[arXiv:1207.1460]

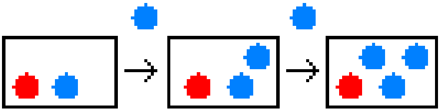
Stein's Method for the Beta distribution with Application to the Pólya Urn

Larry Goldstein
Department of Mathematics
University of Southern California
Joint work with Gesine Reinert, Oxford University

SUTD, May 15^{th}, 2013

The Urn

Pólya-Eggenberger urn at time 0 contains α red and β blue balls, and at every positive integer time a ball is chosen uniformly from the urn and replaced along with one additional ball of the same color.



Asymptotic behavior

If S_n is the number of additional red balls added to the urn by time $n = 0, 1, 2, \ldots$ then as $n \to \infty$

$$\mathcal{L}(W_n) \to_d \mathcal{B}(\alpha, \beta) \quad \text{where} \quad W_n = \frac{S_n}{n}$$

where $\alpha > 0, \beta > 0$ and $\mathcal{B}(\alpha, \beta)$ is the Beta distribution having density

$$p(x; \alpha, \beta) \quad = \quad \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \mathbf{1}_{\{x \in [0,1]\}},$$

where $B(\alpha, \beta)$ is the Beta function, $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$.

Model for

- Opinion formation
- Contagious disease
- Learning
- Signalling games (*Argiento, Pemantle, Skyrms, Volkov 2009*)

Question: How good is the Beta approximation for a given n?

Stein (1972, 1986)

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

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

If a random variable W with $\mathbf{E}W = \mu, \mathrm{Var} W = \sigma^2$ satisfies

$$\sigma^2 \mathbf{E} f'(W) - \mathbf{E}(W - \mu) f(W) \approx 0$$

for many functions f, then $W \approx Z$ in distribution.

Stein Equation

Given a test function h, let $Nh = \mathbf{E}h((Z-\mu)/\sigma)$, and solve for f in the *Stein equation*

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

Replace w by W and evaluate the expectation on the right side, involving two distributions, by taking expectation on the left hand side, involving one distribution.

Bounding the expectation will eventually require bounds on the solution f and its derivatives.

Simplest Example

Let $X, X_1, \ldots, X_n$ be i.i.d. $\mathbf{E}X = 0, \mathrm{Var}X = 1/n$ and $W = \sum_{i=1}^n X_i$, and $W_i = W - X_i = \sum_{j \neq i} X_j$. Then

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

So

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

Advantages

- Yields explicit bounds for finite n, often with computable constants.
- Can be applied in many dependent situations, often using coupling methods.
- The idea can be made to work for other distributions.

Ingredients for Stein's method

Given a target distribution μ and random variable W:

1. Characterization: An operator $\mathcal{A}$ such that $X \sim \mu$ if and only if $\mathbf{E}\mathcal{A}f(X) = 0$ for all smooth functions f.
2. Stein Equation: For h in a class of test functions find solution $f = f_h$ of
$$h(x) - \int h d\mu = \mathcal{A}f(x)$$
3. Bounds: Compute left hand side by manipulating right hand side of
$$\mathbf{E}h(W) - \int h d\mu = \mathbf{E}\mathcal{A}f(W)$$
Usually need bounds of f, f' or Δf, etc., in terms of h.

We will consider smooth test functions; techniques for nonsmooth functions, yielding Kolmogorov bounds, exist and are more complex.

Stein characterizations: the density approach

For a smooth density function $p(x)$ on (a, b), if

$$\psi(x) = \frac{p'(x)}{p(x)}$$

then Z has density p if and only if, for all functions $f \in \mathcal{F}(p)$

$$\mathbf{E}(f'(Z) + \psi(Z)f(Z)) = f(b-)p(b-) - f(a+)p(a+).$$

For $\mathcal{N}(0, 1)$ on $(-\infty, \infty)$ we have $\psi(x) = -x$, yielding the standard Stein equation.

Beta characterization

The Beta distribution, density

$$p(x;\alpha,\beta) \quad = \quad \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}\mathbf{1}_{\{x\in[0,1]\}}$$

gives, for $\{\alpha,\beta\} \subset (1,\infty)$,

$$p'(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-2}(1-x)^{\beta-2}\left\{(\alpha-1)(1-x)-(\beta-1)x\right\},$$

so that

$$\psi(x) = \frac{(\alpha-1)}{x} - \frac{(\beta-1)}{1-x}.$$

Difficulty when α or β is 1, such as for the uniform.

Transformation using fixed function

If $\mathbf{E}\mathcal{A}f(X) = 0$ for all smooth f then for a 'good' function $c(x)$ we have $\mathbf{E}\mathcal{A}c(X)g(X) = 0$ for all smooth g. Choosing $c(x) = x(1-x)$ in the Beta characterization we obtain the Stein equation

$$x(1-x)f'(x) + [\alpha(1-x) - \beta x]f(x) = h(w) - Bh,$$

which holds when α or β, or both, equal 1.
The Beta distribution is the unique stationary distribution of the Fisher Wright model in genetics that models genetic drift in a population, having generator given by the left hand side above with f' replacing f.

