Plan:
1. Stein's method.
2. The Stein equation for the Poisson distribution. Properties of its solutions.
3. Construction of error bounds. Local and coupling approaches.
5. Poisson approximation for the number of points of a point process. Connections to Poisson-Charlier expansions.
6. Poisson-Charlier expansions.

Stein’s method for Poisson and compound Poisson approximations: I

Torkel Erhardsson
Department of Mathematics
KTH
The problem:

Let \((S, \mathcal{S}, \mu)\) be a probability space, let \(\chi\) be the set of measurable functions \(h : S \to \mathbb{R}\), and let \(\chi_0 \subset \chi\) be such that all \(h \in \chi_0\) are \(\mu\)-integrable. We want to compute \(\int_S h d\mu\) for all \(h \in \chi_0\).

Stein’s method [Stein 1972, 1986].

Choose a probability measure \(\mu_0\) such that all \(h \in \chi_0\) are \(\mu_0\)-integrable, and all \(\int_S h d\mu_0\) are easily computed. Find a set of functions \(\mathcal{F}_0\) and a mapping \(T_0 : \mathcal{F}_0 \to \chi\), such that, for each \(h \in \chi_0\), the equation

\[
T_0f = h - \int_S h d\mu_0 \quad (1)
\]

has a solution \(f \in \mathcal{F}_0\). Then,

\[
|\int_S h d\mu - \int_S h d\mu_0| \leq \int_S |(T_0f)| d\mu.
\]

(1) is called a Stein equation.

\(T_0\) is called a Stein operator.

\(f\) is called a Stein transform.

Exchangeable pairs and antisymmetric functions.

To construct a Stein operator \(T_0\) for \(\mu_0\), Stein proposes the following procedure:

1) Choose an exchangeable pair of random variables \((X,Y)\) with marginal distribution \(\mu_0\).
2) Choose a mapping \(\alpha : \mathcal{F}_0 \to \mathcal{F}\), where \(\mathcal{F}\) is the space of measurable antisymmetric functions \(F : S^2 \to \mathbb{R}\) such that

\[
\mathbb{E}(|F(X,Y)|) < \infty.
\]

3) Take \(T_0 = T \circ \alpha\), where \(T : \mathcal{F} \to \chi\) is defined by

\[
(TF)(x) = \mathbb{E}(F(X,Y)|X = x).
\]

This procedure is not guaranteed to give us a Stein operator with good properties, but experience shows that it often will.
The Stein equation for Po(λ) [Chen 1975].

Here \( (S, \mathcal{F}, \mu) = (\mathbb{Z}_+, \mathcal{B}_{\mathbb{Z}_+}, \mu) \), where \( \mathcal{B}_{\mathbb{Z}_+} \)
is the power \( \sigma \)-algebra, and \( \mu_0 = \text{Po}(\lambda) \).
Take \( \mathcal{F}_0 = \chi \), and define \( T_0 : \chi \to \chi \) by
\[
(T_0 f)(k) = \lambda f(k + 1) - k f(k) \quad \forall k \in \mathbb{Z}_+.
\]

**Theorem.** The Stein equation
\[
T_0 f = h - \int_{\mathbb{Z}_+} h d\mu_0
\]
has a unique solution for each \( \mu_0 \)-integrable \( h \) (except for \( f(0) \), which can be chosen arbitrarily). \( f \) can be computed recursively from the Stein equation, and is explicitly given by
\[
f(k) = \frac{(k - 1)!}{\lambda^k} \sum_{i=0}^{k-1} \left( h(i) - \int_{\mathbb{Z}_+} h d\mu_0 \right) \frac{\lambda^i}{i!}
\]
\[
= -\frac{(k - 1)!}{\lambda^k} \sum_{i=k}^{\infty} \left( h(i) - \int_{\mathbb{Z}_+} h d\mu_0 \right) \frac{\lambda^i}{i!}.
\]
Also: if \( h \) is bounded, then \( f \) is bounded.

A characterization of \( \text{Po}(\lambda) \).

**Corollary.** A probability measure \( \mu \) on \( \mathbb{Z}_+ \)
is \( \mu_0 = \text{Po}(\lambda) \) if and only if
\[
\int_{\mathbb{Z}_+} (T_0 f) d\mu = 0
\]
for all bounded \( f : \mathbb{Z}_+ \to \mathbb{R} \).

**Proof:** Necessity: straightforward calculations. Sufficiency: use the fact that the Stein equation has a bounded solution for each bounded \( h : \mathbb{Z}_+ \to \mathbb{R} \).
Preliminary bounds [Barbour and Eagleson 1983].

**Theorem.** Let $f_A$ be the solution of the Stein equation with $h = I_A$ (where $A \subset \mathbb{Z}_+$). Then,

$$\sup_{A \subset \mathbb{Z}_+} \sup_{k \in \mathbb{Z}_+} |f_A(k+1) - f_A(k)| \leq \frac{1 - e^{-\lambda}}{\lambda}.$$

$$\sup_{A \subset \mathbb{Z}_+} \sup_{k \in \mathbb{Z}_+} |f_A(k)| \leq \min(1, \frac{1}{\sqrt{\lambda}}).$$

**Proof of the first bound:** First consider the case $A = \{k\}$. We see from the explicit expressions that $f_{\{k\}}(\cdot)$ is negative and decreasing for $1 \leq i < k$, and positive and decreasing for $i \geq k + 1$. Hence, the only positive value taken by $f_{\{k\}}(i+1) - f_{\{k\}}(i)$ is

$$f_{\{k\}}(k+1) - f_{\{k\}}(k) = \frac{e^{-\lambda}}{\lambda} \left( \sum_{r=k+1}^{\infty} \frac{\lambda^r}{r!} \right) \leq \frac{1 - e^{-\lambda}}{\lambda}.$$

The general case follows from the facts that $f_A = \sum_{k \in A} f_{\{k\}}$ and $f_{Ac} = -f_A$. \qed

The generator interpretation [Barbour 1988].

Let $\{Z_t; t \in \mathbb{R}_+\}$ be a stationary birth-and-death process on $\mathbb{Z}_+$, with constant birth intensity $\lambda$ and death intensities $\mu_k = k$. This process is reversible with stationary distribution $\mu_0 = \text{Po}(\lambda)$. Hence, $(Z_0, Z_t)$ is an exchangeable pair with marginal distribution $\mu_0$.

Choose the mapping $\alpha : \chi \to \mathcal{F}$ defined (at least for functions $g$ which do not grow too fast) by

$$(\alpha g)(x,y) = g(y) - g(x),$$

and the mapping $T : \mathcal{F} \to \chi$ defined by

$$(TF)(k) = \mathbb{E}(F(Z_0, Z_t) | Z_0 = k).$$
Then,
\[ \lim_{t \downarrow 0} (T \circ \alpha g)(k) = \lim_{t \downarrow 0} \frac{\mathbb{E}(g(Z_t)|Z_0 = k) - g(k)}{t} \]
\[ = \lambda g(k + 1) + kg(k - 1) - (\lambda + k)g(k) \]
\[ = (\mathcal{A} g)(k) = (T_0 f)(k), \]
where \( f(k) = \nabla g(k) = g(k) - g(k - 1) \), and \( \mathcal{A} \) is the generator of \( \{Z_t; t \in \mathbb{R}_+\} \).

The corresponding Poisson’s equation is
\[ -\mathcal{A} g = h - \int_{\mathbb{Z}_+} h d\mu_0. \]
If \( h \) is bounded, then this equation has the solution
\[ g(k) = \int_0^\infty (\mathbb{E}(h(Z_t)|Z_0 = k) - \int_{\mathbb{Z}_+} h d\mu_0) dt, \]
and \( f = \nabla g \) is the unique bounded solution of the Stein equation.

Construction of explicit error bounds for Poisson approximation: sums of indicators.

**Notation:** \( \Gamma = \{1, \ldots, n\}; \{X_i; i \in \Gamma\} \) are indicator variables; \( p_i = \mathbb{E}(X_i); W = \sum_{i \in \Gamma} X_i; \lambda = \mathbb{E}(W); \mu = \mathcal{L}(W); \) and \( \mu_0 = \text{Po}(\lambda) \).

**Goal:** to find a bound for
\[ d_{TV}(\mu, \mu_0) = \sup_{A \subset \mathbb{Z}_+} |\mu(A) - \mu_0(A)|. \]

The Stein equation gives, for each \( A \subset \mathbb{Z}_+ \),
\[ \mu(A) - \mu_0(A) = \int_{\mathbb{Z}_+} (T_0 f_A) d\mu \]
\[ = \mathbb{E}(\lambda f_A(W + 1) - W f_A(W)), \]
where \( f_A \) is the solution of the Stein equation with \( h = I_A \).
The “local” approach [Chen 1975].

**Theorem.** For each $i \in \Gamma$, divide $\Gamma \setminus \{i\}$ into two subsets $\Gamma_i^s$ and $\Gamma_i^w$, so that, informally,

$\Gamma_i^s = \{j \in \Gamma \setminus \{i\}; X_j \text{ “strongly” dependent on } X_i \}.$

Let $Z_i = \sum_{j \in \Gamma_i^s} X_j$ and $W_i = \sum_{j \in \Gamma_i^w} X_j$.

