On some examples where Stein's method could be improved

Bero Roos

University of Leicester, Department of Mathematics

Thu 15 Jan 2009

Progress in Stein's Method
(Singapore, 5 Jan – 6 Feb 2009)

Content

Poisson approximation of mixed Poisson distributions

Poisson approximation of multivariate Poisson mixtures

On the distance between random sums

Compound Poisson approximation for convolutions

Poisson approximation of mixed Poisson distributions

Let $\mathrm{MPo}(X)$ be a mixed Poisson distribution with structure random variable $X \geq 0$, i.e.

$$\mathrm{MPo}(X)\{m\} = \mathbb{E}\Big[\mathrm{Pois}(X)\{m\}\Big], \quad (m \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}).$$

Let $\mu = \mathbb{E}(X) > 0$, $\sigma^2 = \mathrm{Var}(X)$.

Theorem 1. (BHJ 1992, p. 12 & 68; Stein's method).
If the respective moments are finite, then, for any $\theta \geq \mathrm{e}$,

$$d_{\mathrm{TV}}(\mathrm{MPo}(X), \mathrm{Pois}(\mu)) \leq \min\left\{\frac{\sigma^2}{\mu}, \sigma^2\right\},$$

$$d_{\mathrm{TV}}(\mathrm{MPo}(X), \mathrm{Pois}(\mu)) \geq \frac{1}{2(2\mathrm{e}^{-3/2} + \theta\mathrm{e}^{-1})}\Big[\frac{\sigma^2}{\mu}(1 - 3\theta^{-1}) - \frac{1}{\theta\mu^2}\Big(7\sigma^2 + 6\mathbb{E}(X-\mu)^3 + \mathbb{E}(X-\mu)^4\Big)\Big].$$

A better upper bound:

Theorem 2. (R. 2003, JSPI). If the respective moments are finite then
$$d_{\mathrm{TV}}(\mathrm{MPo}(X), \mathrm{Pois}(\mu)) \le \min\left\{\frac{3}{2e}\,\mathbb{E}\Big(X \ln \frac{X}{\mu}\Big),\, \sigma^2\right\}.$$

Remarks:

- The constant $\frac{3}{2\mathrm{e}} \approx 0.55$ is optimal.
- For all $\varepsilon > 0$,
$$\mathbb{E}\Big(X \ln \frac{X}{\mu}\Big) \le \frac{1}{\varepsilon\,\mu^{\varepsilon}}\Big(\mathbb{E}(X^{1+\varepsilon}) - \mu^{1+\varepsilon}\Big).$$
For $\varepsilon = 1$, we obtain $\mathbb{E}(X \ln \frac{X}{\mu}) \le \frac{\sigma^2}{\mu}$.

Also available:

- Lower bounds and asymptotic relations, e.g. $d_{\mathrm{TV}}(\dots) \sim \frac{\sigma^2}{\mu\sqrt{2\pi\mathrm{e}}}$ if $\mu \to \infty$ and $\frac{\sqrt{\mu}}{\sigma^2}\mathbb{E}[\frac{|X-\mu|^3}{(\sqrt{X}+\sqrt{\mu})^2}] \to 0$.
- Approximations by finite signed measures of higher order.
- Bounds for other distances.
- Biased approximations.

Idea of Proof of Theorem 2

Step 1: Using a Taylor formula for the counting densities of the distributions, we obtain

$$\mathrm{MPo}(X) - \mathrm{Pois}(\mu) \; = \; \mathbb{E}\Big[\int_{\mu}^{X} (X - t)(I_1 - I_0)^2 \mathrm{Pois}(t)\,\mathrm{d}t\Big],$$

where

- I_x is the Dirac measure at point $x \in \mathbb{R}$,
- integrals of measures are understood setwise,
- products and powers of signed measures are understood in the convolution sense, ($W^0 = I_0$ for any signed measure W),
- $\int_{\mu}^{X} = -\int_{X}^{\mu}$.

Step 2: Let $\|\cdot\|$ denote the total variation norm of finite signed measures. Then

$$\begin{aligned} d_{\mathrm{TV}}(\mathrm{MPo}(X), \mathrm{Pois}(\mu)) &= \frac{1}{2}\|\mathrm{MPo}(X) - \mathrm{Pois}(\mu)\| \\ &= \frac{1}{2}\Big\|\mathbb{E}\Big[\int_{\mu}^{X} (X-t)(I_1 - I_0)^2 \mathrm{Pois}(t)\,\mathrm{d}t\Big]\Big\| \\ &\le \frac{1}{2}\mathbb{E}\Big[\int_{\mu\wedge X}^{\mu\vee X} |X-t|\Big\|(I_1 - I_0)^2 \mathrm{Pois}(t)\Big\|\,\mathrm{d}t\Big]. \end{aligned}$$

The proof is completed using

$$\Big\|(I_1 - I_0)^2 \mathrm{Pois}(t)\Big\| \le \min\Big\{\frac{3}{t\mathrm{e}}, 4\Big\}, \qquad (t \in (0,\infty)),$$

(R. 2001, SPL, proof by direct computations).

