Stein's Method for Chisquare Approximations, Weak Law of Large Numbers, and Discrete Distributions from a Gibbs View Point

Gesine Reinert
Department of Statistics
University of Oxford

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(λ)
$$\lambda Ef(X+1) - EXf(X) = 0$$

General situation

Target distribution μ

1. Find characterization: operator $\mathcal{A}$ such that $X \sim \mu$ if and only if for all smooth functions f, $E\mathcal{A}f(X) = 0$
2. For each smooth function h find solution $f = f_h$ of the *Stein equation*
$$h(x) - \int h d\mu = \mathcal{A}f(x)$$
3. Then for any variable W,
$$Eh(W) - \int h d\mu = E\mathcal{A}f(W)$$

Usually need to bound f, f', or Δf
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 $\mathcal{A}$ as generator of a Markov process with stationary distribution μ
That is:
Let $(X_t)_{t \ge 0}$ be a homogeneous Markov process
Put $T_t f(x) = E(f(X_t)|X(0) = x)$
Generator $\mathcal{A}f(x) = \lim_{t \downarrow 0} \frac{1}{t}\left(T_t f(x) - f(x)\right)$
Facts (see *Ethier and Kurtz (1986)*, for example)

1. μ stationary distribution then $X \sim \mu$ if and only if $E\mathcal{A}f(X) = 0$ for f for which $\mathcal{A}f$ is defined
2. $T_t h - h = \mathcal{A}\left(\int_0^t T_u h du\right)$ and formally

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

   if the r.h.s. exists

Examples

1. $\mathcal{A}h(x) = h''(x) - xh'(x)$ generator of *Ornstein-Uhlenbeck* process, stationary distribution $\mathcal{N}(0, 1)$
2. $\mathcal{A}h(x) = \lambda(h(x+1) - h(x)) + x(h(x-1) - h(x))$ or

$$\mathcal{A}f(x) = \lambda f(x+1) - xf(x)$$

   Immigration-death process, immigration rate λ, unit per capita death rate; stationary distribution Poisson(λ) (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 $EX_i = 0, EX_i^2 = 1$

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 $\mathcal{A}_n$, with $\mathbf{x} = (x_1, \ldots\ x_n)$, f smooth,

$$\mathcal{A}_n f(\Phi_n(\mathbf{x})) = \frac{1}{n}\sum_{i=1}^n Ef(\Phi_n(x_1, \ldots, x_{i-1}, X_i^*, x_{i+1}, \ldots\ x_n)) - f(\Phi_n(\mathbf{x}))$$

Taylor expansion:

$$\begin{aligned}
\mathcal{A}_n f(\Phi_n(\mathbf{x})) &\approx \frac{1}{n}\sum_{i=1}^n E(X_i^* - x_i) f'(\Phi_n(\mathbf{x}))\frac{\partial}{\partial x_i}\Phi_n(\mathbf{x}) \\
&\quad + \frac{1}{2n}\sum_{i=1}^n E(X_i^* - x_i)^2 \left\{ f''(\Phi_n(\mathbf{x}))\left(\frac{\partial}{\partial x_i}\Phi_n(\mathbf{x})\right)^2 + f'(\Phi_n(\mathbf{x}))\frac{\partial^2}{\partial x_i^2}\Phi_n(\mathbf{x})\right\} \\
&= -\frac{1}{n}\sum_{i=1}^n x_i f'(\Phi_n(\mathbf{x}))\frac{\partial}{\partial x_i}\Phi_n(\mathbf{x}) + \frac{1}{2n}\sum_{i=1}^n (1 + x_i^2)\left\{ f''(\Phi_n(\mathbf{x}))\left(\frac{\partial}{\partial x_i}\Phi_n(\mathbf{x})\right)^2 \right. \\
&\quad \left. + f'(\Phi_n(\mathbf{x}))\frac{\partial^2}{\partial x_i^2}\Phi_n(\mathbf{x})\right\}.
\end{aligned}$$

Example: Put $\Phi_n(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i$

Then $\frac{\partial}{\partial x_i}\Phi_n(\mathbf{x}) = \frac{1}{\sqrt{n}}$ and

$$\frac{\partial^2}{\partial x_i^2}\Phi_n(\mathbf{x}) = 0$$

and

$$\begin{aligned}\mathcal{A}_n f(\Phi_n(\mathbf{x})) &\approx -\frac{1}{n^{3/2}}\sum_{i=1}^n x_i f'(\Phi_n(\mathbf{x})) + \frac{1}{2n^2}\sum_{i=1}^n (1+x_i^2) f''(\Phi_n(\mathbf{x})) \\ &= -\frac{1}{n}\Phi_n(\mathbf{x}) f'(\Phi_n(\mathbf{x})) + \frac{1}{2n} f''(\Phi_n(\mathbf{x}))\left(1 + \frac{1}{n}\sum_{i=1}^n x_i^2\right) \\ &\approx \frac{1}{n}\left(f''(\Phi_n(\mathbf{x})) - \Phi_n(\mathbf{x}) f'(\Phi_n(\mathbf{x}))\right)\end{aligned}$$

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. Chisquare distributions

Find generator

$\mathbf{X}_1, \ldots, \mathbf{X}_p$ i.i.d. random vectors, $\mathbf{X}_i = (X_{i,1}, \ldots X_{i,n})$ i.i.d., mean zero, $EX_{i,j}^2 = 1$, finite fourth moment

$$\Phi_n(\mathbf{x}) = \sum_{i=1}^p \left(\frac{1}{\sqrt{n}}\sum_{j=1}^n x_{i,j}\right)^2$$

