# A Short Introduction to Stein's Method

Gesine Reinert

Department of Statistics

University of Oxford

## Overview

Lecture 1: focusses mainly on normal approximation

Lecture 2: other approximations

## 1. The need for bounds

Distributional approximations:

*Example* $X_1, X_2, \ldots, X_n$ i.i.d., $\mathbf{P}(X_i = 1) = p = 1 - \mathbf{P}(X_i = 0)$

$$n^{-1/2} \sum_{i=1}^{n} (X_i - p) \approx_d \mathcal{N}(0, p(1-p))$$

$$\sum_{i=1}^{n} X_i \approx_d Poisson(np)$$

would like to assess distance of distributions; would like bounds

*Weak convergence*

For c.d.f.s $F_n, n \geq 0$ and $F$ on the line we say that $F_n$

converges weakly (converges in distribution) to $F$,

$$F_n \xrightarrow{w} F$$

if

$$F_n(x) \rightarrow F(x) \quad (n \rightarrow \infty)$$

for all continuity points $x$ of $F$

For the associated probability distributions:

$$P_n \xrightarrow{w} P$$

Facts:

1.

$$P_n \xrightarrow{w} P \iff P_n(A) \to P(A)$$

for each $P$-continuity set $A$ (i.e. $P(\partial A) = 0$)

2.

$$P_n \xrightarrow{w} P \iff \int f dP_n \to \int f dP$$

for all functions $f$ that are bounded, continuous, real-valued

3.

$$P_n \xrightarrow{w} P \iff \int f dP_n \to \int f dP$$

for all functions $f$ that are bounded, infinitely often differentiable, continuous, real-valued

4. If $X$ is a random variable, denote its distribution by $\mathcal{L}(X)$. Then

$$\mathcal{L}(X_n) \xrightarrow{w} \mathcal{L}(X) \iff Ef(X_n) \to Ef(X)$$

for all functions $f$ that are bounded, infinitely often differentiable, continuous, real-valued

*Metrics*

Let $\mathcal{L}(X) = P, \mathcal{L}(Y) = Q$; define *total variation distance*

$$d_{TV}(P, Q) = \sup_{A \text{ measurable}} |P(A) - Q(A)|$$

Put

$$\mathcal{L} = \{g : \mathbf{R} \to \mathbf{R}; |g(y) - g(x)| \le |y - x|\}$$

and *Wasserstein distance*

$$\begin{aligned} d_W(P, Q) &= \sup_{g \in \mathcal{L}} |\mathbf{E}g(Y) - \mathbf{E}g(X)| \\ &= \inf \mathbf{E}|Y - X|, \end{aligned}$$

where the infimum is over all *couplings* $X, Y$ such that

$\mathcal{L}(X) = P, \mathcal{L}(Y) = Q$

Using

$$\mathcal{F} = \{f \in \mathcal{L} \text{ absolutely continuous}, f(0) = f'(0) = 0\}$$

we also have

$$d_W(P, Q) = \sup_{f \in \mathcal{F}} |\mathbf{E}f'(Y) - \mathbf{E}f'(X)|.$$

## 2. Stein's Method for Normal Approximation

*Stein (1972, 1986)*

$Z \sim \mathcal{N}(\mu, \sigma^2)$ if and only if for all smooth functions $f$,

$$\mathbf{E}(Z-\mu)f(Z) = \sigma^2 \mathbf{E} f'(Z)$$

For $W$ with $\mathbf{E}W = \mu, \mathrm{Var}W = \sigma^2$, if

$$\sigma^2 \mathbf{E} f'(W) - \mathbf{E}(W-\mu)f(W)$$

is close to zero for many functions $f$, then $W$ should be close to $Z$ in distribution

*Sketch of proof for $\mu = 0, \sigma^2 = 1$:*

Assume $Z \sim \mathcal{N}(0,1)$. Integration by parts:

$$\frac{1}{\sqrt{2\pi}} \int f'(x) e^{-x^2/2} dx$$

$$= \left[ \frac{1}{\sqrt{2\pi}} f(x) e^{-x^2/2} \right] + \frac{1}{\sqrt{2\pi}} \int x f(x) e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int x f(x) e^{-x^2/2} dx$$

