\mu)$, and

$$f(j+1) = \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \sum_{k=0}^{j} e^{V(k)} \frac{\omega^k}{k!} (h(k) - \mu(h)) = - \frac{j!}{\omega^{j+1}} e^{-V(j+1)} \sum_{k=j+1}^{N} e^{V(k)} \frac{\omega^k}{k!} (h(k) - \mu(h)).$$

**Lemma 5**

1. Put

$$M := \sup_{0 \leq k \leq N-1} \max \left( e^{V(k) - V(k+1)}, e^{V(k+1) - V(k)} \right).$$

Assume $M < \infty$. Then for every $j \in \mathbb{N}_0$:

$$|f(j)| \leq 2 \min \left( 1, \frac{M}{\sqrt{\omega}} \right).$$

2. Assume that the birth rates are non-increasing:

$$\exp(V(k+1) - V(k)) \leq \exp(V(k) - V(k-1)),$$

and death rates are unit per capita. For every $j \in \mathbb{N}_0$

$$|\Delta f(j)| \leq \frac{1}{j} \wedge \frac{e^{V(j)}}{\omega e^{V(j+1)}}.$$ 

**Examples**

1. **Poisson-distribution** with parameter $\lambda > 0$: non-uniform bound

$$|\Delta f(k)| \leq \frac{1}{k} \wedge \frac{1}{\lambda},$$

leads to $1 \wedge 1/\lambda$, see Barbour, Holst, Janson 1992

does not compare favourably to $1/\lambda(1 - e^{-\lambda})$

$$\| f \| \leq 2 \min \left( 1, \frac{1}{\sqrt{\lambda}} \right).$$

as in Barbour, Holst, Janson 1992

2. **Pascal distribution** with parameter $\gamma \in \{1, 2, \ldots\}$ and $0 < p < 1$:

$$|\Delta f(k)| \leq \frac{1}{k} \wedge \frac{1}{(1 - p)(k + \gamma)},$$

leads to $1 \wedge \frac{1}{(1 - p)\gamma}$

but $M = \infty$
Note that Brown and Xia (2001) give bounds for $\Delta f$ for a wide class of birth-death processes satisfying some monotonicity condition on the rates.

**Size-Bias coupling**

Recall: $W \geq 0, EW > 0$ then $W^*$ has the $W$-size biased distribution if

$$E_W g(W) = EW g(W^*)$$

for all $g$ for which both sides exist

so

$$E \left\{ \exp \{ V(X + 1) - V(X) \} \omega g(X + 1) - X g(X) \right\}$$

$$= E \left\{ \exp \{ V(X + 1) - V(X) \} \omega g(X + 1) - E X g(X^*) \right\}$$

and

$$EX = \omega E e^{V(X + 1) - V(X)} ,$$

so

**Lemma 6** Let $X \geq 0$ be such that $0 < E(X) < \infty$, let $\mu$ be a discrete Gibbs measure. Then $X \sim \mu$ if and only if for all bounded $g$

$$\omega E e^{V(X + 1) - V(X)} g(X + 1) = \omega E e^{V(X + 1) - V(X)} E g(X^*) .$$

Thus for any $W \geq 0$ with $0 < EW < \infty$

$${Eh(W) - \mu(h)}$$

$$= \omega \{ E e^{V(W + 1) - V(W)} f(W + 1) - E e^{V(W + 1) - V(W)} Ef(W^*) \}$$

where $f$ is the solution of the Stein equation.

Can also compare two discrete Gibbs distributions by comparing their birth rates and their death rates (see also Holmes)

Let $\mu$ have generator $A$ and corresponding $(\omega, V)$, and let $\mu_2$ have generator $A_2$, and corresponding $(\omega_2, V_2)$, both unit per-capita death rates. Then, for $X \sim \mu_2$, $f \in B$, if the solution $f$ of the Stein equation for $\mu$ is such that $A_2 f$ exists,

$${Eh(X) - \mu(h)}$$

$$= E Af(X)$$

$$= E (A - A_2) f(X)$$

$$= Ef(X + 1) (\omega e^{V(X + 1) - V(X)} - \omega_2 e^{V_2(X + 1) - V_2(X)})$$

$$= \omega Ef(X + 1) e^{V_2(X + 1) - V_2(X)} e^{V(X + 1) - V(X)} - (V_2(X + 1) - V_2(X))$$