Choose index uniformly from $\{1, \ldots, p\} \times \{1, \ldots, n\}$

We have

$$\frac{\partial}{\partial x_{i,j}}\Phi_n(\mathbf{x}) = \frac{2}{n}\sum_{k=1}^n x_{i,k}$$

and

$$\frac{\partial^2}{\partial x_{i,j}^2}\Phi_n(\mathbf{x}) = \frac{2}{n}$$

and

$$\begin{aligned}\mathcal{A}_n f(\Phi_n(\mathbf{x})) &\approx -\frac{2}{pn} f'(\Phi_n(\mathbf{x}))\sum_{i=1}^p\sum_{j=1}^n x_{i,j}\frac{1}{n}\sum_{k=1}^n x_{i,k} \\ &\quad + \frac{1}{2dn} f''(\Phi_n(\mathbf{x}))\sum_{i=1}^p\sum_{j=1}^n (1+x_{i,j}^2)\frac{4}{n^2}\left(\sum_{k=1}^n x_{i,k}\right)^2 \\ &\quad + \frac{1}{2dn} f'(\Phi_n(\mathbf{x}))\sum_{i=1}^p\sum_{j=1}^n (1+x_{i,j}^2)\frac{2}{n} \\ &\approx -\frac{2}{pn} f'(\Phi_n(\mathbf{x}))\Phi_n(\mathbf{x}) + \frac{4}{pn} f''(\Phi_n(\mathbf{x}))\Phi_n(\mathbf{x}) + \frac{2}{n} f'(\Phi_n(\mathbf{x}))\end{aligned}$$

by the law of large numbers

Suggests

$$\mathcal{A}f(x) = \frac{4}{p} x f''(x) + 2\left(1 - \frac{x}{p}\right) f'(x)$$

More convenient: Generator for χ_p^2

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

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

$$\mathcal{A}f(x) = xf''(x) + (r - \lambda x)f'(x)$$

and $\chi_p^2 = Gamma(d/2, 1/2)$

Luk also showed that, for χ_p^2, $\mathcal{A}$ 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

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

where $\chi_p^2 h$ is the expectation of h under the χ_p^2-distribution

Lemma 1 *(Pickett 2002)*
Suppose $h : \mathbf{R} \to \mathbf{R}$ is absolutely bounded, $|h(x)| \leq ce^{ax}$ for some $c > 0$ $a \in \mathbf{R}$, and the first k derivatives of h are bounded. Then the equation (χ_p^2) 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, exisiting 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

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

and

$$\begin{aligned} 2EWf''(W) + E(1-W)f'(W) &= Eg'(S) - Ef'(W) + E(1-W)f'(W) \\ &= Eg'(S) - ESg(S) \end{aligned}$$

Now proceed as in $\mathcal{N}(0,1)$:

Put

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

Then

$$\begin{aligned} ESg(S) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} EX_i g(S) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} EX_i g(S_i) + \frac{1}{n} \sum_{i=1}^{n} EX_i^2 g'(S_i) + R_1 \end{aligned}$$

where

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

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

From independence

$$\begin{aligned} ESg(S) &= \frac{1}{n} \sum_{i=1}^{n} Eg'(S_i) + R_1 \\ &= Eg'(S) + R_1 + R_2 \end{aligned}$$

where

$$\begin{aligned} R_2 &= \frac{1}{n^{3/2}} \sum_i EX_i g''(S_i) + \frac{1}{2n^2} \sum_i EX_i^2 g^{(3)} \left(S_i + \theta \frac{X_i}{\sqrt{n}} \right) \\ &= \frac{1}{2n^2} \sum_i EX_i^2 g^{(3)} \left(S_i + \theta \frac{X_i}{\sqrt{n}} \right) \end{aligned}$$

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 $\beta_i = EX_1^i$

$$\begin{aligned} & \frac{1}{2n^2} \sum_i EX_i^2 \left| g^{(3)}(S_i + \theta \frac{X_i}{\sqrt{n}}) \right| \\ \leq \; & \frac{24}{n} \parallel f^{(3)} \parallel \left(1 + \frac{\beta_4}{n} \right) + \frac{6}{n} \parallel f'' \parallel \\ & + \frac{8}{n} \parallel f^{(4)} \parallel \left(6 + \frac{\beta_4}{n} + 4 \frac{\beta_3^2}{\sqrt{n}} + 6 \frac{\beta_4}{n} + \frac{\beta_6}{n^2} \right) \\ = \; & c(f) \frac{1}{n}. \end{aligned}$$

Similarly for $\frac{1}{2n^2} \sum_i EX_i^4 \left| g^{(3)}(S_i + \theta \frac{X_i}{\sqrt{n}}) \right|$, employ β_8

For $\frac{1}{n^{3/2}} \sum_i EX_i^3 g''(S_i)$ have, for some $c(f)$

$$\frac{1}{n^{3/2}} \sum_i EX_i^3 g''(S_i) = \frac{1}{\sqrt{n}} \beta_3 Eg''(S) + c(f)\frac{1}{n}$$

and

$$Eg''(S) = 6ESf''(S^2) + 4ES^3 f^{(3)}(S^2)$$

Note that g'' is antisymmetric, $g''(-s) = -g''(s)$, so for $Z \sim \mathcal{N}(0,1)$ we have

$$Eg''(Z) = 0$$

(Almost) 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 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 δ_0, point mass at 0,

$$\mathcal{A}f(x) = -xf'(x)$$

and the corresponding transition semigroup is given by

$$T_t h(x) = h\left(xe^{-t}\right)$$

Stein equation for point mass at 0

$$(\delta_0) \quad h(x) - h(0) = -xf'(x)$$

Lemma 2 *If $h \in C_b^2(\mathbf{R})$, then the Stein equation (δ_0) has solution $f = f_h \in C_b^2$ such that*

$$\begin{aligned} \| f' \| &\leq \| h' \| \\ \| f'' \| &\leq \| h'' \| \end{aligned}$$