Then,

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} \left( p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i) \right)$$

$$+ \min(1, \frac{1}{\sqrt{\lambda}}) \sum_{i \in \Gamma} \mathbb{E}\left| \mathbb{E}(X_i | W_i) - p_i \right|.$$

**Proof:**

$$\mathbb{E}(\lambda f_A(W + 1) - Wf_A(W))$$

$$= \sum_{i \in \Gamma} \mathbb{E}(p_i f_A(W + 1) - X_i f_A(W))$$

$$= \sum_{i \in \Gamma} \mathbb{E}\left( p_i f_A(W + 1) - p_i f_A(W_i + 1) ight.$$

$$+ p_i f_A(W_i + 1) - X_i f_A(W_i + 1)$$

$$+ X_i f_A(W_i + 1) - X_i f_A(W) \right)$$

$$\leq \sup_{k \in \mathbb{Z}_+} |f_A(k+1) - f_A(k)| \sum_{i \in \Gamma} \left( p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i) \right)$$

$$+ \sup_{k \in \mathbb{Z}_+} |f_A(k)| \sum_{i \in \Gamma} \mathbb{E}\left| \mathbb{E}(X_i | W_i) - p_i \right|.$$
Example. Let \( \{X_i; i \in \Gamma\} \) be independent. Choosing \( \Gamma_i^s = \emptyset \) gives:

\[
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} p_i^2.
\]

Le Cam (1960) showed using Fourier transforms that

\[
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq 4.5 \max_{i \in \Gamma} p_i,
\]

and that, if \( \max_{i \in \Gamma} p_i \leq \frac{1}{4} \), then

\[
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \min(1, \frac{8}{\lambda}) \sum_{i \in \Gamma} p_i^2.
\]

The constant 8 was improved to 1.05 by Kerstan (1964), and to 0.71 by Daley and Vere-Jones (1988).

Barbour and Hall (1984) showed the lower bound:

\[
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \geq \frac{1}{32} \min(1, \frac{1}{\lambda}) \sum_{i \in \Gamma} p_i^2.
\]

Example. (Classical birthday problem.) [Arratia, Goldstein and Gordon 1989]. \( n \) balls (people) are thrown independently into \( d \) equiprobable boxes (days of the year). Let \( W \) be the number of pairs of balls that go into the same box. Then,

\[
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{8 \lambda(1 - e^{-\lambda})}{n - 1},
\]

where \( \lambda = \mathbb{E}(W) = \binom{n}{2} d^{-1} \).

Proof: Let \( \Gamma = \{i \subset \{1, \ldots, n\}; |i| = 2\} \). Let \( X_i \), where \( i = \{i_1, i_2\} \), be the indicator for the event “the balls \( i_1 \) and \( i_2 \) go into the same box”. Clearly \( W = \sum_{i \in \Gamma} X_i \). \{\( X_i; i \in \Gamma\} \) are dissociated, meaning that \( X_i \) and \( X_j \) are independent if \( i \cap j = \emptyset \). Choosing \( \Gamma_i^s = \{j \in \Gamma \setminus \{i\}; i \cap j \neq \emptyset\} \), the last term in the bound vanishes.
Since also $\mathbb{E}(X_i) = d^{-1}$ for all $i \in \Gamma$, and $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)^2 = d^{-2}$ for all $i \neq j$,

\[
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} (p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i))
\]

\[
= \frac{1 - e^{-\lambda}}{\lambda} \left( \binom{n}{2} \left( \frac{2(n-1)+1}{d^2} + \frac{2(n-1)}{d^2} \right) \right)
\]

\[
= \frac{1 - e^{-\lambda}}{\lambda} \left( \binom{n}{2} \frac{1}{d^2} (4n - 3) \right) \leq \frac{8\lambda(1 - e^{-\lambda})}{n - 1},
\]

where $\lambda = \mathbb{E}(W) = \left( \binom{n}{2} \right) d^{-1}$. \hfill \Box

---

The “coupling” approach. [Stein 1986].

**Theorem.** For each $i \in \Gamma$, divide $\Gamma \setminus \{i\}$ into two subsets $\Gamma_i^s$ and $\Gamma_i^w$. Let $Z_i = \sum_{j \in \Gamma_i^s} X_j$ and $W_i = \sum_{j \in \Gamma_i^w} X_j$. Let a random variable $\tilde{W}_i$ such that

\[
\mathcal{L}(\tilde{W}_i) = \mathcal{L}(W_i | X_i = 1)
\]

be defined on the same probability space as $W_i$. Then,

\[
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} (p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i))
\]

\[
+ \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} p_i \mathbb{E} \left| W_i - \tilde{W}_i \right|.
\]
Proof:

$$
\mathbb{E}(\lambda f_A(W + 1) - W f_A(W))
= \sum_{i \in \Gamma} \mathbb{E} \left( \eta_i f_A(W + 1) - \eta_i f_A(W_i + 1) + \eta_i f_A(W_i + 1) - \eta_i f_A(W_i + 1) - \eta_i f_A(W) \right)
\leq \sup_{k \in \mathbb{Z}_+} |f_A(k + 1) - f_A(k)| \sum_{i \in \Gamma} (\eta_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i))
+ \sum_{i \in \Gamma} \eta_i \mathbb{E} \left( f_A(W_i + 1) - \mathbb{E}(f_A(W_i + 1) | X_i = 1) \right),
$$

where the last term equals

$$
\sum_{i \in \Gamma} \eta_i \mathbb{E} \left( f_A(W_i + 1) - f_A(W_i + 1) \right)
\leq \sup_{k \in \mathbb{Z}_+} |f_A(k + 1) - f_A(k)| \sum_{i \in \Gamma} \eta_i \mathbb{E}|W_i - \tilde{W}_i|.
$$

\[ \square \]

Example. (Classical occupancy problem.) [Barbour and Holst 1989] $r$ balls are thrown independently into $n$ equiprobable boxes. Let $W$ be the number of empty boxes. Then,

$$
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \left( 1 - \exp \left( -n \left( \frac{n - 1}{n} \right)^r \right) \right) \times \left( n \left( \frac{n - 1}{n} \right)^r - (n - 1) \left( \frac{n - 2}{n - 1} \right)^r \right).
$$

If $r = n a_n$ with $\lim_{n \to \infty} a_n = \infty$, then

$$
d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) = O(\lambda e^{-a_n}) \text{ as } n \to \infty.
$$

If $a_n = \log n - \log c$, then $\lim_{n \to \infty} \lambda = c$.

Proof: For each $i \in \Gamma = \{1, \ldots, n\}$, let $X_i$ be the indicator for the event “the $i$th box is empty”, so $W = \sum_{i \in \Gamma} X_i$. Let $\Gamma^c_i = \emptyset$ and $\Gamma_i^w = \Gamma \setminus \{i\}$. Define $\{\tilde{X}_{i,j}, j \in \Gamma_i^w\}$ in the following way. Take those balls which have landed in the $i$th box, throw them independently into other boxes,
and let $\tilde{X}_{i,j}$ be the indicator for the event "the $j$th box is empty after this". Then

$$\mathcal{L}(\tilde{X}_{i,j}; j \in \Gamma_i^w) = \mathcal{L}(X_j; j \in \Gamma_i^w | X_i = 1),$$

since for each ball the probability of ending up in a particular box is $\frac{1}{n} + \frac{1}{n(n-1)} = \frac{1}{n-1}$, implying that, for each $\Gamma' \subset \Gamma_i^w$,

$$P(\tilde{X}_{i,j} = 1 \forall j \in \Gamma') = \left(\frac{n - |\Gamma'| - 1}{n - 1}\right)^r$$

$$= P(X_j = 1 \forall j \in \Gamma' | X_i = 1).$$

Let $\tilde{W} = \sum_{j \in \Gamma_i^w} \tilde{X}_{i,j}$. Observing that $\tilde{X}_{i,j} \leq X_j$ for each $j \in \Gamma_i^w$, we get

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda))$$

$$\leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} p_i \left(p_i + \mathbb{E}[W_i - \tilde{W}_i]\right)$$

$$= \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} p_i \mathbb{E}(X_i + \sum_{j \in \Gamma_i^w} (X_j - \tilde{X}_{i,j}))$$

$$= \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} p_i \mathbb{E}(W - \sum_{j \in \Gamma_i^w} \tilde{X}_{i,j})$$

$$= \frac{1 - e^{-\lambda}}{\lambda} \left(\mathbb{E}(W)^2 - \sum_{i \in \Gamma} \sum_{j \in \Gamma_i^w} \mathbb{E}(X_i X_j)\right)$$

$$= \frac{1 - e^{-\lambda}}{\lambda} \left(\mathbb{E}(W)^2 - \mathbb{E}(W^2) + \mathbb{E}(W)\right)$$

$$= \left(1 - \exp\left(-n\left(\frac{n - 1}{n}\right)^r\right)\right) \times \left(n\left(\frac{n - 1}{n}\right)^r - (n - 1)\left(\frac{n - 2}{n - 1}\right)^r\right).$$

$\Box$
Monotone couplings [Barbour and Holst 1989; Barbour, Holst and Janson 1992].

**Theorem.** For each $i \in \Gamma$, let random variables $\{\tilde{X}_{i,j}; j \neq i\}$ such that

$$\mathcal{L}(\tilde{X}_{i,j}; j \neq i) = \mathcal{L}(X_j; j \neq i|X_i = 1)$$

be defined on the same probability space as $\{X_j; j \neq i\}$. If $\Gamma \setminus \{i\}$ can be divided into three subsets $\Gamma_i^+, \Gamma_i^-$, and $\Gamma_i^0$, such that $\tilde{X}_{i,j} \geq X_j$ for $j \in \Gamma_i^+$ and $\tilde{X}_{i,j} \leq X_j$ for $j \in \Gamma_i^-$, then,

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \left( \sum_{i \in \Gamma} p_i^2 \right) + \sum_{i \in \Gamma} \sum_{j \in \Gamma_i^+} \text{Cov}(X_i, X_j) + \sum_{i \in \Gamma} \sum_{j \in \Gamma_i^-} \left| \text{Cov}(X_i, X_j) \right|$$

$$+ \sum_{i \in \Gamma} \sum_{j \in \Gamma_i^0} (\mathbb{E}(X_i X_j) + p_i p_j).$$

**Proof:** For each $i \in \Gamma$, let $\Gamma_i^s = \emptyset$ and $\Gamma_i^w = \Gamma \setminus \{i\}$, and let $\tilde{W}_i = \sum_{j \neq i} \tilde{X}_{i,j}$. Then,

$$p_i \mathbb{E}|W_i - \tilde{W}_i| = p_i \mathbb{E} \left| \sum_{j \in \Gamma_i^w} (X_j - \tilde{X}_{i,j}) \right|$$

$$\leq p_i \mathbb{E} \left( \sum_{j \in \Gamma_i^+} (\tilde{X}_{i,j} - X_j) \right) + p_i \mathbb{E} \left( \sum_{j \in \Gamma_i^-} (X_j - \tilde{X}_{i,j}) \right)$$

$$+ p_i \mathbb{E} \left( \sum_{j \in \Gamma_i^0} (\tilde{X}_{i,j} + X_j) \right)$$

$$= \sum_{j \in \Gamma_i^+} \text{Cov}(X_i, X_j) + \sum_{j \in \Gamma_i^-} \left| \text{Cov}(X_i, X_j) \right|$$

$$+ \sum_{j \in \Gamma_i^0} (\mathbb{E}(X_i X_j) + p_i p_j).$$

$\square$
Definition. \( \{ X_i; i \in \Gamma \} \) are called positively related if the conditions of the previous theorem hold with \( \Gamma_i^- = \Gamma_i^0 = \emptyset \), and negatively related if they hold with \( \Gamma_i^+ = \Gamma_i^0 = \emptyset \).