$$= -E(X) Ef(X^*)$$

$$= \omega \{ E(X) E f(X^*) e^{V(X^*) - V(X^* - 1)} - (V_2(X^*) - V_2(X^* - 1)) \}$$

$$+ \omega \{ e^{V(X^*) - V(X^* - 1)} - (V_2(X^*) - V_2(X^* - 1)) \} - 1$$

14
Thus
\[
\left| Eh(X) - \int h d\mu \right| \\
\leq \| f \| E(X) \left\{ \frac{|\omega - \omega_2|}{\omega_2} + \frac{\omega}{\omega_2} \left| e^{(V(X^*) - V(X^* - 1)) - (v_2(X^*) - v_2(X^* - 1))} - 1 \right| \right\}
\]

Example: Poisson($\lambda_1$) and Poisson($\lambda_2$) gives
\[
\left| Eh(X) - \int h d\mu \right| \leq \| f \| |\lambda_1 - \lambda_2|
\]

Remarks:
1. The normalising constant $Z$ in the Gibbs distribution is often difficult to calculate. Note that it is not needed explicitly in the Stein approach.

4. An S-I-R epidemic

Bartlett (1949), Bailey (1975), Selke (1983)
Population: total size $K$
susceptibles (S), infected (I), removed(R); an individual is infectious when infected at time $t = 0$: $aK$ infected, $bK$ susceptible, $a + b = 1$

$(l_i, r_i)_{i \in \mathbb{N}}$ positive i.i.d. random vectors
$(\hat{r}_i)_{i \in \mathbb{N}}$ positive i.i.d.

Individual $i$:
if infected at time 0: stays infected for a period of length $\hat{r}_i$, then gets removed
if susceptible at time 0: gets infected at time $A_K^{-1} (l_i)$, stays infected for a period of length $r_i$, then gets removed

$l_i$ “resistance to infection”:

$Z_K(t)$ proportion infectives present at time $t$

$\lambda(t, (x(s))_{s \leq t})$ accumulation function

infectious pressure
\[
F_K(t) = \int_{[0,t]} \lambda(s, Z_K) ds
\]

\[
A_K^i = \inf \left\{ t \in \mathbb{R}_+ : \int_{[0,t]} \lambda(s, Z_K) ds = l_i \right\}
\]

Classical case: Bartlett’s GSE

$\lambda(t, x) = x(t)$
for each i, \( l_i \) and \( r_i \) are independent

results in a Markovian model

\[ \lambda(t, x) = \lambda(x(t)) \]

\( (l_i) \), i.i.d. \( \exp(1) \); \( (r_i) \), i.i.d. \( \exp(\rho) \)

for each i, \( l_i \) and \( r_i \) are independent

still Markovian structure

consider the vector of the proportion of S, I, R

**Here**

empirical measure

\[ \xi_K = \frac{1}{K} \sum_{i=1}^{aK} \delta_{[0, \hat{r}_i)} + \frac{1}{K} \sum_{i=1}^{bK} \delta_{[A^K_i, A^K_i + r_i)} \]

Note

\[ \xi_K([0, t] \times (t, \infty)) = \frac{1}{K} \sum_{i=1}^{aK} 1_{[0, \hat{r}_i]}(t) + \frac{1}{K} \sum_{i=1}^{bK} 1_{[A^K_i, A^K_i + r_i]}(t) \]

= the proportion of infected at time \( t \)

\[ \xi_K([0, s] \times (t, \infty)), \ t > s, \]

= the proportion of individuals: infected before time \( s \), not removed before time \( t \)

Limiting behaviour as \( K \to \infty \)?