Proof

May assume $h(0) = 0$. Generator method gives

$$\begin{aligned} f(x) &= -\int_0^\infty h\left(xe^{-t}\right) dt \\ &= -\int_0^x \frac{h(t)}{t} dt \end{aligned}$$

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 $0 < \theta < 1$,

$$\begin{aligned} E\mathcal{A}f(W) &= -EWf'(W) \\ &= -EWf'(0) + EW^2 f''(\theta W) \\ &= EW^2 f''(\theta W) \end{aligned}$$

and

$$|E\mathcal{A}f(W)| \leq \| f'' \| Var(W).$$

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

Remarks

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

Empirical measures

$E = \mathbf{R}, \mathbf{R}^d, \mathbf{R}_+, \ldots$ (locally compact Hausdorff space with countable basis)
can define a metric on E, Borel sets $\mathcal{B}$
for μ signed measure on E define

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

Then

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

is a linear space

Put

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

vague convergence

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

Not equal to weak convergence: $\delta_n \stackrel{v}{\Rightarrow} 0$ 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 $m, f \in C_b^\infty(\mathbf{R}^m)$ and $\phi_i, \ldots, \phi_m \in C_c(E)$

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

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 $E = \mathbf{R}^d$ or connected open or closed subset of $\mathbf{R}^d$, could use C_b^∞ instead of C_c.*

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

Weak law of large numbers for empirical measures

$X_1, \ldots, X_n$ values in E
μ_i law of X_i
$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n \mu_i$
$\xi_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ *empirical measure*
Want to bound distance between $\mathcal{L}(\xi_n)$ and δ_μ, say
Distance here

$$\zeta(\mu, \nu) = \sup_{g \in \mathcal{F}_0} \left| \int g d\mu - \int g d\nu \right|$$

Generator for F of form (F)

$$\begin{aligned} \mathcal{A}F(\nu) &= F'(\nu)[\mu - \nu] \quad \text{Gateaux derivative} \\ &= \sum_{j=1}^m f_{(j)} \left(\int \phi_i d\nu, i = 1, \ldots, m \right) \left(\int \phi_j d\mu - \int \phi_j d\nu \right) \end{aligned}$$

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 ϕ_i's. Furthermore, $\|f'\| \leq \|h'\|$, and $\|f''\| \leq \|h''\|$.*

Proof like before.

Theorem 1 *For all $H \in \mathcal{F}$ we have*

$$\begin{aligned} |EH(\xi_n) - H(\mu)| &\leq \sum_{j=1}^m \| h_{(j)} \| \left| \int \phi_j d\bar{\mu} - \int \phi_j d\mu \right| \\ &+ \sum_{j,k=1}^m \| h_{(j,k)} \| \left\{ \max_{1 \leq j \leq m} \left[\int \phi_j d\bar{\mu} - \int \phi_j d\mu \right]^2 \right. \\ &\left. + Var \left(\frac{1}{n} \sum_{i=1}^n \phi_j(X_i) \right) \right\}. \end{aligned}$$

Connection with mixing

Put

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

and

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

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

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

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 $i = 1, \ldots, m$,

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

This could be extended to neighbourhoods of strong dependence.

Example: A dissociated family

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

$$\Gamma = \{(j_1, \dots, j_k) \in \mathbf{N}^k : j_r \neq j_s \text{ for } r \neq s\};$$
$$\Gamma^{(n)} = \{(j_1, \dots, j_k) \in \Gamma : j_1, \dots, j_k \in \{1, \dots, n\}\}.$$

Suppose ψ is a measurable functions $\mathcal{X}^k \to E$, and put, for $(j_1, \dots, j_k) \in \Gamma$,

$$X_{j_1,\dots,j_k} = \psi(Y_{j_1}, \dots, Y_{j_k}).$$

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

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 $n \in \mathbf{N}$ 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

$$\begin{aligned} |\Gamma(J)| &= k\left(\frac{(n-1)!}{(n-k+1)!} - 1\right) \\ &\leq \frac{r(n)}{n}k^2 \end{aligned}$$

and

$$\begin{aligned} \frac{1}{r(n)^2}\sum_{J\in\Gamma^{(n)}} |\Gamma(J)| &< k\frac{(n-k)!(n-1)!}{n!(n-k+1)!} \\ &\leq \frac{k}{n(n-k+1)}. \end{aligned}$$

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

Can be extended to family of functions $(\psi_{j_1,\dots,j_k})_{(j_1,\dots,j_k)\in\Gamma}$ (*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

$$EWg(W) = EWEg(W^*)$$

for all g for which the expectations exist.