If \( \{ X_i; i \in \Gamma \} \) are positively related, then
\[
d_{TV}(L(W), Po(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \left( \text{Var}(W) - \lambda + 2 \sum_{i \in \Gamma} p_i^2 \right).
\]

If \( \{ X_i; i \in \Gamma \} \) are negatively related, then
\[
d_{TV}(L(W), Po(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \left( \lambda - \text{Var}(W) \right).
\]

Theorem. \( \{ X_i; i \in \Gamma \} \) are positively (negatively) related if and only if, for each \( i \in \Gamma \) and each increasing function \( \phi : \{0,1\}^{n-1} \rightarrow \{0,1\} \),
\[
\mathbb{E}(\phi(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n) | X_i = 1)
\geq (\leq) \mathbb{E}(\phi(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n)).
\]

Proof: Necessity: immediate. Sufficiency: use Strassen’s theorem. \( \square \)

Definition. [Esary, Proschan and Walkup 1967.] The random variables \( \{ X_i; i \in \Gamma \} \) are called associated if they satisfy the FKG inequality: if \( f \) and \( g \) are bounded increasing functions, then
\[
\mathbb{E}(f(X_i; i \in \Gamma)g(X_i; i \in \Gamma))
\geq \mathbb{E}(f(X_i; i \in \Gamma)) \mathbb{E}(g(X_i; i \in \Gamma)).
\]
**Theorem.** Associated indicator variables are positively related.

**Proof:** For each $i \in \Gamma$ and each increasing function $\phi : \{0, 1\}^{n-1} \to \{0, 1\}$, use the FKG inequality with $f(X_i; i \in \Gamma) = \phi(X_1, ..., X_{i-1}, X_{i+1}, ..., X_n)$ and $g(X_i; i \in \Gamma) = X_i$. □

**Theorem.** Independent random variables are associated.

**Theorem.** Increasing functions of associated random variables are associated.

---

**Example.** (Extremes of MA processes.) [Barbour, Holst and Janson 1992]. Let \( \{Z_i; i \in \mathbb{Z}\} \) be I.I.D., and let $\eta_i = \sum_{k=0}^{q} c_k Z_{i-k}$, where $c_k \geq 0$. Let $X_i = I\{\eta_i > a\}$, and let $W = \sum_{i=1}^{n} X_i$. \( \{X_i; i = 1, \ldots, n\} \) are associated, so

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq \frac{1 - e^{-\lambda}}{\lambda} \left( \text{Var}(W) - \lambda + 2 \sum_{i=1}^{n} p_i^2 \right).$$

It is easy to see that

$$\frac{\text{Var}(W)}{\lambda} - 1 = -\frac{\lambda}{n} + 2 \sum_{i=1}^{n} \frac{n-1}{n} \left( \mathbb{P}(\eta_i > a, \eta_0 > a) - \mathbb{P}(\eta_i > a) \right) \mathbb{P}(\eta_i > a).$$

In the special case $Z_i \sim U(0,1)$ and $\eta_i = Z_i + Z_{i-1}$, let $a = 2 - \sqrt{2\lambda/n}$ (where $n \geq 2\lambda$). Then, $p_i = \lambda/n$, and

$$\frac{\text{Var}(W)}{\lambda} - 1 = -\frac{\lambda}{n} + \frac{2(n-1)}{n} \left( \sqrt{\frac{8}{3}} \sqrt{n - \lambda} - \lambda \right),$$

implying that

$$d_{TV}(\mathcal{L}(W), Po(\lambda)) \leq (1 - e^{-\lambda}) \left( \frac{\lambda}{n} + \frac{2(n-1)}{n} \left( \sqrt{\frac{8}{3}} \sqrt{n - \lambda} - \lambda \right) \right).$$
**Definition.** [Joag-Dev and Proschan 1983.] The random variables \( \{X_i; i \in \Gamma\} \) are called \textit{negatively associated} if, whenever \( f \) and \( g \) are bounded increasing functions and \( \Gamma^1 \) and \( \Gamma^2 \) are \textit{disjoint} subsets of \( \Gamma \),

\[
\mathbb{E}(f(X_i; i \in \Gamma^1)g(X_i; i \in \Gamma^2)) \\
\leq \mathbb{E}(f(X_i; i \in \Gamma^1))\mathbb{E}(g(X_i; i \in \Gamma^2)).
\]

**Theorem.** Negatively associated indicator variables are negatively related.

**Theorem.** Independent random variables are negatively associated.

**Theorem.** Increasing (decreasing) functions of disjoint subsets of negatively associated random variables are negatively associated.
The “detailed coupling” approach.

**Theorem.** For each $i \in \Gamma$, divide $\Gamma \setminus \{i\}$ into two subsets $\Gamma_i^x$ and $\Gamma_i^w$. Let $Z_i = \sum_{j \in \Gamma_i^x} X_j$ and $W_i = \sum_{j \in \Gamma_i^w} X_j$. Let a random variable $\sigma_i$ be defined on the same probability space as $X_i$ and $W_i$, and let, for each $s \in \mathbb{R}$, a random variable $\tilde{W}_{i,s}$ such that

$$\mathcal{L}(\tilde{W}_{i,s}) = \mathcal{L}(W_i | X_i = 1, \sigma_i = s)$$

be defined on the same probability space as $W_i$. Then,

$$d_{TV}(\mathcal{L}(W), \text{Po}(\lambda))$$

$$\leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} \left( p_i \mathbb{E}(X_i + Z_i) + \mathbb{E}(X_i Z_i) \right)$$

$$+ \frac{1 - e^{-\lambda}}{\lambda} \sum_{i \in \Gamma} \mathbb{E} \left( X_i \mathbb{E} \left| W_i - \tilde{W}_{i,s} \right|_{s=\sigma_i} \right).$$

The number of points of a point process [Barbour and Brown 1992].

**Theorem.** Let $\xi$ be a point process on $(S, \mathcal{S})$, where $S$ is a locally compact second countable Hausdorff topological space, with a locally finite expectation measure $\nu$. For each $x \in S$, let a Palm process $\xi^x$ be defined on the same probability space as $\xi$. Then, for any relatively compact $B \in \mathcal{S}$,

$$d_{TV}(\mathcal{L}(\xi(B)), \text{Po}(\nu(B)))$$

$$= \sup_{A \subset \mathbb{Z}_+} \left| \mathbb{E}(\xi(B) f_A(\xi(B)) - \nu(B) f_A(\xi(B) + 1)) \right|$$

$$= \sup_{A \subset \mathbb{Z}_+} \left| \int_B \mathbb{E}(f_A(\xi^x(B)) - f_A(\xi(B) + 1)) d\nu(x) \right|$$

$$\leq \frac{1 - e^{-\nu(B)}}{\nu(B)} \int_B \mathbb{E} \left| \xi(B) - \xi^x(B) + 1 \right| d\nu(x).$$
Poisson process approximation
[Barbour and Brown 1992].

Let \( \xi \) be a point process on \((S, \mathcal{S})\) with locally finite expectation measure \( \nu \), and let \( B \subset S \) be relatively compact. Then, \( \xi(B) \) is a measurable function on the space of locally finite counting measures on \((S, \mathcal{S})\). Hence,

\[
d_{TV}(\mathcal{L}(\xi(B)), \mathcal{P}(\nu(B))) \leq d_{TV}(\mathcal{L}(\xi), \mathcal{L}(\xi_0)),
\]

where \( \xi_0 \) is a Poisson point process on \((S, \mathcal{S})\) with expectation measure \( \nu \).

The right-hand side can be bounded using Stein’s method, but these bounds are not in general as sharp as desired. We may also use

\[
d_{TV}(\mathcal{L}(\xi(B)), \mathcal{P}(\nu(B))) \leq d_2(\mathcal{L}(\xi), \mathcal{L}(\xi_0)),
\]

where \( d_2 \) is the Wasserstein distance. For the right-hand side in this inequality, better bounds can be obtained using Stein’s method.

---

Poisson-Charlier expansions [Barbour 1987].