→ mean-field approximation (deterministic system)

**Assumptions**

1. \( \lambda : \mathbb{R}_+ \times \mathbb{D}_+ \to \mathbb{R}_+ \) is uniformly bounded by a constant \( \gamma \), Lipschitz in \( x \in \mathbb{D}_+ \) with Lipschitz constant \( \alpha \), non-anticipating, and for all \( t \in \mathbb{R}_+ \)

\[ \lambda(t, x) = 0 \iff x(t) = 0. \]

2. There is a constant \( \beta > 0 \) such that, for each \( x \in \mathbb{R}_+ \), \( \Psi_x(t) := \mathbb{P}[l_1 \leq t | r_1 = x] \) has a density \( \psi_x(t) \) that is uniformly bounded from above by \( \beta; \)

\[ \psi_x(t) \leq \beta \] for all \( x \in \mathbb{R}_+, t \in \mathbb{R}_+ \).

3. \( (l_i) \), distribution function \( \Psi \) possessing a density \( \psi \).

4. \( r_i, \hat{r}_i \) distribution function \( \Phi, \hat{\Phi}(0) = 0 \) and \( \Phi(0) = 0 \), so that infected individuals do not immediately get removed.
Heuristics

\[ F_K(t) = \int_0^t \lambda(s, Z_K)ds \]

\[ Z_K(t) = \frac{1}{K} \sum_{i=1}^{a_K} 1(\hat{r}_i > t) + \frac{1}{K} \sum_{j=1}^{b_K} 1(F_K^{-1}(l_j) \leq t < F_K^{-1}(l_j) + r_j) \]

Define, for \( f \in C(\mathbb{R}_+, \mathbb{R}), t \in \mathbb{R}_+ \), operators

\[ Zf(t) = a(1 - \Phi(t)) + bP(f(t - r_1) \leq l_1 < f(t)) \]

\[ Lf(t) = \int_{(0,t]} \lambda(s, Zf)ds \]

Then

\[ F_K \approx LF_K \]

Results

Restrict everything to finite time interval \([0, T]\), \( T \) arbitrary (superscript \( T \), subscript \( T \))

**Theorem 3** For \( T \in \mathbb{R}_+ \), the operator \( L \) is a contraction on \([0, T]\), and the equation

\[ f(t) = \int_{(0,t]} \lambda(s, Zf)ds, \quad 0 \leq t \leq T, \]  

(1)

has a unique solution \( G_T \).

For \( T \in \mathbb{R}_+ \), let \( G_T \) be the solution of (1) and \( \tilde{\mu}^T \) be given for \( r, s \in (0, T] \) by

\[ \tilde{\mu}^T([0, r] \times [0, s]) = \mathbf{P}^T[l_1 \leq G_T(r), l_1 \leq G_T(s - r_1)]. \]

Put

\[ \mu^T = a(\delta_0 \times d\Phi)^T + b\tilde{\mu}^T. \]

**Theorem 4** For all \( T \in \mathbb{R}_+ \),

\[ \zeta_{\tau_0}(\mathbf{L}(X_n^T), \delta_{\mu^T}) \leq \frac{\sqrt{\alpha} + \sqrt{\beta}}{\sqrt{K}} + ab\beta T(T + 2) \exp(b[2\alpha\beta T]) \left\{ (1 + b)\sqrt{\frac{1}{K} + \frac{2}{K}} \right\}. \]

where \([x] \) is the smallest integer larger than \( x \).

Arguments:

Glivenko-Cantelli

Contraction theorem

\( F_K \) and \( l_1 \) are not independent, but if \( F_{K,1} \) denotes the similar quantity with individual 1 from the susceptible population left out, then \( F_{K,1} \) and \( l_1 \) are independent
Sketch of Proof

Abbreviate

\[ \zeta_K = \frac{1}{b_K} \sum_{i=1}^{b_K} \delta_{(A^i_K, A^i_K + r_i)} \]

and

\[ \langle \phi, \nu \rangle = \int \phi d\nu \]

Then we have

\[
\sum_{j=1}^{m} E f_j(\langle \xi_K^T, \phi_k \rangle, k = 1, \ldots, m) \mu^T - \xi_K^T, \phi_j) = a \sum_{j=1}^{m} E f_j(\langle \xi_K^T, \phi_k \rangle, k = 1, \ldots, m) (\delta_0 \times \hat{\mu})^T - \frac{1}{a K} \sum_{i=1}^{a K} \delta_{(0, \hat{\rho}_i)}^T, \phi_j) + b \sum_{j=1}^{m} E f_j(\langle \xi_K^T, \phi_k \rangle, k = 1, \ldots, m) (\hat{\mu}^T - \zeta_K^T, \phi_j).
\]