Here the constant $\frac{3}{\mathrm{e}}$ is the best possible ("=" holds for $t = 1$). □

How to get the better order using Stein's method

Let $g := g_{t,A}(\cdot)$ be the solution of the Stein-Chen equation

$$\mathbf{1}_A(j) - \mathrm{Pois}(t)(A) = tg(j+1) - jg(j), \qquad j \in \mathbb{Z}_+$$

($t \in (0,\infty)$, $A \subseteq \mathbb{Z}_+$). We know that (BHJ, 1992, p. 7)

$$\begin{aligned} \|g\| &:= \sup_{j\in\mathbb{N}} |g(j)| \leq \min\left\{1, \sqrt{\frac{2}{\mathrm{e}t}}\right\}, \\ \|\Delta g\| &:= \sup_{j\in\mathbb{N}} |g(j+1) - g(j)| \leq \frac{1-\mathrm{e}^{-t}}{t} \leq \min\left\{1, \frac{1}{t}\right\}. \end{aligned}$$

Let $(N_t \mid t \in [0,\infty))$ be a homogeneous Poisson process with intensity 1 independent of X. A simple proof shows that

Lemma. For $t \in (0,\infty)$,

$$\begin{aligned} (I_1 - I_0)\mathrm{Pois}(t) &= -\mathbb{E}\Big[g_{t,\bullet}(N_t+1)\Big], \\ (I_1 - I_0)^2\mathrm{Pois}(t) &= 2\mathbb{E}\Big[g_{t,\bullet}(N_t+1) - g_{t,\bullet}(N_t+2)\Big]. \end{aligned}$$

Corollary.

(a) For $t \in (0,\infty)$,

$$\begin{aligned} \left\|(I_1 - I_0)\mathrm{Pois}(t)\right\| &\leq 2 \sup_{A\subseteq\mathbb{Z}_+} \|g_{t,A}\|, \\ \left\|(I_1 - I_0)^2\mathrm{Pois}(t)\right\| &\leq 4 \sup_{A\subseteq\mathbb{Z}_+} \|\Delta g_{t,A}\|. \end{aligned}$$

(b) For $A \subseteq \mathbb{Z}_+$,

$$\begin{aligned} &\mathrm{MPo}(X)(A) - \mathrm{Pois}(\mu)(A) \\ &= 2\mathbb{E}\Big[\int_\mu^X (X - t)\Big(g_{t,A}(N_t + 1) - g_{t,A}(N_t + 2)\Big)\,\mathrm{d}t\Big]. \end{aligned}$$

Theorem 3.

$$d_{\mathrm{TV}}(\mathrm{MPo}(X), \mathrm{Pois}(\mu)) \leq \min\left\{2\mathbb{E}\left(X \ln \frac{X}{\mu}\right), \sigma^2\right\}.$$

Open question:

- Can the constant 2 be dropped using Stein's method?

Poisson approximation of multivariate Poisson mixtures

For a random vector $X = (X_1, \ldots, X_k) \in [0, \infty)^k$, ($k \in \mathbb{N}$) let

$$\mathrm{MPo}(X)\{m\} = \mathbb{E}\Big[\mathrm{Pois}(X)\{m\}\Big], \qquad (m = (m_1, \ldots, m_k) \in \mathbb{Z}_+^k),$$

where $\mathrm{Pois}(t)\{m\} = \prod_{r=1}^k \mathrm{Pois}(t_r)\{m_r\}$, ($t \in [0, \infty)^k$).
For $r \in \{1, \ldots, k\}$, let $\mu_r = \mathbb{E}(X_r) > 0$ and $\sigma_r^2 = \mathrm{Var}(X_r)$, $\mu = (\mu_1, \ldots, \mu_k)$.