Assume $\mathbf{E} Z f(Z) = \mathbf{E} f'(Z)$: Can use partial integration to solve differential equation

$$f'(x) - x f(x) = g(x), \quad lim_{x \to -\infty} f(x) e^{-x^2/2} = 0$$

for any bounded function g, giving

$$f(y) = e^{y^2/2} \int_{-\infty}^{y} g(x) e^{-x^2/2} dx$$

Take $g(x) = \mathbf{1}(x \le x_0) - \Phi(x_0)$, then

$$0 = \mathbf{E}(f'(Z) - Zf(Z)) = \mathbf{P}(Z \le x_0) - \Phi(x_0)$$

so $Z \sim \mathcal{N}(0, 1)$.

Let $\mu = 0$. Given a test function $h$, let $Nh = \mathbf{E}h(Z/\sigma)$,

and solve for $f$ in the *Stein equation*

$$\sigma^2 f'(w) - wf(w) = h(w/\sigma) - Nh$$

giving

$$f(y) = e^{y^2/2} \int_{-\infty}^{y} (h(x/\sigma) - Nh)\, e^{-x^2/2} dx$$

Now evaluate the expectation of the r.h.s. of the Stein

equation by the expectation of the l.h.s.

Can bound, e.g. $\| f'' \| \leq 2 \| h' \|$

*Example:* $X, X_1, \ldots, X_n$ i.i.d. mean zero, $\mathrm{Var} X = \frac{1}{n}$

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

Put

$$W_i = W - X_i = \sum_{j \neq i} X_j$$

Then

$$\begin{aligned}
\mathbf{E} W f(W) &= \sum_{i=1}^{n} \mathbf{E} X_i f(W) \\
&= \sum_{i=1}^{n} \mathbf{E} X_i f(W_i) + \sum_{i=1}^{n} \mathbf{E} X_i^2 f'(W_i) + R \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbf{E} f'(W_i) + R
\end{aligned}$$

So

$$\mathbf{E}f'(W) - \mathbf{E}Wf(W) \; = \; \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}\{f'(W) - f'(W_i)\}$$

$$+R$$

and can bound remainder term $R$;

**Theorem 1** *For any smooth $h$*

$$|\mathbf{E}h(W) - Nh| \; \leq \; \|h'\| \left( \frac{2}{\sqrt{n}} + \sum_{i=1}^{n} \mathbf{E}|X_i^3| \right).$$

Extends to local dependence:

Let $X_1, \ldots, X_n$ be mean zero, finite variances, put

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

Assume $VarW = 1$. Suppose that for each $i = 1, \ldots, n$

there exist sets $A_i \subset B_i \subset \{1, \ldots, n\}$ such that

$X_i$ is independent of $\Sigma_{j \notin A_i} X_j$ and

$\Sigma_{j \in A_i} X_j$ is independent of $\Sigma_{j \notin B_i} X_j$

Define

$$\eta_i = \sum_{j \in A_i} X_j$$

$$\tau_i = \sum_{j \notin B_i} X_j$$

**Theorem 2** *For any smooth $h$ with $\|h'\| \leq 1$,*

$$|\mathbf{E}h(W) - Nh| \;\leq\; 2\sum_{i=1}^{n}\left(\mathbf{E}|X_i\eta_i\tau_i| + |\mathbf{E}(X_i\eta_i)|\mathbf{E}|\tau_i|\right)$$

$$+\sum_{i=1}^{n}\mathbf{E}|X_i\eta_i^2|.$$

*Example: Graphical dependence*

$V = \{1, \dots, n\}$ set of vertices in graph $G = (V, E)$

$G$ is a *dependency graph* if, for any pair of disjoint sets $\Gamma_1$ and $\Gamma_2$ of $V$ such that no edge in $E$ has one endpoint in $\Gamma_1$ and the other endpoint in $\Gamma_2$, the sets of random variables $\{X_i, i \in \Gamma_1\}$ and $\{X_i, i \in \Gamma_2\}$ are independent

Let $A_i$ be the set of all $j$ such that $(i, j) \in E$, union with $\{i\}$, $B_i = \cup_{j \in A_i} A_j$. Then the above theorem applies.