First summand: Cauchy-Schwarz and form (F) of functions

\[
\left\| a \sum_{j=1}^{m} E f_j(\langle \xi_K^T, \phi_k \rangle, k = 1, \ldots, m) (\delta_0 \times \hat{\mu})^T - \frac{1}{a K} \sum_{i=1}^{a K} \delta_{(0, \hat{\rho}_i)}^T, \phi_j) \right\| \leq a \sum_{j=1}^{m} \| f_j \| \| E \left\| \frac{1}{a K} \sum_{i=1}^{a K} (\phi_j(0, \hat{\rho}_i) - E\phi_j(0, \hat{\rho}_i)) \right\| \]

\[
\leq a \sum_{j=1}^{m} \| f_j \| \left( \text{Var} \left( \frac{1}{a K} \sum_{i=1}^{a K} (\phi_j(0, \hat{\rho}_i) - E\phi_j(0, \hat{\rho}_i)) \right) \right)^{\frac{1}{2}} \]

\[
\leq \frac{\sqrt{\alpha}}{\sqrt{K}}.
\]

Similarly

\[
b \sum_{j=1}^{m} E f_j(\langle \xi_K^T, \phi_k \rangle, k = 1, \ldots, m) (\hat{\mu}^T - \zeta_K^T, \phi_j)
\]

\[
= b \sum_{j=1}^{m} E f_j(\langle a(\delta_0 \times \hat{\mu})^T + b \zeta_K^T, \phi_k \rangle, k = 1, \ldots, m) (\hat{\mu}^T - \zeta_K^T, \phi_j) + R_1,
\]

where

\[ |R_1| \leq 2b \frac{\sqrt{\alpha}}{\sqrt{K}}.\]

For the remaining summand

\[
b \sum_{j=1}^{m} E f_j(\langle a(\delta_0 \times \hat{\mu})^T + b \zeta_K^T, \phi_k \rangle, k = 1, \ldots, m) (\hat{\mu}^T - \zeta_K^T, \phi_j)
\]

\[
\leq b \sum_{i=1}^{m} \| f_j \| \| E \left\| \hat{\mu}^T - \zeta_K^T, \phi_j \right\| \]

\[18\]
\[
\sum_{j=1}^{bK} \langle f^{(j)} \rangle - \sum_{i=1}^{aK} \phi_i \left( F^{T}\cdot l_i - \phi_j (G^{T}\cdot l_i) \right) - \sum_{j=1}^{bK} \langle f^{(j)} \rangle + r_i
\]

\[
\sum_{i=1}^{aK} \phi_i \left( F^{T}\cdot l_i - \phi_j (G^{T}\cdot l_i) + r_i \right)
\]

Problem: \(F_K\) and \(l_1\) are dependent
Introduce \(F_{K,1}\) like \(F_K\), with susceptible individual 1 omitted
Then \(F_{K,1}^{-1}(l_1) = F_{K,1}^{-1}(l_1)\) and
\[
E \left| (F_K^{T})^{-1}(l_1) - G_T^{-1}(l_1) \right| = E \left| (F_{K,1}^{T})^{-1}(l_1) - G_T^{-1}(l_1) \right|
\]

For \(h \in D([0,T])\), define operators
\[
Z_{K,1}(t) = 1_aK \sum_{i=1}^{aK} 1(\hat{r}_i > t) + 1_bK \sum_{j=2}^{bK} 1(h(t - r_j) < l_j \leq h(t))
\]

\[
L_{K,1}(t) = \int_{[0,t]} \lambda(s, Z_{K,1}) ds.
\]

Note that \(F_{K,1}^{-1}(l_1) = F_{K,1}^{-1}(l_1)\) by construction, and, for all \(t \leq T\),
\[
\left\| F_{K,1} - G_T \right\|_t = \left\| L_{K,1} F_{K,1} - L G_T \right\|_t
\]
\[
\leq \sup_h \left\| L_{K,1} - L \right\|_t + \left\| L F_{K,1} - L G_T \right\|_t
\]