**Theorem.** Let \( \{X_i; i \in \Gamma\} \) be independent. For each \( n \in \mathbb{Z}_+ \), let \( C_n(\lambda, x) \) be the \( n \)th order Charlier polynomial,

\[
C_n(\lambda, x) = \sum_{r=0}^{n} \binom{n}{r} (-1)^{n-r} \lambda^{-r} x(x-1)\ldots(x-r+1).
\]

For each \( l \geq 1 \), define the \( l \)th order Poisson-Charlier signed measure on \( \mathbb{Z}_+ \) by

\[
Q_l(i) = \left( \frac{e^{-\lambda} \lambda^i}{i!} \right)
\]

\[
\times \left( 1 + \sum_{s=1}^{l-1} \sum_{[s]} \prod_{j=1}^{s} \left[ \frac{1}{r_j!} \left( \frac{(-1)^{j} \lambda_{j+1}}{j+1} \right)^{r_j} \right] C_{R+s}(\lambda, i) \right),
\]

where \( \lambda_{j+1} = \sum_{i \in \Gamma} p_{i}^{j+1} \), and \( \sum_{[s]} \) denotes the sum over all \( s \)-tuples \((r_1,\ldots,r_s) \in \mathbb{Z}_+^s\) such that \( \sum_{j=1}^{s} jr_j = s \), and \( R = \sum_{j=1}^{s} r_j \).
Then, for each $h : \mathbb{Z}_+ \to \mathbb{R}$ and $l \geq 1$,
\[
\left| \mathbb{E}(h(W)) - \int_{\mathbb{Z}_+} h dQ_t \right| \leq \frac{1 - e^{-\lambda}}{\lambda} \lambda_{l+1} 2^{2l-1} \|h\|_\infty,
\]
and, if $\frac{1-e^{-\lambda}}{\lambda} \lambda_2 \leq \frac{1}{8}$,
\[
\left| \mathbb{E}(h(W)) - \int_{\mathbb{Z}_+} h dQ_t \right| \leq \frac{1 - e^{-\lambda}}{\lambda} \lambda_{l+1} 2^{2l} \sqrt{\frac{2}{e\lambda}} \|h\|_1.
\]

**Sketch proof:** Let $X$ be an indicator variable with $\mathbb{E}(X) = p$. Then, for any $j \in \mathbb{Z}_+$,
\[
\sum_{s=1}^{l-1} (-1)^s p^{s+1} \mathbb{E}(\triangle^s f(X + j + 1))
\]
\[
= (-1)^{l+1} p^{l+1} \triangle^l f(j+1) - p^2 \triangle f(j+1),
\]
where $\triangle^l f$ is the $l$th forward difference of $f$. Moreover,
\[
\mathbb{E}(X f(X+j) - p f(X+j+1)) = -p^2 \triangle f(j+1).
\]
Letting $X = X_i$ and $j = W - X_i$, taking expectations, and summing over $i$,
\[
\left| \mathbb{E}(\lambda f(W + 1) - W f(W)) - \sum_{s=1}^{l-1} (-1)^s p^{s+1} \mathbb{E}(\triangle^s f(W+1)) \right| \leq \lambda_{l+1} \|\triangle^l f\|.
\]
Choosing $f = S_0 h$ as the solution to the Stein equation for $h : \mathbb{Z}_+ \to \mathbb{R}$, we get
\[
\left| \mathbb{E}(h(W)) - \int_{\mathbb{Z}_+} h d\mu_0 \right|
\]
\[ - \sum_{s=1}^{l-1} (-1)^{s+1} \lambda_{s+1} \mathbb{E}(\triangle^s S_0 h(W+1)) \leq \lambda_{l+1} \| \triangle^l f \|. \]

Using this expression iteratively we get
\[ \mathbb{E}(h(W)) = \sum_{(l)} \left( \prod_{j=1}^{k} (-1)^{s_j+1} \lambda_{s_j+1} \right) \mathbb{E}\left( \left( \prod_{j=1}^{k} (\triangle^s S_0 h)(Z) \right) + \eta, \right), \]
where \( Z \sim \text{Po}(\lambda) \), \( \sum_{(l)} \) denotes the sum over
\[ \{(s_1, \ldots, s_{k+1}) \in (\mathbb{Z}_+)^{k+1}; k \geq 0, \sum_{j=1}^{k+1} s_j = l\}, \]
and
\[ |\eta| \leq \sum_{(l)} \left( \prod_{j=1}^{k+1} \lambda_{s_j+1} \right) \| \prod_{j=1}^{k+1} (\triangle^s S_0 h) \|. \]

Rewriting, using the identities
\[ \mathbb{E}(C_n(\lambda, Z)(\triangle f)(Z)) = \mathbb{E}(C_{n+1}(\lambda, Z) f(Z)); \]
\[ \mathbb{E}(C_n(\lambda, Z)S_0 h(Z)) = -\frac{1}{n+1} \mathbb{E}(C_{n+1}(\lambda, Z) h(Z)), \]
and using the preliminary bounds to bound \( \eta \), the result is obtained. \( \square \)

Poisson-Charlier expansions with small relative errors [Barbour and Jensen 1989].

**Theorem.** Let \( \{X_i; i \in I\} \) be independent. For any \( m \in [1, n-1] \), let \( \{X_{m,1}, \ldots, X_{m,n}\} \) be independent indicator variables with
\[ \mathbb{E}(X_{m,i}) = p_{m,i} = \frac{\phi_m p_i}{1 - p_i + \phi_m p_i}, \]
where \( \phi_m \) is the solution to the equation
\[ \sum_{i \in I} p_{m,i} = \frac{\phi_m p_i}{1 - p_i + \phi_m p_i} = m. \]
Then,
\[ \mathbb{P}(W = k) = \phi_m^{-k} \prod_{i=1}^{n} \left( 1 - p_i + \phi_m p_i \right) \left( Q_{m,i}(k) + \eta_{m,i} \right), \]
where, if \( \frac{1-e^{-m}}{m} \lambda_{m,2} \leq \frac{1}{8}, \)
\[ |\eta_{m,l}| \leq \frac{1-e^{-m}}{m} \sqrt{\frac{2}{em}} 2^{2l} \lambda_{m,l+1}. \]
In particular,
\[ \mathbb{P}(W = m) = \phi_m^{-m} \prod_{i=1}^{n} (1-p_i+\phi_i p_i)(Q_{m,l}(m)+\eta_{m,l}). \]
Since \( Q_{m,1}(m) \sim \sqrt{m} \), it follows that if \( \frac{1}{m} \lambda_{m,2} \) is “small”, the relative error of this approximation is, for fixed \( l \geq 1 \), of order at most \( \frac{1}{m} \lambda_{m,l+1} \).

Moreover: if \( m \leq \lambda \), then \( \lambda_{m,l+1} \leq \lambda_{l+1} \), while if \( \lambda \leq m \leq \lambda^2/(2\lambda_2) \), then
\[ \lambda_{m,l+1} \leq \phi^{l+1}_m \lambda_{l+1} \leq (1+2(m/\lambda-1))^{l+1} \lambda_{l+1}. \]

**Theorem.** Let \( \{X_i; i \in \Gamma\} \) be independent. Then, uniformly in \( m \) satisfying the inequalities \( \lambda \leq m \leq \lambda^2/(2\lambda_2) \) and \( 1 + 4(m - \lambda)^2/\lambda \leq (16\lambda_2/\lambda)^{-1} \),
\[ \mathbb{P}(W = m) = \frac{\lambda^m e^{-\lambda}}{m!} \left( 1 + O\left( \frac{\lambda_2}{\lambda} \right) + O\left( \frac{\lambda^2 (m - \lambda)^2}{\lambda^2} \right) \right). \]
Also, if \( \lambda_2/\lambda \leq \frac{1}{8} \) and \( \max_{i \in \Gamma} p_i \leq \frac{1}{2} \), then, uniformly in \( m \) satisfying \( 0 \leq m \leq \lambda \) and \( (m - \lambda)^2/\lambda \leq (\lambda_2/\lambda)^{-1} \),
\[ \mathbb{P}(W = m) = \frac{\lambda^m e^{-\lambda}}{m!} \left( 1 + O\left( \frac{\lambda_2}{\lambda} \right) + O\left( \frac{\lambda^2 (m - \lambda)^2}{\lambda^2} \right) \right). \]
Similar results hold for tail probabilities. Also, similar results were obtained, independently and using a different method, by Chen and Choi (1992).
References:


Stein’s method for Poisson and compound Poisson approximations: II

Torkel Erhardsson
Department of Mathematics
KTH

Plan:

1. Definition of POIS(π).
2. Why compound Poisson?
3. The Stein equation and properties of its solutions.
5. Further properties of the solutions of the Stein equation.
6. Stein’s method for signed compound Poisson measure approximation.
6. Connections to Poisson process and compound Poisson process approximation.
The POIS(π) distribution.

**Definition.** POIS(π) is the probability distribution on $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$ which has the characteristic function
\[
\varphi(t) = \exp\left(-\int_0^\infty (1 - e^{itx})d\pi(x)\right),
\]
where the measure π satisfies
\[
\int_0^\infty (x \wedge 1)d\pi(x) < \infty.
\]
If $\|\pi\| = \pi(\mathbb{R}_+) < \infty$, then POIS(π) = $\mathcal{L}(\sum_{i=1}^U T_i)$, where all random variables are independent, $\mathcal{L}(T_i) = \pi = \pi/\|\pi\|$ for each $i \geq 1$, and $\mathcal{L}(U) = \text{Po}(\|\pi\|)$.

We call π the *compounding measure*, and $\pi$ the *compounding distribution* (if $\|\pi\| < \infty$).

**Why compound Poisson approximation?**

**Example.** Let \( \{\eta_i; i \in \mathbb{Z}\} \) be an I.I.D. sequence of indicator variables such that $\mathbb{E}(\eta_i) = p$. Let $X_i = I\{\eta_i = \eta_{i-1} = \ldots = \eta_{i-r+1} = 1\}$, where $r \geq 2$. Let $W = \sum_{i=1}^n X_i$. It can be shown using Stein’s method for Poisson approximation (the coupling approach) that
\[
d_{TV}(\mathcal{L}(W), \text{Po}(np^r)) \leq \frac{2p}{1-p} + p^r.
\]

The approximation error is large because the patterns 11...11 tend to appear in *clumps*. (A rather common phenomenon.)

**Idea.** (Aldous’s “Poisson clumping heuristic”.) Consider the *number of clumps* as approximately Poisson distributed, and the *clump sizes* as approximately I.I.D. Leads to a compound Poisson approximation.
The Stein equation for POIS(\(\pi\)) (lattice case) [Barbour, Chen and Loh 1992].

Let \((S, \mathcal{F}, \mu) = (\mathbb{Z}_+, \mathcal{B}_{\mathbb{Z}_+}, \mu)\), where \(\mathcal{B}_{\mathbb{Z}_+}\) is the power \(\sigma\)-algebra. Let \(\chi\) be the set of all functions \(f : \mathbb{Z}_+ \to \mathbb{R}\). Let \(\mu_0 = \text{POIS}(\pi)\), where \(\sum_{i=1}^{\infty} i \pi_i < \infty\). Define \(T_0 : \mathcal{F}_0 \to \chi\) by

\[
(T_0 f)(k) = kf(k) - \sum_{i=1}^{\infty} i \pi_i f(k+i) \quad \forall k \in \mathbb{Z}_+,
\]

where \(\mathcal{F}_0\) is the set of functions \(f : \mathbb{Z}_+ \to \mathbb{R}\) such that the right-hand side is finite.