Example: If $W \sim Be(p)$ then

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

so $W^* = 1$

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

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

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

In weak law of large numbers:

$$E\mathcal{A}f(W) = E(EW - W)f'(W) = EW(Ef'(W) - Ef'(W^*))$$

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 ξ be a random measure on E, $E[\xi] = \mu$, let $\phi \in C_c$ be nonnegative with $\int \phi d\mu > 0$.
We say that ξ^ϕ has the ξ size biased distribution in direction ϕ 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 ϕ is one-to-one. Then

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

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

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

If $V = v$ take $\delta^*_{X_v}$ to have the δ_{X_v}-size bias distribution in direction ϕ
If $\delta^*_{X_v} = \eta$ then choose $\hat{\delta}_{X_u}, u \neq v$ according to

$$\mathcal{L}(\hat{\delta}_{X_u}, u \neq v) = \mathcal{L}(\delta_{X_u}, u \neq v | \delta^*_{X_v} = \eta)$$

In generator with F of form (F)

$$\begin{aligned} E\mathcal{A}F[\xi] &= EF'(\xi)[\mu - \xi] \\ &= \sum_{j=1}^m \int \phi_j d\mu E\left\{ f_{(j)}\left(\int \phi_i d\xi_n^{\phi_j}, i = 1, \ldots, n \right) - f_{(j)}\left(\int \phi_i d\xi_n, i = 1, \ldots, n \right) \right\} \end{aligned}$$

Construction depends on ϕ, 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(λ): $\mathcal{A}f(k) = \lambda f(k+1) - kf(k)$
Chen 1975

Binomial(n, p): $\mathcal{A}f(k) = (n-k)pf(k+1) - k(1-p)f(k)$
Ehm 1991

Hypergeometric (n, a, b):

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

$\mathcal{A}f(k) = (n-k)(a-k)f(k+1) - k(b-n+k)f(k)$
Künsch; R. + Schoutens (1998) preprint

Geometric(p) with start at 0: for $f(0) = 0$
$\mathcal{A}f(k) = (1-p)f(k+1) - f(k)$
Peköz 1996

General pattern?
Connection with birth-death processes
See also *Brown and Xia, Holmes, Weinberg*

Discrete Gibbs measure μ:
Assume $\text{supp}(\mu) = \{0, \ldots, N\}$, where $N \in \mathbf{N}_0 \cup \{\infty\}$,

$$\mu(k) = \frac{1}{\mathbf{Z}} \exp(V(k)) \frac{\omega^k}{k!}, \quad k = 0, 1, \ldots, N,$$

with $\mathbf{Z} = \sum_{k=0}^{N} \exp(V(k)) \frac{\omega^k}{k!}$, where $\omega > 0$ is fixed
Assume $\mathbf{Z}$ exists

Example: $Po(\lambda)$
$\omega = \lambda$, $V(k) = -\lambda$, $k \geq 0$, $\mathbf{Z} = 1$
or $V(k) = 0$, $\omega = \lambda$, $\mathbf{Z} = e^{\lambda}$

For a given probability distribution $(\mu(k))_{k \in \mathbf{N}_0}$

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

with $V(0) = \log \mu(0) + \log \mathbf{Z}$

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 μ
generator

$$(\mathcal{A}h)(k) = (h(k+1) - h(k)) \exp\{V(k+1) - V(k)\}\omega + k(h(k-1) - h(k))$$

or

$$(\mathcal{A}f)(k) = f(k+1) \exp\{V(k+1) - V(k)\}\omega - kf(k)$$

Brown and Xia (2001) discuss many choices for birth and death rates. *Holmes (2003)* uses different birth and death rates.

Examples

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

$$(\mathcal{A}f)(k) = f(k+1)\, \lambda - kf(k)$$

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

$$(\mathcal{A}f)(k) = f(k+1)\, \frac{p(n-k)}{(1-p)} - kf(k).$$

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

$$(\mathcal{A}f)(k) = f(k+1)\,(a-k)(n-k) - (b-n-x)\,kf(k).$$

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

$$(\mathcal{A}f)(k) = f(k+1)\,(1-p)(k+\gamma) - kf(k).$$

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

$$(\mathcal{A}f)(k) = f(k+1)\,(1-p)(k+1) - kf(k).$$

Bounds

Solution of Stein equation f for h: $f(0) = 0$, $f(k) = 0$ for $k \notin \mathrm{supp}(\mu)$, and

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

Lemma 5 *1. Put*

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

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

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

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

$$\exp\bigl(V(k+1) - V(k)\bigr) \le \exp\bigl(V(k) - V(k-1)\bigr),$$

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

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

Examples

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

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

   leads to $1 \wedge 1/\lambda$, see *Barbour, Holst, Janson 1992*
   does not compare favourably to $1/\lambda(1 - e^{-\lambda})$
   $\| f \| \le 2 \min\left(1, \frac{1}{\sqrt{\lambda}}\right)$.
   as in *Barbour, Holst, Janson 1992*

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

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

   leads to $1 \wedge \frac{1}{(1-p)\gamma}$
   but $M = \infty$

Note that *Brown and Xia (2001)* give bounds for Δf 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

$$EWg(W) = EWEg(W^*)$$

for all g for which both sides exist

so

$$\begin{aligned} &E\Big\{\exp\{V(X+1)-V(X)\}\,\omega\, g(X+1) - X\, g(X)\Big\} \\ &= \quad E\Big\{\exp\{V(X+1)-V(X)\}\,\omega\, g(X+1) - EXEg(X^*)\Big\} \end{aligned}$$

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 μ be a discrete Gibbs measure. Then $X \sim \mu$ if and only if for all bounded g*

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

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

$$\begin{aligned} &Eh(W) - \mu(h) \\ &= \quad \omega\{E e^{V(W+1)-V(W)} f(W+1) - E e^{V(W+1)-V(W)} Ef(W^*)\} \end{aligned}$$