For each \(h\)
\[
\left\| L_{K,1} - L \right\|_T \leq \alpha \int_0^T \sup_{s \leq x} |Z_{K,1}(h) - Z(h)| ds
\]
\[
\leq \alpha T \left( aR_1 + 2bR_2 + \frac{2}{K} \right),
\]

where
\[
R_1 = \sup_\varepsilon \left\{ \frac{1}{aK} \sum_{i=1}^{aK} 1(\hat{r}_i \leq \varepsilon) - \Phi(\varepsilon) \right\}
\]
and
\[
R_2 = \sup_\varepsilon \left\{ \frac{1}{bK - 1} \sum_{i=2}^{bK} 1(l_i \leq \varepsilon) - \Psi(\varepsilon) \right\}
\]

Massart (1990)
\[
ER_1 \leq \frac{1}{\sqrt{aK}}
\]
\[
ER_2 \leq \frac{1}{\sqrt{bK}}
\]

Thus
\[
E \sup_h \left\| L_{K,1} - L \right\|_T \leq \alpha T \left( (1 + b) \frac{1}{\sqrt{K}} + \frac{2}{K} \right) = S(K).
\]
For $E \| LF_{K,1} - LG_T \|_1$: Contraction argument

$$LF_{K,1}(t) - LG_T(t) \leq \alpha b \beta \int_0^t \| F_{K,1} - G_T \|_x (1 + \Phi(x))dx$$

where $\Phi(x) = P(r_1 \leq x)$.

Hence

$$E \| LF_{K,1} - LG_T \|_1 \leq S(K) + \alpha b \beta \int_0^t \| F_{K,1} - G_T \|_x (1 + \Phi(x))dx$$

Fix some $c \geq b$, put $\eta = \frac{1}{2c\alpha\beta}$

then

$$E \| LF_{K,1} - LG_T \|_1 \leq \frac{c}{c - b} S(K).$$

Induction:

$$E \| LF_{K,1} - LG_T \|_{k\eta} \leq \left( \frac{c}{c - b} \right)^k S(K).$$

Now $k = \lceil \frac{T}{\eta} \rceil$:

$$E \| LF_{K,1} - LG_T \|_{k\eta} \leq \exp(\lceil 2c\alpha\beta T \rceil)(\log c - \log(c - b)) S(K).$$

$c \to \infty$ gives the assertion.

**Remarks**

- First bound on distance at all, and explicit
- More realistic model than Markovian
- Factor $\frac{1}{\sqrt{K}}$ seems optimal - Gaussian approximation
- Waiting time until epidemic dies out is, very roughly, $\log K$, so deterministic approximation may not be good for whole time course
- When considering only a time interval when there is a substantial proportion of infectives present, then the bound on the approximation much improves, growing only linear in time. See R. 2001
- Initially infected: could assume that $(\hat{r}_i)_i$ are not identically distributed (and do not have distribution function $\Phi$)
- nonsmooth test functions
- spatial?

**5. The density approach**

*Stein 2003?*

**Situation:** Let $p$ be a strictly positive density on the whole real line having a derivative $p'$ in the sense that, for all $x$,

$$p(x) = \int_{-\infty}^x p'(y)dy = -\int_x^\infty p'(y)dy,$$
and assume that
\[ \int_{-\infty}^{\infty} |p'(y)| \, dy < \infty. \]

Let
\[ \psi(x) = \frac{p'(x)}{p(x)}. \]

**Proposition 1** Then, in order that a random variable \( Z \) be distributed according to the density \( p \) it is necessary and sufficient that, for all functions \( f \) that have a derivative \( f' \) and for which
\[ \int_{-\infty}^{\infty} |f'(z)| \, p(z) \, dz < \infty, \]
we have
\[ E(f'(Z) + \psi(Z) f(Z)) = 0. \]

**Example:** \( N(0,1) \)
\[ \psi(x) = -x, \] and the above condition is satisfied

gives classical Stein equation

**Example:** Gamma
\[ p_{\lambda,a}(x) = \frac{\lambda^a x^{a-1} e^{-\lambda x}}{\Gamma(a)}, \]
\[ \psi(x) = \frac{a-1-\lambda x}{x}, \]
the above condition is satisfied

This yields the characterization of type
\[ E f'(X) + \frac{a-1-\lambda X}{X} f(X) = 0. \]