**Theorem.** If \(h : \mathbb{Z}_+ \to \mathbb{R}\) is bounded, the Stein equation

\[
T_0 f = h - \int_{\mathbb{Z}_+} h d\mu_0
\]

has a unique bounded solution \(f\) (except that \(f(0)\) can be chosen arbitrarily).

The solution is given by

\[
f(k) = \frac{1}{k} \sum_{i=0}^{\infty} ||\pi||^i \mathbb{E}\left((h(k+S_i) - \int_{\mathbb{Z}_+} h d\mu_0) \frac{\prod_{j=1}^{i} Y_j}{\prod_{j=1}^{i} (k + S_j)}\right),
\]

where \(\{Y_i; i \in \mathbb{Z}_+\}\) are I.I.D. with distribution \(\pi\), and \(S_i = \sum_{j=1}^{i} Y_j\).

Alternatively, the solution is given by

\[
f(k) = \sum_{i=k}^{\infty} a_{i,k}(h(i) - \int_{\mathbb{Z}_+} h d\mu_0),
\]

where \(a_{k,k} = 1/k\) and

\[
a_{k+i,k} = \sum_{j=1}^{i} \frac{j \pi_j}{k+i} a_{k+i-j,k} \quad \forall i \geq 1.
\]

**Corollary.** A probability measure \(\mu\) on \(\mathbb{Z}_+\) is POIS(\(\pi\)) if and only if

\[
\int_{\mathbb{Z}_+} (T_0 f) d\mu = 0
\]

for all bounded \(f : \mathbb{Z}_+ \to \mathbb{R}\).
Preliminary bounds [Barbour, Chen and Loh 1992].

**Theorem.** Let \( f_A \) be the solution of the Stein equation with \( h = I_A \) (where \( A \subset \mathbb{Z}_+ \)). Define

\[
H_1(\pi) = \sup_{A \subset \mathbb{Z}_+} \sup_{k \in \mathbb{Z}_+} |f_A(k + 1) - f_A(k)|, \\
H_0(\pi) = \sup_{A \subset \mathbb{Z}_+} \sup_{k \in \mathbb{Z}_+} |f_A(k)|.
\]

Then,

\[
\max(H_0(\pi), H_1(\pi)) \leq \left(1 + \frac{1}{\pi_1}\right)e^{\|\pi\|}. \tag{1}
\]

Moreover, if

\[i\pi_i - (i + 1)\pi_{i+1} \geq 0 \quad \forall i \geq 1,
\]

then,

\[
H_1(\pi) \leq 1 \wedge \frac{1}{D_\pi} \left(\frac{1}{4D_\pi} + \log^+ 2D_\pi\right); \\
H_0(\pi) \leq \begin{cases} 
\frac{1}{\sqrt{D_\pi}} \left(2 - \frac{1}{\sqrt{D_\pi}}\right), & \text{if } D_\pi > 1; \\
1, & \text{if } D_\pi \leq 1,
\end{cases} \tag{2}
\]

where \( D_\pi = \pi_1 - 2\pi_2 \).

**Sketch proof:** (1) can be shown analytically with some effort using the second representation of \( f_A \). However, from the first representation of \( f_A \) we get

\[
|f_A(k)| \leq \sum_{i=0}^{\infty} \|\pi\|^i e^{\|\pi\|} \left(\prod_{j=1}^{i} Y_j\right) \leq \sum_{i=0}^{\infty} \|\pi\|^i \frac{1}{i!} \leq e^{\|\pi\|}.
\]

For (2), we see that if \( f(k) = \nabla g(k) = g(k) - g(k - 1) \) and if all sums converge,

\[
(T_0 f)(k) = -\sum_{i=1}^{\infty} \left(i\pi_i - (i + 1)\pi_{i+1}\right)g(k + i)
\]

\[-kg(k-1)+(\pi_1+k)g(k) = -\mathcal{A}g(k) \quad \forall k \in \mathbb{Z}_+,
\]

where \( \mathcal{A} \) is the generator of a batch immigration-death process \( \{Z_t; t \in \mathbb{R}_+\} \), with stationary distribution POIS(\(\pi\)).
The corresponding Poisson's equation is
\[ -\mathcal{L}g = h - \int_{Z_+} h d\mu_0. \]
If \( h \) is bounded, then this equation has the solution
\[ g(k) = \int_0^\infty \left( \mathbb{E}(h(Z_t)|Z_0 = k) - \int_{Z_+} h d\mu_0 \right) dt, \]
and \( f = \nabla g \) is the unique bounded solution of the Stein equation.
For each \( k \geq 1 \), define four coupled batch immigration-death processes of the above kind, such that \( Z_t^{(0)} \) starts at \( k \), and
\[
\begin{align*}
Z_t^{(1)} &= Z_t^{(0)} + I\{\tau_1 > t\}, \\
Z_t^{(2)} &= Z_t^{(0)} + I\{\tau_2 > t\}, \\
Z_t^{(3)} &= Z_t^{(1)} + I\{\tau_2 > t\},
\end{align*}
\]
where \( \tau_1 \sim \exp(1) \) and \( \tau_2 \sim \exp(1) \) are independent of each other and of \( Z^{(0)} \).
Then,
\[
f_{A}(k+2) - f_{A}(k+1) = \int_0^\infty \mathbb{E} \left( I_{A}(Z_t^{(3)}) - I_{A}(Z_t^{(2)}) \right) dt - I_{A}(Z_t^{(1)}) + I_{A}(Z_t^{(0)}) dt = \int_0^\infty e^{-2t} \left( \mathbb{P}(Z_t^{(0)} \in A-2) - 2\mathbb{P}(Z_t^{(0)} \in A-1) + \mathbb{P}(Z_t^{(0)} \in A) \right) dt.
\]
We can write \( Z_t^{(0)} = W_t + Y_t \), where \( W_t \) and \( Y_t \) are independent and \( Y_t \) is the number of individuals who immigrated in batches of size 1 after time 0 and are still alive at time \( t \). Then,
\[
\mathbb{P}(Z_t^{(0)} \in A-2) - 2\mathbb{P}(Z_t^{(0)} \in A-1) + \mathbb{P}(Z_t^{(0)} \in A)
\]
\[
= \sum_{k=0}^\infty \mathbb{P}(W_t = k) \sum_{l \in A-k} \left( \mathbb{P}(Y_t = l - 2) - 2\mathbb{P}(Y_t = l - 1) + \mathbb{P}(Y_t = l) \right)
\]
\[
\leq \left( 2 \wedge \frac{1}{(1 - e^{-t}) D_\pi} \right).
\]
\( \square \)
Construction of explicit error bounds in compound Poisson approximation: sums of nonnegative integer valued random variables.

**Notation:** \( \Gamma = \{1, \ldots, n\} \); \( \{X_i; i \in \Gamma\} \) are nonnegative integer valued random variables with finite means; \( W = \sum_{i \in \Gamma} X_i \); \( \mu = \mathcal{L}(W) \); and \( \mu_0 = \text{POIS}(\pi) \), where the compounding measure \( \pi \) will be defined below.

**Goal:** to find a bound for

\[
d_{TV}(\mu, \mu_0) = \sup_{A \in \mathcal{Z}_+} \mu(A) - \mu_0(A) = \sup_{A \in \mathcal{Z}_+} \int_{\mathcal{Z}_+} (T_0 f_A) d\mu
\]

\[
= \sup_{A \in \mathcal{Z}_+} \mathbb{E} \left( \sum_{i=1}^{\infty} i \pi_i f_A(W + i) - W f_A(W) \right),
\]

where \( f_A \) is the unique bounded solution of the Stein equation with \( h = I_A \). Alternatively, we may try to bound

\[
d_k(\mu, \mu_0) = \sup_{m \in \mathcal{Z}_+} \mu([m, \infty)) - \mu_0([m, \infty))
\]

\[
= \sup_{m \in \mathcal{Z}_+} \mathbb{E} \left( \sum_{i=1}^{\infty} i \pi_i f_{[m, \infty)}(W + i) - W f_{[m, \infty)}(W) \right).
\]

The “local”, “coupling”, and “detailed coupling” approaches [Barbour, Chen and Loh 1992; Roos 1994; Barbour and Utev 1999].

**Theorem.** For each \( i \in \Gamma \), divide \( \Gamma \setminus \{i\} \) into three subsets \( \Gamma_i^{vs}, \Gamma_i^w, \) and \( \Gamma_i^b \), so that, informally,

\( \Gamma_i^{vs} = \{ j \in \Gamma \setminus \{i\}; X_j \text{ "very strongly" dependent on } X_i \}; \)

\( \Gamma_i^w = \{ j \in \Gamma \setminus \{i\}; X_j \text{ "weakly" dependent on } \{X_k; k \in \{i\} \cup \Gamma_i^{vs}\} \}). \)

Let \( Z_i = \sum_{j \in \Gamma_i^{vs}} X_j, W_i = \sum_{j \in \Gamma_i^w} X_j \) and \( U_i = \sum_{j \in \Gamma_i^b} X_j \). Define the “canonical” compounding measure \( \pi \) by

\[
\pi_k = \frac{1}{k} \sum_{i \in \Gamma} \mathbb{E}(X_i I\{X_i + Z_i = k\}) \quad \forall k \in \mathcal{Z}_+.
\]
Then,

\[ d_{TV}(\mathcal{L}(W), \text{POIS}(\pi)) \]

\[ \leq H_1(\pi) \sum_{i=1}^{n} \left( \mathbb{E}(X_i) \mathbb{E}(X_i + Z_i + U_i) + \mathbb{E}(X_i U_i) \right) \]

\[ + H_0(\pi) \sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{P}(X_i = j, Z_i = k-j\mid W_i) \]

\[ - \mathbb{P}(X_i = j, Z_i = k-j) \].