Theorem 4. (R. 2003, JAP). If the respective moments are finite then

$$d_{\mathrm{TV}}(\mathrm{MPo}(X), \mathrm{Pois}(\mu)) \leq \min\Big\{\frac{1}{\sqrt{2}} \sum_{r=1}^k \mathbb{E}\Big(X_r \ln \frac{X_r}{\mu_r}\Big), \Big(\sum_{r=1}^k \sigma_r\Big)^2\Big\}.$$

Tool: If e_r denotes the rth standard unit vektor in $\mathbb{R}^k$, then

$$\mathrm{MPo}(X) - \mathrm{Pois}(\mu) = \mathbb{E}\Big[\int_0^1 (1-t)\Big(\sum_{r=1}^k (X_r - \mu_r)(I_{e_r} - I_0)\Big)^2 * \mathrm{Pois}\Big(\Big(tX_r + (1-t)\mu_r\Big)_{r=1,\ldots,k}\Big)\,\mathrm{d}t\Big]$$

Open question: Can Stein's method be used to prove Thm. 4?

On the distance between random sums

Notation:

- Let $X_1, X_2, \ldots$ be iid random variables in $\mathbb{R}$, $S_n = \sum_{j=1}^{n} X_j$,
- M, N be random variables in $\mathbb{Z}_+$ independent of the X_j.

Problem: Give a bound for $d_{\mathrm{TV}}(S_M, S_N)$.

The bound should reflect the following simple fact:

Lemma. Let $p = 1 - q = \mathbb{P}(X_1 \neq 0)$. Then

$$d_{\mathrm{TV}}(S_M, S_N) = 0 \quad \Leftrightarrow \quad [\, M \overset{\mathrm{d}}{=} N \quad \text{or} \quad p = 0 \,].$$

In what follows we consider the case $\mathbb{E}[M] = \mathbb{E}[N] < \infty$.

Theorem 5. (R. & Pfeifer 2003, JAP).
Let $a_n = \mathbb{P}(M = n) - \mathbb{P}(N = n)$, $(n \in \mathbb{Z}_+)$. Then, for all $m \in \mathbb{Z}_+$,

$$d_{\mathrm{TV}}(S_M, S_N) \leq p^2 \sum_{n=0}^{\infty} (n - m)(n - m - 1)|a_n|,$$

$$d_{\mathrm{TV}}(S_M, S_N) \leq \frac{p}{q\sqrt{2}} \sum_{n=0}^{\infty} \left[m - n + n \log\left(\frac{n + 1/2}{m + 1/2}\right)\right] |a_n|,$$

Let Z be a random variable with

$$\mathbb{P}(Z = n) = \frac{|a_n|}{\sum_{k=0}^{\infty} |a_n|}, \qquad \mu = \mathbb{E}Z < \infty.$$

Corollary. Letting $m = \lfloor \mu \rfloor$, we obtain

$$d_{\mathrm{TV}}(S_M, S_N) \leq p^2 \sum_{n=0}^{\infty} (n - \lfloor \mu \rfloor)(n - \lfloor \mu \rfloor - 1)|a_n|,$$

$$d_{\mathrm{TV}}(S_M, S_N) \leq \frac{p}{q\sqrt{2}(\lfloor \mu \rfloor + \frac{1}{2})} \sum_{n=0}^{\infty} (n - \lfloor \mu \rfloor)^2 |a_n|.$$

Also available:

- Approximations by finite signed measures of higher order.
- (Somewhat complicated) bounds with sharp constants.
- Bounds with $|a_n|$ replaced by $|\mathbb{P}(M \leq n) - \mathbb{P}(N \leq n)|$.
- Better bounds when $M \leq_{\mathrm{ST}} N$ or $M \leq_{\mathrm{SL}} N$.

Example:

(a) Assume that
- $X_j \sim \mathrm{Bi}(1, p)$, $(j \in \mathbb{N})$,
- $M = m \in \mathbb{Z}_+$ is a constant, and
- N is an arbitrary random variable with $\mathbb{E}(N) = m$.

Then

$$d_{\mathrm{TV}}(S_M, S_N) \leq \mathrm{Var}(N)\, p^2 \min\left\{\frac{1}{\sqrt{2}\, pq(m + 1/2)}, 1\right\}.$$

(b) If additionally $N \sim \mathrm{Pois}(m)$, then $\mathrm{Var}(N) = m$.
If $p < C < 1$, the bound has the right order, since, here,

$$\frac{1}{32}\, mp^2 \min\left\{\frac{1}{mp}, 1\right\} \leq d_{\mathrm{TV}}(S_M, S_N) \leq mp^2 \min\left\{\frac{1}{mp}, 1\right\},$$

(Barbour & Hall 1984).