Compare with the Luk-characterization:
equivalent; putting \( g(x) = xf(x) \)

Let for convenience
\[ \phi(x) = -\psi(x). \] (2)

**Theorem 5** Suppose \( Z \) has probability density function \( p \) satisfying the assumptions of the above proposition. Let \( (W,W') \) be an exchangeable pair such that \( E(\phi(W))^2 = \sigma^2 < \infty, \) and let
\[ \lambda = \frac{E(\phi(W') - \phi(W))^2}{2\sigma^2}. \]

Then, for all piecewise continuous functions \( h \) on \( \mathbb{R} \) to \( \mathbb{R} \) for which \( E|h(Z)| < \infty, \)
\[ Eh(W) - Eh(Z) = E(\phi(W)) \]
\[ -\frac{1}{\lambda \sigma^2} E(\phi(W') - \phi(W))(Eh(W') - (Uh)(W)) - EE^{-W} \left( \frac{\phi(W') - (1-\lambda)\phi(W)}{\lambda} \right) (Uh)(W), \]
where \( Uh \) and \( Vh \) are defined by
\[ (Uh)(w) = \int_{-\infty}^{w} \left( h(x) - \int_{-\infty}^{x} h(y)p(y) \, dy \right) p(x) \, dx \]
and
\[ (Vh)(w) = (Uh)'(w). \]
6. Distributional transformations

joint work with Larry Goldstein

The zero bias distributional transformation:

**Definition 1** Let $X$ be a mean zero random variable with finite, nonzero variance $\sigma^2$. We say that $X^*$ has the $X$-zero biased distribution if for all differentiable $f$ for which $E X f(X)$ exists,

$$EXf(X) = \sigma^2 Ef'(X^*).$$

(3)

The zero bias distribution $X^*$ exists for all $X$ that have mean zero and finite variance.

Goldstein and R. 1997

**General Biasing**

**Theorem 6** Let $m \in \{1, 2, \ldots\}$, and $P$ a function with exactly $m$ sign changes, positive on its rightmost interval. Then for every random variable $X$ with $EX^{2m} < \infty$ such that for some $\alpha > 0$,

$$\frac{1}{m!} EX^j P(X) = \alpha \delta_{j,m} \quad j = 0, \ldots, m,$$

there exists a random variable $X^{(m)}$, such that

$$EP(X)G(X) = \alpha E G^{(m)}(X^{(m)}) \quad \text{for all smooth } G.$$

The $X^{(m)}$ distribution is named the $X - P$ biased distribution.

**Example:** $P(x) = x$: for any variable $X$ such that $\sigma^2 = EX^2 < \infty$ and so that $EX = 0$, there exists a random variable $X^{(1)}$ such that, for all smooth $G$, we have $EXG(X) = \sigma^2 EG'(X^{(1)})$: zero bias distribution

**Biasing using orthogonal polynomials**

Suppose $P$ member of an orthogonal polynomial system

Consider infinitely divisible random variables $\{Z_\lambda\}_{\lambda > 0}$ so that if $Z_\lambda$ and $Z_\mu$ are independent, then their sum has distribution $Z_{\lambda + \mu}$.

Assume corresponding collection $\{P_\lambda^k\}_{k \geq 1}$ of polynomials where $P_\lambda^k$ has $k$ distinct roots, is positive on its rightmost interval, and the collection is orthogonal with respect to the law of $Z_\lambda$

Define

$$\alpha_\lambda^k = \frac{1}{k!} EZ_\lambda^k P_\lambda^k(Z_\lambda),$$

and

$$M_\lambda^k = \{X : EX^{2k} < \infty, \quad EX^j = EZ_\lambda^j, \quad 0 \leq j \leq 2k\}.$$

For every $X \in M_\lambda^k$, for $j = 0, \ldots, k$,

$$\frac{1}{k!} EX^j P_\lambda^k(X) = \frac{1}{k!} EZ_\lambda^j P_\lambda^k(Z_\lambda) = \alpha_\lambda^k \delta_{j,k}$$
Corollary 1  For all \( X \in M_{\lambda}^k \) there exists a random variable \( X_{\lambda}^k \) such that
\[
EP_{\lambda}^k (X) G(X) = \alpha^k_{\lambda} E G^k(X^k).
\]