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 μ have generator $\mathcal{A}$ and corresponding (ω, V), and let μ_2 have generator $\mathcal{A}_2$, and corresponding (ω_2, V_2), both unit per-capita death rates. Then, for $X \sim \mu_2$, $f \in \mathcal{B}$, if the solution f of the Stein equation for μ is such that $\mathcal{A}_2 f$ exists,

$$\begin{aligned} &Eh(X) - \mu(h) \\ &= \quad E\mathcal{A}f(X) \\ &= \quad E(\mathcal{A} - \mathcal{A}_2)f(X) \\ &= \quad Ef(X+1)\left(\omega e^{V(X+1)-V(X)} - \omega_2 e^{V_2(X+1)-V_2(X)}\right) \\ &= \quad \omega Ef(X+1)e^{V_2(X+1)-V_2(X)} e^{V(X+1)-V(X)-(V_2(X+1)-V_2(X))} - E(X)Ef(X^*) \\ &= \quad \frac{\omega}{\omega_2} E(X)Ef(X^*)e^{(V(X^*)-V(X^*-1))-(V_2(X^*)-V_2(X^*-1))} \\ &\quad\quad -E(X)Ef(X^*) \\ &= \quad \frac{\omega-\omega_2}{\omega_2} E(X)Ef(X^*) + \frac{\omega}{\omega_2} E(X)Ef(X^*)\left\{e^{(V(X^*)-V(X^*-1))-(V_2(X^*)-V_2(X^*-1))} - 1\right\} \end{aligned}$$

Thus

$$\left|Eh(X) - \int hd\mu\right|$$
$$\leq \quad \| f \| \, E(X) \left\{ \frac{|\omega - \omega_2|}{\omega_2} + \frac{\omega}{\omega_2} E \left| e^{(V(X^*) - V(X^*-1)) - (V_2(X^*) - V_2(X^*-1))} - 1 \right| \right\}$$

Example: Poisson(λ_1) and Poisson(λ_2) gives

$$\left|Eh(X) - \int hd\mu\right| \quad \leq \quad \| f \| \, |\lambda_1 - \lambda_2|$$

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.
2. *Eichelsbacher and R. (2003)* generalise the above to point processes.

4. An S-I-R epidemic

Bartlett (1949), Bailey (1975), Sellke (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, $a + b = 1$

$(l_i, r_i)_{i \in \mathbf{N}}$ positive i. i. d. random vectors

$(\hat{r}_i)_{i \in \mathbf{N}}$ 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 $A_i^K = F_K^{-1}(l_i)$, 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

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

$$A_i^K = \inf \left\{ t \in \mathbf{R}_+ : \int_{(0,t]} \lambda(s, Z_K) ds = l_i \right\}$$

Classical case: Bartlett's GSE

$\lambda(t, x) = x(t)$

$(l_i)_i$ i. i. d. $exp(1)$; $(r_i)_i, (\hat{r}_i)_i$ i. i. d. $\exp(\rho)$

for each i, l_i and r_i are independent

results in a Markovian model

Wang (1975, 1977)

$\lambda(t, x) = \lambda(x(t))$

$(l_i)_i$ i. i. d. $exp(1)$

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

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

Note

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

= 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 $K \to \infty$?

$\to$ mean-field approximation (deterministic system)

Assumptions

1. $\lambda : \mathbf{R}_+ \times D_+ \to \mathbf{R}_+$ is uniformly bounded by a constant γ, Lipschitz in $x \in D_+$ with Lipschitz constant α, non-anticipating, and for all $t \in \mathbf{R}_+$
$$\lambda(t, x) = 0 \iff x(t) = 0.$$
2. There is a constant $\beta > 0$ such that, for each $x \in \mathbf{R}_+$, $\Psi_x(t) := \mathbf{P}[l_1 \leq t | r_1 = x]$ has a density $\psi_x(t)$ that is uniformly bounded from above by β;
$$\psi_x(t) \leq \beta \text{ for all } x \in \mathbf{R}_+, t \in \mathbf{R}_+.$$
3. $(l_i)_i$ distribution function Ψ possessing a density ψ.
4. r_i, $\hat{r}_i$ distribution function Φ; $\hat{\Phi}(0) = 0$ and $\Phi(0) = 0$, so that infected individuals do not immediately get removed.

Heuristics

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

$$Z_K(t) \quad = \quad \frac{1}{K}\sum_{i=1}^{aK} \mathbf{1}(\hat{r}_i > t) + \frac{1}{K}\sum_{j=1}^{bK} \mathbf{1}(F_K^{-1}(l_j) \le t < F_K^{-1}(l_j) + r_j)$$

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

$$\begin{aligned} \mathcal{Z}f(t) &= a(1 - \Phi(t)) + b\mathbf{P}(f(t - r_1) \le l_1 < f(t)) \\ Lf(t) &= \int_{(0,t]} \lambda(s, \mathcal{Z}f)ds \end{aligned}$$

Then

$$F_K \approx LF_K$$

Results

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

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

$$f(t) = \int_{(0,t]} \lambda(s, \mathcal{Z}f)ds, \quad 0 \le t \le T, \tag{1}$$

has a unique solution G_T.

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

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

Put

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

Theorem 4 *For all $T \in \mathbf{R}_+$,*

$$\zeta_{\mathcal{F}_0}(\mathcal{L}(\xi_n^T), \delta_{\mu^T}) \quad \le \quad \frac{\sqrt{a} + \sqrt{b}}{\sqrt{K}} + \alpha b\beta T(T + 2)\exp(b\lceil 2\alpha\beta T\rceil)\left\{(1 + b)\sqrt{\frac{1}{K}} + \frac{2}{K}\right\}.$$

where $\lceil x\rceil$ 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

$$\zeta_K = \frac{1}{bK}\sum_{i=1}^{bK}\delta_{(A_i^K, A_i^K + r_i)}$$

and

$$\langle \phi, \nu \rangle = \int \phi d\nu$$

Then we have

$$\begin{aligned}
&\sum_{j=1}^{m} \mathbf{E} f_{(j)}(\langle {\xi_K}^T, \phi_k\rangle, k=1,\ldots,m)\langle \mu^T - {\xi_K}^T, \phi_j\rangle \\
&= \quad a\sum_{j=1}^{m} \mathbf{E} f_{(j)}(\langle {\xi_K}^T, \phi_k\rangle, k=1,\ldots,m)\left\langle (\delta_0\times\hat{\mu})^T - \frac{1}{aK}\sum_{i=1}^{aK}\delta^T_{(0,\hat{r}_i)}, \phi_j\right\rangle \\
&\quad\quad + b\sum_{j=1}^{m} \mathbf{E} f_{(j)}(\langle {\xi_K}^T, \phi_k\rangle, k=1,\ldots,m)\langle \tilde{\mu}^T - \zeta_K^T, \phi_j\rangle.
\end{aligned}$$