Tool in the proof of Theorem 5

Let $Q = \mathbb{P}(X_1 \in \cdot \,|\, X_1 \neq 0)$, i.e. $\mathcal{L}(X_1) = qI_0 + pQ$.
Then

$$\mathcal{L}(S_M) - \mathcal{L}(S_N) = p^2 \sum_{j=0}^{\infty} g_{j,m}(Q - I_0)^2 (I_0 + p(Q - I_0))^j,$$

where

$$g_{j,m} = \begin{cases} -\sum_{n=0}^{j}(n-j-1)a_n & , \quad \text{if } 0 \leq j < m, \\ \sum_{n=j+1}^{\infty}(n-j-1)a_n & , \quad \text{if } m \leq j. \end{cases}$$

Conjecture: Stein's method can be used here.

Useful smoothness bounds in this context

Remark:

- We used that (R. 2000, TPA)

$$\left\|(I_1 - I_0)^2 (I_0 + p(I_1 - I_0))^j\right\| \leq \left[\binom{j+2}{2}(pq)^2\right]^{-1/2}.$$

  See R. (Bernoulli, accepted) for a generalization.
- The following bound can also be used here (Mattner & R. 2007, PTRF, Cor. 1.6)

$$\left\|(I_1 - I_0)\prod_{k=1}^{n} F_k\right\| \leq 2\left[\frac{\frac{2}{\pi}}{\frac{1}{4} + \sum_{k=1}^{n}[1 - d_{\mathrm{TV}}(F_k,\, F_k I_1)]}\right]^{1/2},$$

  where $F_1, \ldots, F_n$ are probability distributions on the line. (This improves Barbour & Xia (1999, Prop. 4.6, Mineka coupling).)

Compound Poisson approximation for convolutions

Notation:

- Let $X_j \in \mathbb{Z}_+$, $(j = 1, \ldots, n)$ be independent but not necessarily identically distributed random variables,
- $S_n := \sum_{j=1}^{n} X_j$,
- $p_j := \mathbb{P}(X_j \neq 0)$, $Q_j(B) := \mathbb{P}(X_j \in B \mid X_j \neq 0)$.
- Then $\mathcal{L}(S_n) = \prod_{j=1}^{n} \Big((1 - p_j) I_0 + p_j Q_j\Big)$.

Problem: Approximate $\mathcal{L}(S_n)$ by a compound Poisson distribution

$$\mathrm{CPo}(\lambda,\, Q) = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} Q^i = \exp(\lambda(Q - I_0))$$

with

$$\lambda = \sum_{j=1}^{n} p_j, \qquad Q = \sum_{j=1}^{n} \frac{p_j}{\lambda} Q_j.$$

Theorem 6. (Barbour, Chen & Loh 1992; Stein's Method).

(a) The following inequality holds:

$$d_{\mathrm{TV}}(\mathcal{L}(S_n), \mathrm{CPo}(\lambda, Q)) \le H(\lambda,\, Q) \sum_{j=1}^{n} p_j^2\, \mu_j^2,$$

where $\mu_j := \sum_{r=1}^{\infty} r Q_j\{r\} \ge 1$ and $H(\lambda,\, Q)$ is a quantity, which was derived with the help of Stein's method.

(b) If the $q_r := \sum_{j=1}^{n} \frac{p_j}{\lambda} Q_j\{r\}$ satisfy $r q_r \downarrow 0$, then $a := q_1 - 2q_2 \ge 0$ and

$$H(\lambda,\, Q) \le \min\left\{1,\, \frac{1}{\lambda a}\left(\frac{1}{4\lambda a} + \log^+(2\lambda a)\right)\right\}.$$

Theorem 7. (Barbour & Xia 1999; Stein's Method).

Let $\mu = \sum_{r=1}^{\infty} r q_r$. If $\vartheta = \frac{1}{\mu}\sum_{r=2}^{\infty} r(r-1) q_r < \frac{1}{2}$, then

$$H(\lambda,\, Q) \le \frac{1}{(1-2\vartheta)\lambda\mu}.$$

Remark:

- Stein's method was also used by Barbour & Utev (1998, 1999) and Barbour & Xia (2000).
- A very deep bound is due to Zaitsev (1983):

$$d_{\mathrm{KM}}(\mathcal{L}(S_n), \mathrm{CPo}(\lambda, Q)) \le C \max_{1\le j\le n} p_j.$$

(Kolmogorov distance).

Theorem 8. (R. 2003, AP).

Let μ_j and q_r be defined as before. Set $\nu_j = \sum_{r\in\mathbb{N}:q_r>0} \frac{[Q_j\{r\}]^2}{q_r}$.