Hence, if \( X_i \in M_{\lambda_i}^{m_i} \) and are independent, then letting \( x = (x_1, \ldots, x_n) \) and \( m = (m_1, \ldots, m_n) \) and defining
\[
\alpha_m^\lambda = \prod_{i=1}^n \alpha_{m_i}^{\lambda_i},
\]
and
\[
P_{m}^\lambda (x) = \prod_{i=1}^n P_{m_i}^\lambda (x_i),
\]
the vector
\[
X_m^\lambda = ((X_1)^{\lambda_1}_{m_1}, \ldots, (X_n)^{\lambda_n}_{m_n})
\]
satisfies
\[
E P_{m}^\lambda (X) G(X) = \alpha_m^\lambda E G^m (X_m^\lambda),
\]
where \( G : R^n \to R \) and
\[
G^m(x) = \frac{\partial m_1 + \cdots + m_n G(x)}{\partial x_1 \cdots \partial x_n}.
\]

Construction

Theorem 7  Let \( m \in \{0, 1, \ldots\} \). Let \( X_1, \ldots, X_n \) be independent variables with
\[
X_i \in M_{\lambda_i}^{m_i}
\]
for some \( \lambda_1, \ldots, \lambda_n \), and let \( \lambda = \lambda_1 + \cdots + \lambda_n \) and
\[
W = \sum_{i=1}^n X_i.
\]

Suppose that there exists weights \( w_m \) on non-negative integer sequences \( m = (m_1, \ldots, m_n) \) with \( m = m_1 + \cdots + m_n \) such that with \( w = x_1 + \cdots + x_n \) we have
\[
\left( \alpha_m^\lambda \right)^2 = \sum_m w_m \left( \alpha_m^\lambda \right)^2 \quad \text{and} \quad \alpha_m^\lambda P_m^\lambda (w) = \sum_m w_m \alpha_m^\lambda P_m^\lambda (x).
\]

Then, if \( I \) is independent of all other variables, with distribution
\[
P(I = m) = \frac{w_m \left( \alpha_m^\lambda \right)^2}{(\alpha_m^\lambda)^2},
\]
we have
\[
W_m^\lambda = \sum_m \left( X_i \right)^{I_i}_{\lambda_i}
\]

Examples

Hermite biasing:
For \( \sigma^2 = \lambda > 0 \), define the collection of Hermite polynomials \( \{ H^\lambda_n \}_{n \geq 0} \) through the generating function
\[
e^{-\frac{1}{2}t^2} = \sum_{n=0}^{\infty} \frac{H^\lambda_n(x)}{n!},
\]
23
Then Stein equation

\[ h(x) - Nh = \phi'(x)H_{m-1}(x) - H_m(x)\phi(x) \]

or

\[ h(x) - Nh = \phi^{(m)}(x) - H_m(x)\phi(x) \]

This gives an infinite number of Stein characterisations for the standard normal distribution.

Charlier biasing: For \( \lambda > 0 \), let \( \{C^\lambda_m\}_{m \geq 0} \) be the collection of Charlier polynomials defined by the generating function

\[ e^{-\sqrt{\lambda}t} \left( 1 + \frac{t}{\sqrt{\lambda}} \right)^x = \sum_{m=0}^{\infty} C^\lambda_m(x) \frac{t^m}{m!}. \]

Corresponds to Poisson distribution with parameter \( \lambda \)

Laguerre biasing: For \( \lambda > 0 \), let \( \{L^\lambda_m\}_{m \geq 0} \) be the collection of monic Laguerre polynomials defined by the generating function

\[ (1 + t)^{-\lambda} \exp \left\{ \frac{xt}{1 + t} \right\} = \sum_{m=0}^{\infty} L^\lambda_m(x) \frac{t^m}{m!} \]

corresponds to the Gamma distribution \( \propto x^{\lambda - 1}e^{-x} \)

see also Diaconis and Zabell 1991 for connections between distributions and orthogonal polynomials

Note that there are many other applications of Stein’s method to other distributions. Persi Diaconis’ work for probabilities on groups and for rates of convergence of Markov chains would be a good starting point.

**References**