*Size-Bias coupling:* $\mu > 0$

If $W \geq 0, \mathbf{E}W > 0$ then $W^s$ has the $W$-*size biased distribution* if

$$\mathbf{E}Wf(W) = \mathbf{E}W\mathbf{E}f(W^s)$$

for all $f$ for which both sides exist

Put $f(x) = \mathbf{1}(x = k)$ then

$$k\mathbf{P}(W = k) = \mathbf{E}W\mathbf{P}(W^s = k)$$

so

$$\mathbf{P}(W^s = k) = \frac{k\mathbf{P}(W = k)}{\mathbf{E}W}$$

*Example:* If $X \sim Bernoulli(p)$, then $\mathbf{E}Xf(X) =$

$pf(1)$ and so $X^s = 1$

*Example:* If $X \sim Poisson(\lambda)$, then

$$\mathbf{P}(X^s = k) = \frac{ke^{-\lambda}\lambda^k}{k!\lambda} = \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!}$$

and so $X^s = X + 1$, where the equality is in distribution

*Construction*

(*Goldstein + Rinott 1996*) Suppose $W = \Sigma_{i=1}^{n} X_i$ with $X_i \geq 0$, $EX_i > 0$, all $i$.

Choose index $V$ proportional to the mean, $\mathbf{E}X_v$. If $V = v$: replace $X_v$ by $X_v^s$ having the $X_v$-size biased distribution, independent, and if $X_v^s = x$: adjust $\hat{X}_u, u \neq v$, such that

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

Then $W^s = \Sigma_{u \neq V} \hat{X}_u + X_V^s$

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

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

See Poisson approximation, *Barbour, Holst, Janson 1992*

$$X, X_1, \ldots, X_n \geq 0 \text{ i.i.d.}, \mathbf{E}X = \mu, \mathrm{Var}X = \sigma^2$$

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

Then

$$\begin{aligned}
\mathbf{E}(W-\mu)f(W) &= \mu\mathbf{E}(f(W^s) - f(W)) \\
&\approx \mu\mathbf{E}(W^s - W)f'(W) \\
&= \mu\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}(X_i^s - X_i)f'(W) \\
&\approx \mu\mathbf{E}f'(W)\mathbf{E}(X^s - \mu) \\
&= \mu\mathbf{E}f'(W)\left\{\frac{1}{\mu}\mathbf{E}X^2 - \mu\right\} \\
&= \sigma^2\mathbf{E}f'(W).
\end{aligned}$$

*Zero bias coupling*

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 $\mathbf{E}Xf(X)$ exists,

$$\mathbf{E}Xf(X) = \sigma^2 \mathbf{E}f'(X^*).$$

The zero bias distribution $X^*$ exists for all $X$ that have mean zero and finite variance. (*Goldstein and R. 1997*)

It is easy to verify that $W^*$ has density

$$p^*(w) = \sigma^{-2}\mathbf{E}\{W\mathbf{I}(W > w)\}$$

*Example:* If $X \sim Bernoulli(p) - p$, then

$$\mathbf{E}\{X\mathbf{I}(X > x)\} = p(1-p) \text{ for } -p < x < 1-p$$

and is zero elsewhere, so $X^* \sim Uniform(-p, 1-p)$

*Connection with Wasserstein distance*

We have for $W$ mean zero, variance 1,

$$
\begin{aligned}
|\mathbf{E}h(W) - Nh| &= |\mathbf{E}\left[f'(W) - Wf(W)\right]| \\
&= |\mathbf{E}\left[f'(W) - f'(W^*)\right]| \\
&\leq ||f''||\mathbf{E}|W - W^*|,
\end{aligned}
$$

where $||\cdot||$ is the supremum norm. As $||f''|| \leq 2||h'||$

$$|\mathbf{E}h(W) - Nh| \leq 2||h'||\mathbf{E}|W - W^*|;$$

thus

$$d_W(\mathcal{L}(W), \mathcal{N}(0,1)) \leq 2\mathbf{E}|W - W^*|.$$

*Construction* in the case of $W = \Sigma_{i=1}^n X_i$ sum of independent mean zero finite variance $\sigma_i^2$ variables: Choose an index $I$ proportional to the variance, zero bias in that variable,