Solution of the Beta Stein equation

For any $\{\alpha, \beta\} \subset (0, \infty)$ and real valued function h on $[0, 1]$ such that $Bh < \infty$, the solution of the Stein equation is

$$f(x) = \frac{1}{x^\alpha (1-x)^\beta} \int_0^x u^{\alpha-1}(1-u)^{\beta-1}(h(u) - Bh)du.$$

For $\alpha, \beta \geq 1$ we can bound the solution and its first derivative in terms of the first derivative of h.

Pólya's Urn, exact distribution

The number S_n of red balls in the urn at time n with α and β the number of red and blue balls, respectively, in the urn at time 0, has distribution supported on $\{0, \ldots, n\}$ with

$$p_k = p_{k;\alpha,\beta} = \binom{n}{k} \frac{(\alpha)_k (\beta)_{n-k}}{(\alpha+\beta)_n},$$

where $(\alpha)_k = \alpha(\alpha+1)\cdots(\alpha+k-1)$. The mean is $\alpha/(\alpha+\beta)$.

Coupling Methods and Exact Distributions

Typically Stein's method employs coupling methods to bound the expectation on the right hand side of the Stein equation. Recall the simplest example presented for the normal:

Simplest Example Revisited

Let $X, X_1, \ldots, X_n$ be i.i.d. $\mathbf{E}X = 0, \mathrm{Var}X = 1/n$ and $W = \sum_{i=1}^n X_i$, and $W_i = W - X_i = \sum_{j \neq i} X_j$. Then

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

So

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

This calculation is making use of the 'Leave One Out' coupling.

Comparison of Generators, Holmes and Reinert

When we have explicit forms for both the approximating and the approximand distribution, we can use the comparison of generators approach.
Let X and X_n be characterized by $\mathcal{A}$ and $\mathcal{A}_n$, respectively,

$$\mathbf{E}\mathcal{A}f(X) = 0 \quad \text{and} \quad \mathbf{E}\mathcal{A}_n f(X_n) = 0,$$

and let f be the solution of the Stein equation for X . Then

$$\mathbf{E}h(X_n) - \mathbf{E}h(X) = \mathbf{E}\mathcal{A}f(X_n) = \mathbf{E}\left(\mathcal{A} - \mathcal{A}_n\right) f(X_n).$$

Stein discrete density characterization: Ley and Swan

Theorem: Let p be a mass function having support the integer interval $I = [a, b] \cap \mathbb{Z}$, and set

$$\mathcal{F}(p) = \{f : f(x)p(x) \text{ bounded}, f(a) = 0\}.$$

Then X has mass function p if and only if

$$\mathbf{E}(\Delta f(X-1) + \psi(X)f(X)) = 0$$

for all $f \in \mathcal{F}(p)$, where $\Delta g(x) = g(x+1) - g(x)$ and $\psi(x) = \Delta p(x)/p(x)$.

Pólya urn characterization

For the Pólya urn we have

$$\psi(k) = \frac{(n-k)(\alpha+k) - (k+1)(\beta+n-k-1)}{(k+1)(\beta+n-k-1)}$$

for $k = 0, \ldots, n-1$ and $\psi(n) = -1$. Here

$$\mathbf{E}(\Delta f(X-1) + \psi(X) f(X)) = 0,$$

not very appealing.

Transformation using fixed function

Let p be a mass function with integer interval support $I = [a, b] \cap \mathbb{Z}$ and let $c : I \to \mathbb{R}_+$.
Then $X \sim p$ if and only if for all functions $f \in \mathcal{F}(p)$

$$\mathbf{E}\left[c(X-1)\Delta f(X-1) \\ +[c(X)\psi(X) + c(X) - c(X-1)]f(X)\right] = 0.$$

Pólya urn characterization, c function

Let

$$c(k) = (k+1)(\beta + n - k - 1) \quad \text{for} \quad k = 0, \dots, n-1,$$

and $c(n) = \beta(n+1)$. Then a random variable S_n has distribution p, the number of additional red balls drawn from the Pólya's urn at time n with initial state $\alpha \geq 1$ red and $\beta \geq 1$ blue balls, if and only if for all functions $f \in \mathcal{F}(p)$

$$\begin{aligned} \mathbf{E}\, [S_n(\beta + n - S_n)\Delta f(S_n - 1) \\ + \{(n - S_n)(\alpha + S_n) - S_n(\beta + n - S_n)\} f(S_n)] = 0. \end{aligned}$$

Pólya urn characterization, c function

For $y > 0$ set $\Delta_y f(x) = f(x+y) - f(x)$ and $W_n = S_n/n$. Replacing $f(z)$ by $f(z/n)$ and dividing by n in the Polya Stein equation yields the characterization operator (expectation zero iff)

$$W_n\left(\frac{1}{n}\beta + 1 - W_n\right) n\Delta_{1/n} f\left(W_n - \frac{1}{n}\right) + \left(\alpha(1 - W_n) - \beta W_n\right) f(W_n)$$