If, for each \( i \in \Gamma \), a random variable \( \sigma_i \) is defined on the same probability space as \( X_i, Z_i \) and \( W_i \), and if, for each \( i \in \Gamma, j \in \mathbb{Z}_+^I, k \in \mathbb{Z}_+^I, \) and \( s \in \mathbb{R}, \tilde{W}_i^{j,k,s} \) is defined on the same probability space as \( W_i \), and

\[ \mathcal{L}(\tilde{W}_i^{j,k,s}) = \mathcal{L}(W_i \mid X_i = j, Z_i = k-j, \sigma_i = s), \]

then the last term can be replaced by

\[ H_1(\pi) \sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{P}(X_i = j, Z_i = k-j) \mathbb{E}(\|W_i - \tilde{W}_i^{j,k,s}\|_{s=\sigma_i}). \]
Proof:

\[
\mathbb{E}
\left(\left(W f_{\mathcal{A}}(W) - \sum_{k=1}^{\infty} k \pi_{k} f_{\mathcal{A}}(W + k)\right)\right)
= \sum_{i=1}^{n} \mathbb{E}
\left(\sum_{k=1}^{\infty} \mathbb{E}(X_i f_{\mathcal{A}}(W) | \text{ } \{X_i + Z_i = k\}) f_{\mathcal{A}}(W + k)\right)
= \sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{E}
\left(\sum_{k=1}^{\infty} \mathbb{E}(X_i I\{X_i + Z_i = k\}) f_{\mathcal{A}}(W_i + U_i + k)\right)
- \mathbb{P}(X_i = j, Z_i = k - j) f_{\mathcal{A}}(W_i + k)
= \sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{E}
\left(\sum_{k=1}^{\infty} \mathbb{E}(X_i I\{X_i + Z_i = k\}) f_{\mathcal{A}}(W_i + U_i + k)\right)
- \mathbb{P}(X_i = j, Z_i = k - j) f_{\mathcal{A}}(W_i + k)
+ \mathbb{P}(X_i = j, Z_i = k - j) f_{\mathcal{A}}(W_i + k)
\right).
\]

The sum of the first and third terms can be bounded by

\[
H_1(\pi) \sum_{i=1}^{n} \mathbb{E}(X_i) \mathbb{E}(X_i + Z_i + U_i) + E(X_i U_i)\).
\]

The second term can be bounded by

\[
H_0(\pi) \sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{P}(X_i = j, Z_i = k - j) \left| \mathbb{P}(X_i = j, Z_i = k - j) \right|
\]

or by

\[
\sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{P}(X_i = j, Z_i = k - j)
\]

\[
\times \mathbb{E}(f_{\mathcal{A}}(W_i + k) | X_i = j, Z_i = k - j) - f_{\mathcal{A}}(W_i + k)
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{P}(X_i = j, Z_i = k - j) \mathbb{E}(f_{\mathcal{A}}(\widetilde{W}_i^{j,k} + k) - f_{\mathcal{A}}(W_i + k))
\]

\[
\leq H_1(\pi) \sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{P}(X_i = j, Z_i = k - j) \mathbb{E}\left| W_i - \widetilde{W}_i^{j,k} \right|,
\]

or (analogously) by

\[
H_1(\pi) \sum_{i=1}^{n} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} j \mathbb{P}(X_i = j, Z_i = k - j) \mathbb{E}\left| W_i - \widetilde{W}_i^{j,k} \right|_{s=\sigma_i}.
\]
**Example.** Let \( \{X_i; i \in \Gamma\} \) be independent. Choosing \( \Gamma _i^{vs} = \Gamma _i^b = \emptyset \), we get:

\[
d_{TV}(\mathcal{L}(W), \text{POIS}(\pi)) \leq H_1(\pi) \sum_{i \in \Gamma} \mathbb{E}(X_i)^2.
\]

**Earlier results.** According to Le Cam (1965), Khintchine (1933) showed that

\[
d_{TV}(\mathcal{L}(W), \text{POIS}(\pi)) \leq \sum_{i \in \Gamma} \mathbb{P}(X_i > 0)^2.
\]

Le Cam also gives examples for which this bound is sharp.

If \( \mathcal{L}(X_i|X_i > 0) \) is the same for all \( i \in \Gamma \), Michel (1988) showed that

\[
d_{TV}(\mathcal{L}(W), \text{POIS}(\pi)) \leq \frac{1}{\|\pi\|} \sum_{i \in \Gamma} \mathbb{P}(X_i > 0)^2.
\]

**Example.** (Head runs.) [Roos 1993]. Let \( \{\eta_i; i \in \mathbb{Z}\} \) be an I.I.D. sequence of indicator variables such that \( \mathbb{E}(\eta_i) = p \). Let

\[
X_i = I\{\eta_i = \eta_{i-1} = \ldots = \eta_{i-r+1} = 1\},
\]

where we identify \( i + kn \) with \( i \) for each \( n \in \mathbb{Z} \). Let \( W = \sum_{i=1}^n X_i \). Choosing

\[
\Gamma _i^{vs} = \{j \in \Gamma; 1 \leq |i - j| \leq r - 1\};
\]

\[
\Gamma _i^b = \{j \in \Gamma; r \leq |i - j| \leq 2(r - 1)\};
\]

\[
\Gamma _i^w = \Gamma \setminus \{i\} \cup \Gamma _i^{vs} \cup \Gamma _i^b,
\]

we get:

\[
d_{TV}(\mathcal{L}(W), \text{POIS}(\pi)) \leq H_1(\pi) \sum_{i=1}^n \left( \mathbb{E}(X_i)\mathbb{E}(X_i+Z_i+U_i)+\mathbb{E}(X_iU_i) \right)
\]\n
\[
= H_1(\pi)(6r - 5)np^{2r}.
\]
The canonical compounding measure $\pi$ is

$$\pi_k = \begin{cases} n p^{r+k-1} (1-p)^2, & \text{if } k = 1, \ldots, r-1; \\ \frac{1}{2} n p^{r+k-1} (1-p) \times (2 + (2r - k - 2)(1-p)), & \text{if } k = r, \ldots, 2(r-1); \\ n p^{3r-2} \frac{1}{2r-1}, & \text{if } k = 2r - 1. \end{cases}$$

As $n \to \infty$, if $p = p(n)$ and $r = r(n)$, we get:

if $\mathbb{E}(W) = n p^r \leq C < \infty$ and $p \leq p' < 1$, the bound is $O(rp^r)$;
if $\mathbb{E}(W) \to \infty$ and $p \leq p' < \frac{1}{3}$, the bound is $O(rp^r \log(np^r))$;
if $\mathbb{E}(W) \to \infty$ and $p \leq p' < \frac{1}{5}$, the bound is $O(rp^r)$.

Examples of similar flavour from reliability theory, random graph theory, etc., can be found in Barbour and Chryssaphinou (2001).

**Example.** (Visits to rare sets by Markov chains.) [Erhardsson 1997]. Let $\{\eta_i; i \in \mathbb{Z}\}$ be a stationary irreducible discrete time Markov chain on the finite state space $S$, with stationary distribution $\nu$. Let $W = \sum_{i=1}^{n} I\{\eta_i \in B\}$, where $B \subset S$.

The problem can be reformulated as follows. Choose $a \in B^c$, and let $\tau_i^a$ be the first return time to $a$ by $\eta$ after time $i$. Let

$$X_i = I\{\eta_i = a\} \sum_{j=i+1}^{\tau_i^a} I\{\eta_j \in B\}.$$ 

Clearly, if $B$ is a rare set, $W \approx W' = \sum_{i=1}^{n} X_i$.

Now use the coupling approach. For each $i \in \Gamma$, choose $\Gamma^\nu_i = \Gamma^b_i = \emptyset$. For each $i \in \Gamma$ and $j \in \mathbb{Z}_+^\prime$, we construct a random sequence $\{\tilde{\eta}_k^i; k \in \mathbb{Z}\}$ on the same probability space as $\eta$, such that

$$\mathcal{L}(\tilde{\eta}_k^i; j) = \mathcal{L}(\eta|X_i = j).$$
This $\tilde{\eta}^i,j$ consists of a few short segments which are independent of $\eta$, and long segments which are time-shifted segments of $\eta$. Let

$$\tilde{X}^i,j_k = I\{\tilde{\eta}^i,j_k = a\} \sum_{l=k+1}^{\tau^{*}(\tilde{\eta}^i,j)} I\{\tilde{\eta}^i,j_l \in B\},$$

and let $\tilde{W}^j_i = \sum_{k\in\Gamma^w_i} \tilde{X}^i,j_k$. This gives:

$$d_{TV}(\mathcal{L}(W), \text{POIS}(\pi)) \leq 2H_1(\pi) \left( \mathbb{E}(\tau^a | \eta_0 \in B) + \mathbb{E}(\tau^a | \eta_0 \in B) + \frac{\mathbb{E}(\tau^a)}{\nu(a)} n\nu(B)^2 + 2\mathbb{P}(\tau^B < \tau^a) \right),$$

where $\nu(a)$ may be replaced by $\frac{1}{2}$ if the first term is multiplied by $\frac{3}{2}$, and

$$\pi_k = n\mathbb{P}(X_i = k) \quad \forall k \geq 1.$$

**Example.** (Clusters of random points.) [Barbour and Månnsson 2000]. Let $n$ random points be uniformly and independently distributed in the unit square $A$. Let $\Gamma$ be the set of $k$-subsets of $\{1, \ldots, n\}$. For each $i \in \Gamma$, let $X_i$ be the indicator that the points with indices in $i$ are covered by a translate of a square $C$ with side length $c$. Let $W = \sum_{i \in \Gamma} X_i$.

Use the detailed coupling approach, with random neighbourhoods. For each $i \in \Gamma$, let $R^{(4)}_i$ be the square centred at a randomly chosen point with index in $i$, with side length $4c$. Choose

$$\Gamma^w_i = \{ j \in \Gamma \setminus \{i\} ; \text{all points with indices in } j \text{ lie in } R^{(4)}_i \};$$

$$\Gamma^u_i = \{ j \in \Gamma ; j \cap i = \emptyset; \text{no points with indices in } j \text{ lie in } R^{(4)}_i \};$$

$$\Gamma^\prime_i = \Gamma \setminus (\{i\} \cup \Gamma^w_i \cup \Gamma^u_i).$$

The random variable $\sigma_i$ is chosen as the number of points with indices not in $i$ that lie in $R^{(4)}_i$. 