If $rq_r \downarrow 0$, then

$$d_{\mathrm{TV}}(\mathcal{L}(S_n), \mathrm{CPo}(\lambda, Q)) \le 9.2 \sum_{j=1}^{n} p_j^2 \min\left\{\frac{\mu_j^2}{\lambda}, \frac{\nu_j}{\lambda}, 1\right\}.$$

In general, we have:

$$d_{\mathrm{TV}}(\mathcal{L}(S_n), \mathrm{CPo}(\lambda, Q)) \le 8.8 \sum_{j=1}^{n} p_j^2 \min\left\{\frac{\mu_j^2}{q_1\lambda}, \frac{\nu_j}{\lambda}, 1\right\}.$$

Remark: The q_1 (or $a = q_1 - 2q_2$, BCL 1992) in the denominators can be improved by using a better smoothness bound (Mineka coupling).

Key lemma. We have

$$d_{\mathrm{TV}}\Big(\mathcal{L}(S_n), \mathrm{CPo}(\lambda, Q)\Big) \leq \frac{1}{2}\sum_{j=1}^{n}\frac{1}{j!}\Big(\sum_{i=1}^{n}\|W_{i,j}\|\Big)^{j},$$

for finite signed measures $W_{i,j} = V_i \exp\Big(\frac{\lambda}{j}(Q - I_0)\Big)$,

$$V_i = \Big((I_0 + R_i)\mathrm{e}^{-R_i} - I_0\Big), \qquad R_i := p_i(Q_i - I_0).$$

Remarks:

- No exponential series.
- $W_{i,j}$ contains the convolution factor $p_i^2(Q_i - I_0)^2 \exp(\frac{\lambda}{j}(Q - I_0))$.
- Theorem 7 is shown by using the key lemma and suitable smoothness bounds for $\|W_{i,j}\|$.

Open question: Can Stein's method be adapted to prove Thm. 8?

Proof of the key lemma (modified Kerstan expansion)

We have

$$
\begin{aligned}
\mathcal{L}(S_n) - \exp(\lambda(Q - I_0)) &= \Big(\prod_{i=1}^{n}(V_i + I_0) - I_0\Big)\exp(\lambda(Q - I_0)) \\
&= \Big(\sum_{\substack{J\subseteq\{1,\dots,n\}:\\ \#J\geq 1}} \prod_{i\in J} V_i\Big)\exp(\lambda(Q - I_0)) \\
&= \sum_{j=1}^{n} \sum_{\substack{J\subseteq\{1,\dots,n\}:\\ \#J=j}} \prod_{i\in J} W_{i,j}.
\end{aligned}
$$

Using the properties of the total variation norm and the polynomial theorem, we obtain

$$
\begin{aligned}
d_{\mathrm{TV}}\Big(\mathcal{L}(S_n), \mathrm{CPo}(\lambda,\ Q)\Big) &\leq \frac{1}{2}\sum_{j=1}^{n} \sum_{\substack{J\subseteq\{1,\dots,n\}:\\ \#J=j}} \prod_{i\in J} \|W_{i,j}\| \\
&\leq \frac{1}{2}\sum_{j=1}^{n} \frac{1}{j!}\Big(\sum_{i=1}^{n}\|W_{i,j}\|\Big)^{j}. \qquad \square
\end{aligned}
$$

Remark

- If $Q = \sum_{r=1}^{\infty} q_r I_r$ with $rq_r \geq (r+1)q_{r+1}$, for all $r \in \mathbb{N}$, then for all $\eta \in (0,1)$ there exists a distribution R_η on $\mathbb{Z}_+$ such that

$$\mathrm{CPo}(t, Q) = \mathrm{CPo}(t, Q(\eta))R_\eta, \qquad Q(\eta) = \sum_{r=1}^{\infty} q_r(I_0 + \eta(I_1 - I_0))^r,$$

  i.e. $\mathrm{CPo}(t, Q)$ is discrete self-decomposable (Steutel & van Harn 1979).

- Two different smoothness bounds can be found in R. (2003, La. 4 & 6). In particular

$$\|(I_1 - I_0)^j \mathrm{CPo}(t, Q)\| \leq 2^j \mathbb{E}\binom{Y + j}{j}^{-1/2}, \quad (j \in \mathbb{Z}_+),$$

  where $Y \sim \mathrm{CPo}(t, Q)$, which implies the weaker bound

$$\|(I_1 - I_0)^2 \mathrm{CPo}(t, Q)\| \leq 4\sqrt{2}\frac{1 - \mathrm{e}^{-t}}{t}.$$

Thank you very much for your attention!