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

$$\begin{aligned}
&\left| a\sum_{j=1}^{m} \mathbf{E} f_{(j)}(\langle {\xi_K}^T, \phi_k\rangle, k=1,\ldots,m)\langle (\delta_0\times\hat{\mu})^T - \frac{1}{aK}\sum_{i=1}^{aK}\delta^T_{(0,\hat{r}_i)}, \phi_j\rangle \right| \\
&\leq \quad a\sum_{j=1}^{m} \parallel f_{(j)} \parallel \mathbf{E}\left|\frac{1}{aK}\sum_{i=1}^{aK}(\phi_j(0,\hat{r}_i) - \mathbf{E}\phi_j(0,\hat{r}_i))\right| \\
&\leq \quad a\sum_{j=1}^{m} \parallel f_{(j)} \parallel \left(\mathrm{Var}\left(\frac{1}{aK}\sum_{i=1}^{aK}(\phi_j(0,\hat{r}_i) - \mathbf{E}\phi_j(0,\hat{r}_i))\right)^{\frac{1}{2}}\right) \\
&\leq \quad \frac{\sqrt{a}}{\sqrt{K}},
\end{aligned}$$

Similarly

$$\begin{aligned}
&b\sum_{j=1}^{m} \mathbf{E} f_{(j)}(\langle {\xi_K}^T, \phi_k\rangle, k=1,\ldots,m)\langle \tilde{\mu}^T - \zeta_K^T, \phi_j\rangle \\
&= \quad b\sum_{j=1}^{m} \mathbf{E} f_{(j)}(\langle a(\delta_0\times\hat{\mu})^T + b\zeta_K^T, \phi_k\rangle, k=1,\ldots,m)\langle \tilde{\mu}^T - \zeta_K^T, \phi_j\rangle + R_1,
\end{aligned}$$

where

$$|R_1| \leq 2b\frac{\sqrt{a}}{\sqrt{K}}.$$

For the remaining summand

$$\begin{aligned}
&b\sum_{j=1}^{m} \mathbf{E} f_{(j)}(\langle a(\delta_0\times\hat{\mu})^T + b\zeta_K^T, \phi_k\rangle, k=1,\ldots,m)\langle \tilde{\mu}^T - \zeta_K^T, \phi_j\rangle \\
&\leq \quad b\sum_{i=1}^{m} \parallel f_{(j)} \parallel \mathbf{E}\left|\langle \tilde{\mu}^T - \zeta_K^T, \phi_j\rangle\right|
\end{aligned}$$

$$
\begin{aligned}
&= b\sum_{j=1}^{m} \parallel f_{(j)} \parallel \mathbf{E}\left|\frac{1}{bK}\sum_{i=1}^{bK}\phi_i((F_K^T)^{-1}(l_i),(F_K^T)^{-1}(l_i)+r_i)-\phi_j(G_T^{-1}(l_i),G_T^{-1}(l_i)+r_i)\right| \\
&\leq b\sum_{j=1}^{m} \parallel f_{(j)} \parallel\parallel \phi_j' \parallel \mathbf{E}\left|(F_K^T)^{-1}(l_1)-(G_T)^{-1}(l_1)\right|.
\end{aligned}
$$

Problem: F_K and l_1 are dependent

Introduce $F_{K,1}$ like F_K, with susceptible individual 1 omitted

Then $F_K^{-1}(l_1) = F_{K,1}^{-1}(l_1)$ and

$$
\mathbf{E}\left|(F_K^T)^{-1}(l_1)-G_T^{-1}(l_1)\right| = \mathbf{E}\left|(F_{K,1}^T)^{-1}(l_1)-G_T^{-1}(l_1)\right|
$$

For $h \in D([0,T])$, define operators

$$
\begin{aligned}
\mathcal{Z}_{K,1}h(t) &= \frac{1}{K}\sum_{i=1}^{aK}\mathbf{1}(\hat{r}_i > t) + \frac{1}{K}\sum_{j=2}^{bK}\mathbf{1}(h(t-r_j) < l_j \leq h(t)) \\
L_{K,1}h(t) &= \int_{(0,t]}\lambda(s,\mathcal{Z}_{K,1}h)\,ds.
\end{aligned}
$$

Note that $F_K^{-1}(l_1) = F_{K,1}^{-1}(l_1)$ by construction, and, for all $t \leq T$,

$$
\begin{aligned}
\parallel F_{K,1} - G_T \parallel_t &= \parallel L_{k,1}F_{K,1} - LG_T \parallel_t \\
&\leq \sup_h \parallel L_{k,1}h - Lh \parallel_t + \parallel LF_{K,1} - LG_T \parallel_t
\end{aligned}
$$

For each h

$$
\begin{aligned}
\parallel L_{K,1}h - Lh \parallel_T &\leq \alpha\int_0^T \sup_{s\leq x}|\mathcal{Z}_{K,1}h(s)-\mathcal{Z}h(s)|ds \\
&\leq \alpha T\left(aR_1 + 2bR_2 + \frac{2}{K}\right),
\end{aligned}
$$

where

$$
R_1 = \sup_s\left|\frac{1}{aK}\sum_{i=1}^{aK}\mathbf{1}(\hat{r}_i \leq s) - \Phi(s)\right|
$$

and

$$
R_2 = \sup_s\left|\frac{1}{bK-1}\sum_{i=2}^{bK}\mathbf{1}(l_i \leq s) - \Psi(s)\right|.
$$

Massart (1990)

$$
\begin{aligned}
\mathbf{E}R_1 &\leq \frac{1}{\sqrt{aK}} \\
\mathbf{E}R_2 &\leq \frac{1}{\sqrt{bK}}
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathbf{E}\sup_h \parallel L_{K,1}h - Lh \parallel_T &\leq \alpha T\left\{(1+b)\frac{1}{\sqrt{K}} + \frac{2}{K}\right\} \\
&=: S(K).
\end{aligned}
$$

For $\mathbf{E} \parallel LF_{K,1} - LG_T \parallel_t$: Contraction argument

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

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

Hence

$$\mathbf{E} \parallel LF_{K,1} - LG_T \parallel_t \quad \leq \quad S(K) + \alpha b \beta \int_0^t \parallel F_{K,1} - G_T \parallel_x (1 + \Phi(x)) dx$$

Fix some $c \geq b$, put

$$\eta = \frac{1}{2c\alpha\beta}$$

then

$$\mathbf{E} \parallel LF_{K,1} - LG_T \parallel_\eta \quad \leq \quad \frac{c}{c-b} S(K).$$

Induction:

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

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

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

$c \to \infty$ 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 Φ)
- 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*

$$\int_{-\infty}^{\infty} \left| f'(z) \right| p(z) dz < \infty,$$

we have

$$\mathbf{E}(f'(Z) + \psi(Z) f(Z)) = 0.$$

Example: $\mathbf{N}(0,1)$
$\psi(x) = -x$, and the above condition is satisfied
gives classical Stein equation

Example: Gamma
$p_{\lambda,a}(x) = \frac{\lambda^a e^{-\lambda x} x^{a-1}}{\Gamma(a)}$
$\psi(x) = \frac{a-1-\lambda x}{x}$
the above condition is satisfied
This yields the characterization of type

$$\mathbf{E} f'(X) + \frac{a - 1 - \lambda X}{X} f(X) = 0.$$

Compare with the Luk-characterization:
equivalent; putting $g(x) = x f(x)$

Let for convenience

$$\phi(x) = -\psi(x). \tag{2}$$

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 $\mathbf{E}(\phi(W))^2 = \sigma^2 < \infty$, and let*

$$\lambda = \frac{\mathbf{E}(\phi(W') - \phi(W))^2}{2\sigma^2}.$$

Then, for all piecewise continuous functions h on $\mathbf{R}$ to $\mathbf{R}$ for which $\mathbf{E}|h(Z)| < \infty$,

$$\begin{aligned} \mathbf{E}h(W) - \mathbf{E}h(Z) &= \mathbf{E}(Vh)(W) \\ &\quad - \frac{1}{\lambda\sigma^2} \mathbf{E}(\phi(W') - \phi(W))((Uh)(W') - (Uh)(W)) - \mathbf{E}\mathbf{E}^W \left(\frac{\phi(W') - (1-\lambda)\phi(W)}{\lambda} \right) (Uh)(W), \end{aligned}$$

where Uh and Vh are defined by

$$(Uh)(w) = \frac{\int_{-\infty}^{z} \left(h(x) - \int_{-\infty}^{\infty} h(y) p(y) dy \right) p(x) dx}{p(z)}$$

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 σ^2. We say that X^* has the X-zero biased distribution if for all differentiable f for which $EXf(X)$ exists,*

$$EXf(X) = \sigma^2 Ef'(X^*). \tag{3}$$

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

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

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_λ and Z_μ are independent, then their sum has distribution $Z_{\lambda+\mu}$

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

Define

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

and

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

For every $X \in \mathcal{M}_k^\lambda$, for $j = 0, \ldots, k$,

$$\begin{aligned} \frac{1}{k!} EX^j P_k^\lambda(X) &= \frac{1}{k!} EZ_\lambda^j P_k^\lambda(Z_\lambda) \\ &= \alpha_k^\lambda \delta_{j,k} \end{aligned}$$

Corollary 1 *For all $X \in \mathcal{M}_k^\lambda$ there exists a random variable X_k^λ such that*

$$EP_k^\lambda(X)G(X) = \alpha_k^\lambda EG^{(k)}(X_k^\lambda).$$

Hence, if $X_i \in \mathcal{M}_{m_i}^{\lambda_i}$ and are independent, then letting $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{m} = (m_1, \ldots, m_n)$ and defining

$$\alpha_{\mathbf{m}}^\lambda = \prod_{i=1}^n \alpha_{m_i}^{\lambda_i},$$

and

$$P_{\mathbf{m}}^\lambda(\mathbf{x}) = \prod_{i=1}^n P_{m_i}^{\lambda_i}(x_i),$$

the vector

$$X_{\mathbf{m}}^\lambda = ((X_1)_{m_1}^{\lambda_1}, \ldots, (X_n)_{m_n}^{\lambda_n})$$

satisfies

$$EP_{\mathbf{m}}^\lambda(\mathbf{X})G(\mathbf{X}) = \alpha_{\mathbf{m}}^\lambda EG^{(\mathbf{m})}(X_{\mathbf{m}}^\lambda),$$

where $G : R^n \to R$ and

$$G^{(\mathbf{m})}(\mathbf{x}) = \frac{\partial^{m_1+m_2\ldots+m_n} G(\mathbf{x})}{\partial x_1^{m_1} \ldots \partial x_n^{m_n}}.$$

Construction

Theorem 7 *Let $m \in \{0, 1, \ldots\}$. Let $X_1, \cdots, X_n$ be independent variables with*

$$X_i \in \mathcal{M}_m^{\lambda_i}$$

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 $w_{\mathbf{m}}$ on non-negative integer sequences $\mathbf{m} = (m_1, \ldots, m_n)$ with $m = m_1 + \cdots + m_n$ such that with $w = x_1 + \cdots + x_n$ we have

$$\begin{aligned} \left(\alpha_m^\lambda\right)^2 &= \sum_{\mathbf{m}} w_{\mathbf{m}} \left(\alpha_{\mathbf{m}}^\lambda\right)^2 \quad \text{and} \\ \alpha_m^\lambda P_m^\lambda(w) &= \sum_{\mathbf{m}} w_{\mathbf{m}} \alpha_{\mathbf{m}}^\lambda P_{\mathbf{m}}^\lambda(\mathbf{x}). \end{aligned}$$

Then, if $\mathbf{I}$ is independent of all other variables, with distribution

$$P(\mathbf{I} = \mathbf{m}) = \frac{w_{\mathbf{m}}(\alpha_{\mathbf{m}}^\lambda)^2}{(\alpha_m^\lambda)^2},$$

we have

$$W_m^\lambda = \sum_{\mathbf{m}} (X_i)_{\lambda_i}^{(I_i)}$$

Examples

Hermite biasing:

For $\sigma^2 = \lambda > 0$, define the collection of Hermite polynomials $\{H_n^\lambda\}_{n \geq 0}$ through the generating function

$$e^{xt - \frac{1}{2}\lambda t^2} = \sum_{n=0}^\infty H_n^\lambda(x) \frac{t^n}{n!},$$

Then Stein equation

$$h(x) - \mathcal{N}h = \phi'(x)H_{m-1}(x) - H_m(x)\phi(x)$$

or

$$h(x) - \mathcal{N}h = \phi^{(m)}(x) - H_m(x)\phi(x)$$

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

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

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

Corresponds to Poisson distribution with parameter λ

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

$$(1+t)^{-\lambda} \exp\left\{\frac{xt}{1+t}\right\} = \sum_{m=0}^{\infty} L_m^\lambda(x)\frac{t^m}{m!}$$

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

Barbour, A.D. (1974). On a functional central limit theorem for Markov population processes. *Adv. Appl. Probab.* **6**, 21–39.

Barbour, A.D. (1975). The duration of the closed stochastic epidemic. *Biometrika* **62**, 477–482.

Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford Science Publications.

Brown, T. and Xia, A. (2001). Stein's method and birth-death processes. *Ann. Probab.* **29**, 1373-1403.

Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

Dawson, D.A. (1993). Measure-valued Markov processes. In *Ecole d'Eté de Probabilité de Saint-Flour 1991*, LNM 1541, Hennequin, P.L., eds., Springer-Verlag, 1–260.

Diaconis, P., and Zabell, S. (1991). Closed form summation for classical distributions: variations on a theme of de Moivre. *Statistical Science* **6**, 284-302.

Eichelsbacher, P. and Reinert, G. (2003). Stein's method for discrete Gibbs measures. Submitted.

Eichelsbacher, P. and Reinert, G. (2003). Stein's method for spatial Gibbs measures. Preprint.

Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statist. Probab. Lett.* **11**, 7-16.

Ethier, S.N. and Kurtz, T. (1986). *Markov Processes, Characterization and Convergence.* Wiley, New York.

Goldstein, L., and Reinert, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935-952.

Goldstein, L. and Reinert, G. (2003). Distributional transformations. Preprint.

Goldstein, L., and Rinott, Y. (1996). On multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1–17.

Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19**, 724–739.

Holmes, S. (2003). Stein's method for birth and death chains. Preprint.

Luk, H.M. (1994). Stein's method for the gamma distribution and related statistical applications. Ph.D. thesis. University of Southern California, Los Angeles, USA.

Peköz, E. (1996). Stein's method for geometric approximation. *J. Appl. Probab.* **33**, 707-713.

Pickett, A. (2002). Stein's method for chisquare approximations. Transfer thesis, University of Oxford.

Pickett, A. and Reinert, G. (2003). Stein's method for chisquare approximations. Preprint.

Reinert, G. (1994). A weak law of large numbers for empirical measures via Stein's method. *Ann. Probab.* **23**, 334–354.

Reinert, G. (1995) The asymptotic evolution of the General Stochastic Epidemic. *Ann. Appl. Probab.* **5**, 1061–1086.

Reinert, G. (1998). Stein's method in application to empirical measures. In *Modelos Estocasticos.* J.M. Gonzales Barrios and L.G.Gorostiza, eds. Sociedad Matematica Mexicana, 65–120.

Reinert, G. (2001). Stein's method in application for epidemic processes. In *Complex Stochastic Systems.* O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg, eds., Chapman and Hall, Boca Raton etc., 235-275.

Reinert, G. and Schoutens, W. (1998). Stein's method for the hypergeometric. Preprint.

Stein, C. (1986). *Approximate Computation of Expectations.* IMS, Hayward, California.

Stein, C. (2003). Use of Normal Approximations in the Analysis of Simulations. Preprint.

Weinberg, G.V. (2000). Stein factor bounds for random variables. *J. Appl. Probab.* **37**, 1181-1187.