21
After some computations, they obtain:

\[
d_{TV}(\mathcal{L}(W), \text{POIS}(\pi)) = 2H_1(\pi) \binom{n}{k} 16 \binom{n-k}{k} k^4 |C|^{2k-1} \\
+ k^4 |C|^{2k-1} \left( \sum_{l=1}^{k-1} \binom{k}{l} \binom{n-k}{k-l} + 25 \binom{n-k}{k} |C| + 1 \right) \\
+ 32k^4 |C|^k \left( \frac{|C|}{1 - 16|C|} \right)^{k-1} (n-k) \binom{n-k}{k-1}.
\]

As \( n \to \infty \), if \( C = C(n) \) so that\n
\[
\mathbb{E}(W) = \binom{n}{k} k^2 |C|^{k-1} \leq K < \infty,
\]

the bound is \( O(n^{2k-1} |C|^{2k-2}) \).

Unfortunately, the canonical compounding measure \( \pi \) does not satisfy the condition

\[ k\pi_k - (k+1)\pi_{k+1} \geq 0 \quad \forall i \in \mathbb{Z}_+^t. \]

A modified Stein’s method for compound Poisson approximation [Barbour and Utev 1999].

The bound (1) cannot be much improved in general, since for some \( \pi \) such that the condition

\[ k\pi_k - (k+1)\pi_{k+1} \geq 0 \quad \forall i \in \mathbb{Z}_+^t \]

does not hold, there exists \( \beta > 0 \) and \( C(\pi') \) so that \( H_1(\pi) \geq C(\pi') e^{\beta \|\pi\|} \).

For each \( a \geq 1 \), define

\[
H_0^a(\pi) = \sup_{A \subseteq \mathbb{Z}_+} \sup_{k \geq a} |f_A(k+1) - f_A(k)|, \\
H_0^a(\pi) = \sup_{A \subseteq \mathbb{Z}_+} \sup_{k \geq a} |f_A(k)|.
\]
**Theorem.** Let $W$ be a nonnegative integer valued random variable, let $\mu = \mathcal{L}(W)$, and let $\mu_0 = \text{POIS}(\pi)$, where we assume that $m_2 = \sum_{i=1}^{\infty} i^2 \pi_i < \infty$. For each $0 < a < b < \infty$, let

$$u_{a,b}(x) = \left(\frac{x-a}{b-a}\right) I_{(a,b]}(x) + I_{(b,\infty)}(x).$$

Then,

$$d_{TV}(\mu, \mu_0) \leq \sup_{A \subseteq \mathbb{Z}_+} \left| \mathbb{E}\left( \sum_{i=1}^{\infty} i \pi_i f_A(W+i) u_{a,b}(W+i) \right) - W f_A(W) u_{a,b}(W) \right| + \mathbb{P}(W \leq b) \left( 1 + \frac{\|\pi\| m_2 H_0^a(\pi)}{b-a} \right).$$

In particular, we may choose $a = c\|\pi\| m_1$ and $b = \frac{1}{2} (1 + c)\|\pi\| m_1$, for $0 < c < 1$.

To bound the first term we can use the local/coupling approaches together with explicit bounds on $H_0^a(\pi)$ and $H_1^a(\pi)$, since it can be shown that

$$\sup_{A \subseteq \mathbb{Z}_+} \sup_{k \in \mathbb{Z}_+} f_A(k+1) u_{a,b}(k+1) - f_A(k) u_{a,b}(k) \leq H_1^a(\pi) + \frac{H_0^a(\pi)}{b-a};$$

$$\sup_{A \subseteq \mathbb{Z}_+} \sup_{k \in \mathbb{Z}_+} f_A(k) u_{a,b}(k) \leq H_0^a(\pi).$$

For the second term we also need an explicit bound for $\mathbb{P}(W \leq b)$. We may use e.g. Chebyshev’s inequality or Janson’s inequality.
**Theorem.** Assume that the generating function $\varphi_{\pi}(z) = \sum_{k=1}^{\infty} \pi_k z^k$ has radius of convergence $R > 1$, that $||\pi|| \geq 2$, and that

$$\min\left(\rho_1^*(\zeta), \frac{1}{2} \rho_2^*(\zeta)\right) > 0 \quad \forall 0 < \zeta \leq \pi,$$

where

$$\rho_1^*(\zeta) = \inf_{\zeta \leq \theta \leq \pi} \left(1 - \sum_{k=1}^{\infty} \pi_k \cos(k\theta)\right),$$

$$\rho_2^*(\zeta) = \inf_{\zeta \leq \theta \leq \pi} \left(1 - \frac{1}{m_1} \sum_{k=1}^{\infty} k\pi_k \cos(k\theta)\right),$$

and $m_1 = \sum_{k=1}^{\infty} k\pi_k$. Then there exist explicit constants $C_0(\pi)$, $C_1(\pi)$ and $C_2(\pi) < 1$, such that, for each $a \geq C_2(\pi)||\pi||m_1 + 1$,

$$H_0^a(\pi) \leq C_0(\pi)||\pi||^{-1/2},$$

$$H_1^a(\pi) \leq C_1(\pi)||\pi||^{-1}.$$

(3)

However, $C_0(\pi)$ and $C_1(\pi)$ are defined through complicated expressions, and can be quite large.

Starting points of (the long and complicated) proof: The unique bounded solution of the Stein equation with $h(k) = z^k$, where $z$ is a complex number such that $|z| \leq 1$, is

$$f^z(k) = e^{||\pi||\varphi_{\pi}(z)} \int_{\Gamma_{z,1}} w^{k-1} e^{-||\pi||\varphi_{\pi}(w)} dw,$$

where $\Gamma_{z,1}$ is any path between $z$ and 1. Moreover, for each $A \subset \mathbb{Z}_+$,

$$f_A(k) = \sum_{l \in A} \frac{1}{2\pi i} \int_{z = 1} z^{-l-1} f^z(k) dz$$

$$= \sum_{l \in A} \frac{1}{2\pi i} \int_0^1 t^{k-1} e^{||\pi||(1-\varphi_{\pi}(t))(\mu_t(l-k)-\mu_0(l))} dt,$$

where $\mu_t = \text{POIS}(\pi^t)$, and $\pi^t_k = \pi_k (1 - t^k)$ for each $k \in \mathbb{Z}_+$. Using these two representations of $f_A$, the bounds can be derived with much effort.
Preliminary bounds [Barbour and Xia 2000].

**Theorem.** Let \( f_{[m,\infty)} \) be the unique bounded solution of the Stein equation with \( h = I_{[m,\infty)} \). Define

\[
J_1(\pi) = \sup_{m \in \mathbb{Z}_+} \sup_{k \in \mathbb{Z}_+} |f_{[m,\infty)}(k+1) - f_{[m,\infty)}(k)|,
\]

\[
J_0(\pi) = \sup_{m \in \mathbb{Z}_+} \sup_{k \in \mathbb{Z}_+} |f_{[m,\infty)}(k)|.
\]

If

\[
i \pi_i - (i+1) \pi_{i+1} \geq 0 \quad \forall i \in \mathbb{Z}_+,
\]

then

\[
J_1(\pi) \leq \frac{1}{2} \wedge \frac{1}{\pi_1 + 1},
\]

\[
J_0(\pi) \leq 1 \wedge \sqrt{\frac{2}{e \pi_1}}. \tag{4}
\]

**Sketch proof:** As in the proof of (2), we use the probabilistic representation of the solution to the Stein equation, and get

\[
f_{[m,\infty)}(k+2) - f_{[m,\infty)}(k+1)
\]

\[
= \int_0^\infty e^{-2t} \left( \mathbb{P}(Z_t^{(0)} \geq m-2) - 2\mathbb{P}(Z_t^{(0)} \geq m-1) 
\right.
\]

\[
+ \mathbb{P}(Z_t^{(0)} \geq m) \big) dt
\]

\[
= \mathbb{E} \left( \int_0^\infty e^{-2t} (I\{Z_t^{(0)} = m-2\} - I\{Z_t^{(0)} = m-1\}) dt \right),
\]

where \( \{Z_t^{(0)}; t \in \mathbb{R}_+\} \) is a batch immigration-death process with generator \( \mathcal{A} \), starting at \( k \). The last expectation can be bounded using the strong Markov property. \( \square \)
Stein’s method for signed compound Poisson measure approximation [Barbour and Xia 1999; Barbour and Čekanavičius 2002].

Let $\pi$ be a signed measure on $(\mathbb{Z}, \mathcal{B}_{\mathbb{Z}})$ with finite support, and such that $\pi_0 = 0$. For each $\gamma \in \mathbb{Z}$, denote by $\mu_{\pi, \gamma}$ the (possibly signed) measure on $(\mathbb{Z}, \mathcal{B}_{\mathbb{Z}})$ with generating function

$$\varphi_{\pi, \gamma}(z) = z^\gamma \exp \left( - \sum_{k \in \mathbb{Z}} (1 - z^k) \pi_k \right).$$

Let $\chi$ be all functions $f : \mathbb{Z} \rightarrow \mathbb{R}$. Define a Stein operator $T_{\pi, \gamma} : \chi \rightarrow \chi$ by

$$(T_{\pi, \gamma} f)(k) = \sum_{i \in \mathbb{Z}} i \pi_i f(i + k) - (k - \gamma) f(k).$$

**Theorem.** Assume that

$$\lambda = \sum_{k \in \mathbb{Z}} k \pi_k > 0; \quad \theta = \frac{1}{\lambda} \sum_{k \in \mathbb{Z}} k(k - 1) \pi_k < \frac{1}{2}.$$ 