$$W^* = W - X_I + X_I^*.$$

*Proof:* For any smooth $f$,

$$\begin{aligned}
\mathbf{E}Wf(W) &= \sum_{i=1}^{n} \mathbf{E}X_i f(W) \\
&= \sum_{i=1}^{n} \mathbf{E}X_i f(X_i + \sum_{t \neq i} X_t) \\
&= \sum_{i=1}^{n} \sigma_i^2 \mathbf{E}f'(X_i^* + \sum_{t \neq i} X_t) \\
&= \sigma^2 \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma^2} \mathbf{E}f'(W - X_i + X_i^*) \\
&= \sigma^2 \mathbf{E}f'(W - X_I + X_I^*) = \sigma^2 \mathbf{E}f'(W^*),
\end{aligned}$$

where we have used independence of $X_i$ and $X_t, t \neq i$.

Immediate consequence:

$$\mathbf{E}|W - W^*| \le \frac{1}{\sigma^2} \sum_{i=1}^n \sigma_i^2 \{\mathbf{E}|X_i| + \mathbf{E}|X_i^*|\}$$

and

**Proposition 1** *Let $X_1, \ldots, X_n$ be independent mean zero variables with variances $\sigma_1^2, \ldots, \sigma_n^2$ and finite third moments, and let $W = (X_1 + \ldots + X_n)/\sigma$ where $\sigma^2 = \sigma_1^2 + \ldots + \sigma_n^2$. Then for all absolutely continuous test functions $h$,*

$$|\mathbf{E}h(W) - Nh| \le \frac{2||h'||}{\sigma^3} \sum_{i=1}^n \mathbf{E}\left(|X_i| + \frac{1}{2}|X_i|^3\right) \sigma_i^2,$$

*so in particular, when the variables are identically distributed with variance 1,*

$$|\mathbf{E}h(W) - Nh| \leq \frac{3||h'||\mathbf{E}|X_1|^3}{\sqrt{n}}.$$

*General construction:* (*Goldstein + R. 1997, Goldstein 2003*)

Let $Y', Y''$ be exchangeable pair with distribution $F(y', y'')$ such that

$$\mathbf{E}(Y''|\mathcal{F}) = (1-\lambda)Y'$$

for some $\mathcal{F}$ such that $\sigma(Y') \subset \mathcal{F}$, and for some $0 < \lambda < 1$

Let $\hat{Y}', \hat{Y}''$ have distribution

$$dG(\hat{y}', \hat{y}'') = \frac{(\hat{y}' - \hat{y}'')^2}{\mathbf{E}(\hat{Y}' - \hat{Y}'')^2} dF(\hat{y}', \hat{y}'')$$

and let $U \sim \mathcal{U}(0,1)$ be independent of $\hat{Y}', \hat{Y}''$, then

$$Y^* = U\hat{Y}' + (1-U)\hat{Y}''$$

has the $Y^*$-distribution.

If in addition $Y' = V + T'$ and $Y'' = V + T''$ for some $T', T''$, and on the same state space $\hat{Y}' = V + \hat{T}'$ and $\hat{Y}'' = V + \hat{T}''$ for some $\hat{T}', \hat{T}''$ with $|\hat{T}'| \leq B$ and $|\hat{T}''| \leq B$, then we can couple such that

$$|Y' - Y^*| \leq 3B.$$

*Example: Simple random sampling*

Population of characteristics $\mathcal{A}$ with $\Sigma_{a\in\mathcal{A}}\, a = 0$ and

$|a| \leq n^{-1/2}B$ for all $a \in \mathcal{A}$. Let

$$X', X'', X_2, \ldots, X_n$$

be a simple random sample of size $n+1$

Use notation $\| \mathbf{Z} \| = \Sigma_{z\in\mathbf{Z}}\, z$, put

$$Y' = \| \mathbf{X}' \|, \quad Y'' = \| \mathbf{X}'' \|$$

So $Y' - Y'' = X' - X''$

Choose $\hat{X}', \hat{X}'' \propto (\hat{x}' - \hat{x}'')^2 \mathbf{I}(\{\hat{x}', \hat{x}''\} \in \mathcal{A}$

Intersection $\mathcal{R} = \{\hat{X}', \hat{X}''\} \cap \{X_2, \ldots, X_n\}$ and

$$V = \| \{X_2, \ldots, X_n\} \setminus \mathcal{R} \|; \quad T' = \| X' \cap \mathcal{R} \|$$

Let $\mathcal{S}$ be a simple random sample of size $|\mathcal{R}|$ from $\mathcal{A} \setminus \{\hat{X}', \hat{X}'', X_2, \ldots, X_n\}$ and put $\hat{T}' = \| \hat{X}' \cap \mathcal{S} \|$; similarly for $T'', \hat{T}''$

As $|\mathcal{R}| \leq 2$ we have $|\hat{T}'| \leq 2n^{-1/2}B$ and $|\hat{T}''| \leq 2n^{-1/2}B$

When third moments vanish, fourth moments exist: Order $n^{-1}$ bound

Also: Berry-Esseen bound, combinatorial central limit theorem *(Goldstein 2004)*

# Lecture 2: Other distributions

## Recap

Would like bounds on distributional distance

Use test functions to assess distance

For *standard normal distribution* $\mathcal{N}(0,1)$, distribution function $\Phi(x)$ *Stein (1972, 1986)*

1. $Z \sim \mathcal{N}(0,1)$ if and only if for all smooth functions $f$

$$\mathbf{E}f'(Z) = \mathbf{E}Zf(Z)$$

2. For any smooth function $h$ there is a smooth function $f = f_h$ solving the *Stein equation*

$$h(x) - \int h d\Phi = f'(x) - xf(x)$$

(and bounds on $f$ in terms of $h$)

3. For any random variable $W$, smooth $h$

$$\mathbf{E}h(W) - \int h d\Phi = \mathbf{E}f'(W) - \mathbf{E}Wf(W)$$

Use Taylor expansion or couplings to quantify weak dependence

If $W \geq 0, \mathbf{E}W > 0$ then $W^s$ has the $W$-*size biased distribution* if

$$\mathbf{E}Wf(W) = \mathbf{E}W\mathbf{E}f(W^s)$$

for all $f$ for which both sides exist

*Construction*(*Goldstein + Rinott 1996*)

Suppose $W = \Sigma_{i=1}^{n} X_i$ with $X_i \geq 0$, $EX_i > 0$, all $i$.

Choose index $V \propto \mathbf{E}X_v$. If $V = v$: replace $X_v$ by $X_v^s$ having the $X_v$-size biased distribution, independent, and if $X_v^s = x$: adjust $\hat{X}_u, u \neq v$, such that

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

Then $W^s = \Sigma_{u \neq V} \hat{X}_u + X_V^s$

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

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

## 3. General situation

Target distribution $\mu$

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 $\Delta f$

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 $\mathcal{A}$ 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* $\mathcal{A}f(x) = \lim_{t\downarrow 0} \frac{1}{t}(T_t f(x) - f(x))$

*Facts* (see *Ethier and Kurtz (1986)*, for example)

1. $\mu$ stationary distribution then $X \sim \mu$ if and only if

   $E\mathcal{A}f(X) = 0$ 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 $\lambda$, unit per capita death rate; stationary distribution Poisson($\lambda$)

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

## 4. Chisquare distributions

Generator for $\chi_p^2$:

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

(*Luk 1994*: *Gamma*$(r, \lambda)$ ) $\mathcal{A}$ 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

Stein equation

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

where $\chi_p^2 h$ is the expectation of $h$ under the $\chi_p^2$-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* $(\chi_p^2)$ *has a solution* $f = f_h$ *such that*

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

*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

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

$$= \; Eg'(S) - Ef'(W) + E(1-W)f'(W)$$

$$= \; Eg'(S) - ESg(S)$$

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

Put

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

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

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

where

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

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

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

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 $\beta_i = EX_1^i$

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

Similarly for $\frac{1}{2n^2}\Sigma_i EX_i^4\left|g^{(3)}(S_i+\theta\frac{X_i}{\sqrt{n}})\right|$, employ $\beta_8$

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

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

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* $\frac{1}{n}$

Also: Pearson's chisquare statistic

## 5. Discrete Gibbs measure (*R. + Eichelsbacher*)

Let $\mu$ be a probability measure with support $\mathrm{supp}(\mu) =$

$\{0, \ldots, N\}$, where $N \in \mathbf{N}_0 \cup \{\infty\}$. Write as

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

with $\mathbf{Z} = \Sigma_{k=0}^{N} \exp(V(k)) \frac{\omega^k}{k!}$, 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 $k, k+1 \in \mathrm{supp}(\mu)$

then invariant measure $\mu$

generator

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

$$+k(h(k-1) - h(k))$$

or

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

*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

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

## Bounds

Solution of Stein equation $f$ for $h$: $f(0)=0$, $f(k)=0$

for $k \notin \mathrm{supp}(\mu)$, and

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

$$\left(h(k)-\mu(h)\right).$$

**Lemma 2** *1. Put*

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

$$e^{V(k+1)-V(k)}\Big).$$

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

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

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

$$e^{V(k+1)-V(k)} \leq e^{V(k)-V(k-1)},$$

*and death rates are unit per capita. For every $j \in$*

$\mathbf{N}_0$

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

*Example: 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*

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

as in *Barbour, Holst, Janson 1992*

**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

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

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

and

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

**Lemma 3** *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 $g$*

$$\omega\, Ee^{V(X+1)-V(X)} g(X+1)$$

$$= \omega\, Ee^{V(X+1)-V(X)} Eg(X^*).$$

*For any $W \geq 0$ with $0 < EW < \infty$*

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

$$= \omega\{Ee^{V(W+1)-V(W)} g(W+1)$$

$$-Ee^{V(W+1)-V(W)} Eg(W^*)\}$$

*where $g$ 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*)

*Example:* Poisson($\lambda_1$) and Poisson($\lambda_2$) gives

$$|Eh(X) - \int h d\mu| \ \leq \ \| \ f \ \| \ |\lambda - \lambda_2|$$

## 6. Final remarks

If $X_1, X_2, \ldots, X_n$ i.i.d., $\mathbf{P}(X_i = 1) = p = 1 - \mathbf{P}(X_i = 0)$, using Stein's method we can show that

$$\sup_x |P((np(1-p))^{-1/2} \sum_{i=1}^n (X_i - p) \le x)$$

$$-P(\mathcal{N}(0,1) \le x)|$$

$$\le 6\sqrt{\frac{p(1-p)}{n}}$$

and

$$\sup_x |P(\sum_{i=1}^n X_i = x) - P(Po(np) = x)|$$

$$\le \min(np^2, p)$$

So, if $p < \frac{36}{n+36}$, the bound on the Poisson approximation

is smaller than the bound on the normal approximation

- Exchangeable pair couplings, also used for variance reduction in simulations

- Multivariate, also coupling approaches

- General distributional transformations

- Bounds in the presence of dependence

- In the i.i.d. case: Berry-Esseen inequality not quite recovered

# Further reading


1. Arratia, R., Goldstein, L., and Gordon, L. (1989). Two moments suffice for Poisson approximation: The Chen-Stein method. *Ann. Probab.* **17**, 9-25.

2. Barbour, A.D. (1990). Stein's method for diffusion approximations, *Probability Theory and Related Fields* **84** 297-322.

3. Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford University Press.

4. Barbour, A.D. and Chen. L.H.Y. eds (2005). *An Introduction to Stein's Method.* Lecture Notes Series **4**, Inst. Math. Sciences, World Scientific Press, Singapore.

5. Barbour, A.D. and Chen. L.H.Y. eds (2005). *Stein's Method and Applications.* Lecture Notes Series **5**, Inst. Math. Sciences, World Scientific Press, Singapore.

6. Diaconis, P., and Holmes, S. (eds.) (2004). *Stein's Method: Expository Lectures and Applications.* IMS Lecture Notes **46**.

7. Raič, M. (2003). Normal approximations by Stein's method. In *Proceedings of the Seventh Young Statisticians Meeting*, Mvrar, A. (ed.), Metodološki zveski, 21, Ljubljana, FDV, 71-97.

8. Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proc. Sixth Berkeley Symp. Math. Statist. Probab. **2** 583-602, Univ. California Press, Berkeley.

9. Stein, C. (1986). Approximate Computation of Expectations. IMS, Hayward, CA.