Comparing distributions via their Stein characterizations

Polya Urn characterization

$$W_n \left(\frac{1}{n}\beta + 1 - W_n \right) n\Delta_{1/n} f \left(W_n - \frac{1}{n} \right) + (\alpha(1 - W_n) - \beta W_n) f(W_n)$$

Beta characterization

$$x(1-x)f'(x) + [\alpha(1-x) - \beta x]f(x)$$

Applying Polya urn characterization in the Beta Stein equation

$$
\begin{aligned}
&\mathbf{E}h(W_n) - \mathcal{B}h \\
&\quad = \quad \mathbf{E}\left(W_n(1-W_n)f'(W_n) + [\alpha(1-W_n) - \beta W_n]f(W_n)\right) \\
&\quad = \quad \mathbf{E}\left(W_n(1-W_n)f'(W_n)\right. \\
&\qquad\qquad \left. -W_n\left(\frac{1}{n}\beta + 1 - W_n\right) n\Delta_{1/n} f\left(W_n - \frac{1}{n}\right)\right) \\
&\quad \approx \quad \mathbf{E}\left(W_n(1-W_n)f'(W_n) - W_n\left(1-W_n\right) n\Delta_{1/n} f\left(W_n - \frac{1}{n}\right)\right).
\end{aligned}
$$

Bound in terms of derivative in f, hence in terms of the derivatives of the original test function h.

Wasserstein distance

Let X, Y be random variables, $\mathcal{L}(X) = P, \mathcal{L}(Y) = Q$ and let $\mathcal{L}$ be the set of Lipschitz-1-functions on the real line,

$$\mathcal{L} = \{f : |f(x) - f(y)| \leq |x - y|\}.$$

Define the *Wasserstein distance*

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

Also,

$$d_W(P, Q) = \int_{-\infty}^{\infty} |P(X \leq z) - P(Y \leq z)| dz$$

Wasserstein Bounds

Let $\alpha \geq 1$ and $\beta \geq 1$, let $Z \sim Beta(\alpha, \beta)$, and $W_n = S_n/n$, put

$$a_1 = (\alpha + \beta - 2)/(\alpha \wedge \beta - 1) \quad \text{and} \quad a_2 = a_1^2(2(\alpha \wedge \beta) - 1).$$

1. If neither α nor β take the value 1,
$$d_W(W_n, Z) \leq \frac{1}{n}(1 + 2a_1(\alpha + \beta) + 3(a_1 + 2a_2)(1 + \alpha \vee \beta)).$$
2. If $\alpha = 1$ and $\beta > 1$ then
$$d_W(W_n, Z) \leq \frac{1}{n}\left(5 + 14\beta + 8\beta^2\right),$$
and if $\alpha > 1$ and $\beta = 1$ then replace β by α.
3. If $\alpha = \beta = 1$,
$$d_W(W_n, Z) \leq \frac{27}{n}.$$

Order cannot be improved

Take $h(x) = x(1-x)$ for $x \in [0,1]$, then

$$\begin{aligned} \mathbf{E}(h(W_n)) - \mathbf{E}h(W) &= \left(\frac{[n]_2}{n^2} - 1\right) \mathbf{E}\left(W(1-W)\right) \\ &= -\frac{1}{n}\frac{\alpha\beta}{(\alpha+\beta)_2}. \end{aligned}$$

Thus, for all $\alpha \geq 1, \beta \geq 1$,

$$d_W(W_n, Z) \geq \frac{1}{n}\frac{\alpha\beta}{(\alpha+\beta)_2}.$$

Arcsine law, $\mathcal{B}(1/2, 1/2)$

Asymptotic distribution of the last return time L_{2n} to zero

$$L_{2n} = \sup\{m \leq 2n : T_m = 0\}$$

of a simple symmetric random walk $T_n = X_1 + \cdots X_n$, where $X_1, \ldots, X_n$ are independent symmetric Bernoulli.

The number U_{2n} of segments of the walk that lie above the x axis, and R_{2n}, the first time the walk visits the terminal point S_{2n}, are both equal in distribution to L_{2n}.

Stein and the Arcsine law

Finding the characterizing equation for L_{2n} and applying the methods above yield the following result (see also Dobler):

Theorem

Let L_{2n} be the last return time to zero of a simple symmetric random walk of length of length $2n$ and let Z have the Arcsine distribution. Then

$$d_W\left(\frac{L_{2n}}{2n}, Z\right) \leq \frac{27}{2n} + \frac{8}{n^2}.$$

The same bound holds with L_{2n} replaced by U_{2n} or R_{2n}. The $O(1/n)$ rate of the bound cannot be improved.

Remarks

- The idea of comparing distributions via Stein characterizations is due to *Holmes 2004, Eichelsbacher and Reinert 2008*.
- *Drinane (2008)* gives an $O(n^{-1/2})$ bound in Kolmogorov distance.
- Stein's method has been applied to many other distributions, and also in multivariate settings.