Then, for each bounded $h$ there exists a $f$ such that

$$f(i) = 0 \quad \forall i \leq 0;$$

$$(T_{\pi, 0} f) - (h(i) - \int_{\mathbb{Z}} h d \mu_{\pi, 0}) \leq \frac{2}{1 - 2\theta} \sum_{j < 0} \mu_{\pi, 0}(j) \|h\| \quad \forall i \geq 0;$$

$$\|f\| \leq \frac{2}{1 - 2\theta} \min(1, \frac{1}{\sqrt{\lambda}}) \|h\|;$$

$$\| \triangle f \| \leq \frac{2}{1 - 2\theta} \min(1, \frac{1}{\lambda}) \|h\|,$$

where $\triangle f(k) = f(k + 1) - f(k)$, and $\| \cdot \|$ denotes the supremum norm.
Proof: Define the bounded operator $U : \chi \rightarrow \chi$ by

$$(U f)(k) = (T_{\pi,0} f)(k) - \lambda f(k+1) + k f(k).$$

For each bounded $f$, let $S f$ denote the unique bounded solution $\bar{f}$ to

$$\lambda \bar{f}(k+1) - k \bar{f}(k) = h(k) - (U f)(k) - \int_{\mathbb{Z}_+} (h-U f) d\bar{\mu} \quad \forall k \geq 0;$$

$$\bar{f}(k) = 0 \quad \forall k \leq 0,$$

where $\bar{\mu} = \text{Po}(\lambda)$. Define $f_0 \equiv 0$, $f_n = S f_{n-1}$, and $g_n = f_n - f_{n-1}$. It follows that

$$\lambda g_n(k+1) - k g_n(k) = -(U g_{n-1})(k) + \int_{\mathbb{Z}_+} (U g_{n-1}) d\bar{\mu},$$

implying that

$$\|g_n\| \leq 2 \left(1 + \frac{1}{\sqrt{\lambda}}\right) \|U g_{n-1}\| \leq 2 \lambda\theta \left(1 + \frac{1}{\sqrt{\lambda}}\right) \|\Delta g_{n-1}\| \leq 2(2\theta)^n \left(1 + \frac{1}{\sqrt{\lambda}}\right) \|h\|.$$ 

Hence, the $f_n$ converge uniformly as $n \rightarrow \infty$ to $f$, which can be shown to have the required properties. \hfill \Box

Corollary. If $W$ is an integer valued random variable, and $\pi$ is a signed measure satisfying the conditions of the previous theorem, then

$$\|\mathcal{L}(W) - \mu_{\pi,\gamma}\| \leq \frac{2}{1 - 2\theta} \left(\sup_{f \in \mathcal{F}} |\mathbb{E}(T_{\pi,\gamma} f(W))| + \sum_{j < 0} |\mu_{\pi,0}(j)| + (1 + \sum_{j < 0} |\mu_{\pi,0}(j)|) \mathbb{P}(W < \gamma)\right),$$

where $\mathcal{F}$ is the set of bounded functions $f : \mathbb{Z} \rightarrow \mathbb{R}$ of the previous theorem corresponding to functions $h : \mathbb{Z} \rightarrow \mathbb{R}$ such that $\|h\| \leq 1$.  

33
**Theorem.** (Centred Poisson approximation.) Let \(X_1, \ldots, X_n\) be independent integer valued random variables with \(\mathbb{E}(X_i) = \beta_i\), \(\mathbb{V}(X_i) = \sigma_i^2\) and \(\mathbb{E}|X_i^3| < \infty\). Let \(W = \sum_{i=1}^n X_i\). Define \(\pi_i = 0\) for \(i \neq 1\), and

\[
\pi_1 = \sigma^2 + \delta; \quad \gamma = |m - \sigma^2|,
\]

where \(m = \mathbb{E}(W)\), \(\sigma^2 = \mathbb{V}(W)\), and \(\delta = (m - \sigma^2) - |m - \sigma^2|\). Let also \(W_i = W - X_i\), \(d = 2 \max_{i=1, \ldots, n} d_{TV}(\mathcal{L}(W_i), \mathcal{L}(W_i + 1))\), and

\[
\psi_i = \sigma_i^2 \mathbb{E}(X_i(X_i - 1)) + |\mu_i - \sigma_i^2| \mathbb{E}((X_i - 1)(X_i - 2))
\]

\[
+ \mathbb{E}|X_i(X_i - 1)(X_i - 2)|.
\]

Then,

\[
\|\mathcal{L}(W) - \mu_{\pi, \gamma}\| \leq \frac{2}{\sigma^2} \left( \frac{d}{2} \sum_{i=1}^n \psi_i + 1 + \delta \right).
\]

**Sketch proof:**

\[
\mathbb{E}(T_{\pi, \gamma} f(W)) = \mathbb{E}(\pi_1 f(W + 1) - (W - \gamma) f(W))
\]

\[
= \sum_{i=1}^n (\sigma_i^2 \mathbb{E}(f(W + 1)) + (\beta_i - \sigma_i^2) \mathbb{E}(f(W + 1))
\]

\[
- \mathbb{E}(X_i f(W + 1)) + \delta \mathbb{E}(\Delta f(W))
\]

Newton’s expansion gives

\[
f(W_i + l) = f(W_i + 1) + (l - 1) \Delta f(W_i + 1)
\]

\[
+ \begin{cases}
\sum_{s=1}^{l-2} (l - 1 - s) \Delta^2 f(W_i + s), & l \geq 3; \\
0, & 1 \leq l \leq 2; \\
\sum_{s=0}^{-l-s+1} (l - s + 1) \Delta^2 f(W_i - s), & l \leq 0,
\end{cases}
\]

from which we get

\[
\mathbb{E}(f(W_i + l)) - \mathbb{E}(f(W_i + 1)) - (l - 1) \mathbb{E}(\Delta f(W_i + 1))
\]

\[
\leq \frac{1}{2} (l - 1)(l - 2) \|f\|,
\]

since, for each \(j \in \mathbb{Z}\),

\[
\mathbb{E}(\Delta^2 f(W_i + j)) \leq 2 \|f\| d_{TV} (\mathcal{L}(W_i), \mathcal{L}(W_i + 1)).
\]

Hence,

\[
\mathbb{E}(f(W + 1)) = \sum_{j \in \mathbb{Z}} \mathbb{P}(X_i = j) \mathbb{E}(f(W_i + j + 1))
\]

\[
= \mathbb{E}(f(W_i + 1)) + \beta_i \mathbb{E}(\Delta f(W_i + 1)) + r_{i,1},
\]
where $|r_{i,1}| \leq \frac{1}{2} \mathbb{E}(X_i(X_i - 1))d\|f\|$, and similarly for the other terms. This gives the result. □

**Corollary.** In the above example, if $\sigma_i^2 \geq a > 0$, $\min\{\frac{1}{2}, 1 - d_{TV}(\mathcal{L}(X_i), \mathcal{L}(X_i))\} \geq b > 0$, and $\psi_i/\sigma_i^2 \leq c < \infty$, then

$$
\|\mathcal{L}(W) - \mu_{\pi,\gamma}\| \leq \frac{2c}{\sqrt{nb - \frac{1}{2}}} + \frac{2(1 + \delta)}{na}.
$$

If $X_i \sim \text{Be}(p)$ with $p \leq \frac{1}{2}$, then we may take $a = p(1 - p)$, $b = p$, and $c = 2p$.

If $np^2 < 1$, since $\mathbb{P}(W < \gamma) = 0$, the second term can be replaced by

$$
\frac{2\delta}{na} = \frac{2np^2}{na} = \frac{2p}{1 - p}.
$$


Let $\{Y_i; i \in \mathbb{Z}_+\}$ be nonnegative integer valued random variables with finite means, and let $W = \sum_{i=1}^{\infty} iY_i$. Then,

$$
\mathcal{L}(W) = \mathcal{L}\left(\int_{\mathbb{Z}_+} id\xi(i)\right),
$$

where $\xi$ is a point process on the space $\mathbb{Z}_+$ with locally finite expectation measure. A natural idea is to approximate $\mathcal{L}(W)$ with $\mathcal{L}\left(\int_{\mathbb{Z}_+} id\xi_0(i)\right)$, where $\xi_0$ is a Poisson point process on $\mathbb{Z}_+$ with expectation measure $\pi_k = \mathbb{E}(Y_k)$.
Since \( \int_{\mathbb{Z}^+} \text{id}\xi(i) \) is a measurable function on the space of locally finite counting measures on \( \mathbb{Z}^+ \),

\[
d_{TV}\left( \mathcal{L}\left( \int_{\mathbb{Z}^+} \text{id}\xi(i) \right), \mathcal{L}\left( \int_{\mathbb{Z}^+} \text{id}\xi_0(i) \right) \right) \leq d_{TV}\left( \mathcal{L}(\xi), \mathcal{L}(\xi_0) \right).
\]

The right-hand side can be bounded using Stein’s method, but these bounds are not in general as good as desired. Switching to the Wasserstein \( d_2 \)-distance does not help, since

\[
d_{TV}\left( \mathcal{L}\left( \int_{\mathbb{Z}^+} \text{id}\xi(i) \right), \mathcal{L}\left( \int_{\mathbb{Z}^+} \text{id}\xi_0(i) \right) \right) \not\leq d_2\left( \mathcal{L}(\xi), \mathcal{L}(\xi_0) \right).
\]

**Compound Poisson process approximation [Barbour and Månsso 2002].**

Let \( \{X_i; i \in \Gamma\} \) be nonnegative integer valued random variables, and let \( W = \sum_{i \in \Gamma} X_i \). Then \( \mathcal{L}(W) = \mathcal{L}(\Xi(\Gamma)) \), where \( \Xi \) is a point process on the space \( \Gamma \). We may approximate \( \mathcal{L}(W) \) with \( \mathcal{L}(\Xi_0(\Gamma)) \), where \( \Xi_0 \) is a compound Poisson point process on \( \Gamma \) with expectation measure \( \nu_i = \mathbb{E}(X_i) \), and \( \pi_k = \sum_{i \in \Gamma} \mathbb{P}(X_i = k) \). Then,

\[
d_{TV}(\mathcal{L}(\Xi(\Gamma)), \mathcal{L}(\Xi_0(\Gamma))) \leq d_2(\mathcal{L}(\Xi), \mathcal{L}(\Xi_0))
\]

The right-hand side can again be bounded using Stein’s method.
References:


