Math/Stat 235A, Fall 2025 (J. Gravner)
November 25, 2025

Notes on the Chen-Stein method for Poisson convergence

These notes are based on the book [**BHJ**], which is the most comprehensive text on this subject. Perhaps more accessible is the survey [**Ross**], which also gives applications of this method to convergence to Normal and other random variables.

1 Total variation distance

Let X and Y be integer-valued random variables. The *total variation distance* between two laws μ_X and μ_Y (or, with an abuse of terminology, between X and Y, or between X and μ_Y, etc.)

$$d_{\mathrm{TV}}(X,Y) = d_{\mathrm{TV}}(\mu_X,\mu_Y) = \sup_{A\subset\mathbb{Z}} |P(X\in A) - P(Y\in A)|.$$

Proposition 1.1. *We have*

$$d_{\mathrm{TV}}(X,Y) = \frac{1}{2}\sum_{k\in\mathbb{Z}} |P(X=k) - P(Y=k)|.$$

Proof. Denote the right-hand side by M. Use $|x| = 2x_+ - x = 2x_- + x$, to get

$$M = \sum_{k\in\mathbb{Z}} (P(X=k) - P(Y=k))_+ = \sum_{k\in\mathbb{Z}} (P(X=k) - P(Y=k))_-,$$

as the sum without the absolute value is 0. Let

$$a = \sum_{k\in A} (P(X=k) - P(Y=k))_+, \quad b = \sum_{k\in A} (P(X=k) - P(Y=k))_-.$$

Then $0 \le a, b \le M$, so $|P(X\in A) - P(Y\in A)| = |a-b| \le M$. This demonstrates the "$\le$" part, to prove the "$\ge$" part, take $A = \{k : P(X=k) > P(Y=k)\}$. □

From now on, let P_λ denote the Poisson probability function with parameter λ, that is $P_\lambda(k) = e^{-\lambda}\lambda^k/k!$ for $k \ge 0$, and $P_\lambda(A) = \sum_{k\in A} P_\lambda(k)$.

Proposition 1.2. *For any $\alpha, \lambda > 0$,*

$$d_{\mathrm{TV}}(P_\lambda, P_{\lambda+\alpha}) \le \alpha.$$

In fact, we also have the upper bound $\alpha/\sqrt{\lambda+\alpha}$. For the proof, see [**BHJ**]. We will only need the inequality in Proposition 1.2.

Proof. The law $P_{\lambda+\alpha}$ is obtained by the independent sum of two Poissons, with laws P_λ and P_α, as is easily seen by characteristic functions. Trivially, P_λ is the independent sum of a P_λ random variable and a random variable with law $\delta_0 = 1_{\{0\}}$. Therefore,

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} |P_{\lambda+\alpha}(k) - P_\lambda(k)| \\
&= \sum_k |\sum_\ell (P_\lambda(\ell)P_\alpha(k-\ell) - P_\lambda(\ell)\delta_0(k-\ell)| \\
&\leq \sum_\ell P_\lambda(\ell) \sum_k |P_\alpha(k-\ell) - \delta_0(k-\ell)| \\
&= \sum_\ell P_\lambda(\ell) \sum_k |P_\alpha(k) - \delta_0(k)| \\
&= \sum_k |P_\alpha(k) - \delta_0(k)| \\
&= 2d_{TV}(P_\alpha, \delta_0) = 2\sum_k (P_\alpha(k) - \delta_0(k))_- = 2(1-e^{-\alpha}) \leq 2\alpha.
\end{aligned}$$

□

2 The key estimate

Fix a set $A \subset \mathbb{Z}_+$. Then the *Stein's equation* for the function $f_A : \mathbb{Z}_+ \to \mathbb{R}$ is

$$\begin{aligned}
1_{\{k\in A\}} - P_\lambda(A) &= \lambda f_A(k+1) - kf_A(k), \quad k = 0, 1, \ldots \\
f_A(0) &= 0.
\end{aligned} \tag{2.1}$$

(Recall that P_λ is the Poisson probability.) Note that f_A is uniquely determined by (2.1).

Another useful interpretation is as follows. Let L be the operator on functions $f : \mathbb{Z}_+ \to \mathbb{R}$, given by $Lf(k) = \lambda f(k+1) - kf(k)$, $k \geq 0$. This is a linear operator. Furthermore, let $g_A(k) = 1_{\{k\in A\}} - P_\lambda(A)$. Then f_A is the unique function f that solves $Lf = g_A$ and satisfies $f(0) = 0$. Note that L is linear, and

$$\begin{aligned}
g_{A\cup B} &= g_A + g_B \quad \text{if } A \cap B = \emptyset, \\
g_{A^c} &= -g_A.
\end{aligned}$$

Then, by linearity of L,

$$\begin{aligned}
f_{A\cup B} &= f_A + f_B \quad \text{if } A \cap B = \emptyset, \\
f_{A^c} &= -f_A.
\end{aligned} \tag{2.2}$$

If fact, f_A can be computed — by induction we get, for any $k \geq 0$,

$$\begin{aligned}
f_A(k+1) &= \frac{1}{\lambda}1_{\{k\in A\}} + \frac{k}{\lambda^2}1_{\{k-1\in A\}} + \frac{k(k-1)}{\lambda^3}1_{\{k-2\in A\}} + \cdots + \frac{k!}{\lambda^{k+1}}1_{\{0\in A\}} \\
&\quad - \left(\frac{1}{\lambda} + \frac{k}{\lambda^2} + \frac{k(k-1)}{\lambda^3} + \ldots + \frac{k!}{\lambda^{k+1}}\right) P_\lambda(A) \\
&= \frac{k!}{\lambda^{k+1}} \left(\sum_{i=0}^k \frac{\lambda^i}{i!}1_{\{i\in A\}} - P_\lambda(A) \sum_{i=0}^k \frac{\lambda^i}{i!}\right).
\end{aligned}$$

Then, if we set $U_k = \{0, 1, \ldots, k\}$,

$$
\begin{aligned}
f_A(k+1) &= \frac{k!}{\lambda^{k+1}} e^{\lambda} [P_\lambda(A \cap U_k) - P_\lambda(A) P_\lambda(U_k)] \\
&= \frac{k!}{\lambda^{k+1}} e^{\lambda} [P_\lambda(A \cap U_k) - P_\lambda(A \cap U_k) P_\lambda(U_k) + P_\lambda(A \cap U_k) P_\lambda(U_k) - P_\lambda(A) P_\lambda(U_k)] \\
&= \frac{k!}{\lambda^{k+1}} e^{\lambda} [P_\lambda(A \cap U_k) P_\lambda(U_k^c) - P_\lambda(A \cap U_k^c) P_\lambda(U_k)].
\end{aligned} \tag{2.3}
$$

For $A \subset \mathbb{Z}_+$, write $A_n = A \cap U_n$ and $A'_n = A \setminus A_n$. Then it follows from the first line of (2.3) that, for every fixed k, $f_{A'_n}(k+1) \to 0$ as $n \to \infty$. Therefore, by the first line of (2.2),

$$
f_{A_n}(k+1) \to f_A(k+1) \quad \text{as } n \to \infty, \tag{2.4}
$$

pointwise in k.

For $f : \mathbb{Z}_+ \to \mathbb{R}$, let

$$
\Delta f = \sup\{|f(k+1) - f(k)| : k \geq 1\}.
$$

Lemma 2.1. *For any $A \subset \mathbb{Z}_+$,*

$$
\Delta f_A \leq \lambda^{-1}(1 - e^{-\lambda}) \leq \min(1, \lambda^{-1}).
$$

Proof. First we claim that it suffices to prove that

$$
f_A(k+1) - f_A(k) \leq \lambda^{-1}(1 - e^{-\lambda}), \tag{2.5}
$$

for all A and $k \geq 1$. Indeed, if this holds, the second line of (2.2) implies that

$$
f_A(k+1) - f_A(k) = -(f_{A^c}(k+1) - f_{A^c}(k)) \geq -\lambda^{-1}(1 - e^{-\lambda}),
$$

and thus $|f_A(k+1) - f_A(k)| \leq \lambda^{-1}(1 - e^{-\lambda})$.

To prove (2.5) we may, by (2.4), assume that A is finite. In this case, using the abbreviation $f_j = f_{\{j\}}$, (2.2) implies

$$
f_A = \sum_{j \in A} f_j. \tag{2.6}
$$

By the first line of (2.3),

$$
f_j(k+1) = \frac{k!}{\lambda^{k+1}} e^{\lambda} P_\lambda(j) [1_{\{j \leq k\}} - P_\lambda(U_k)]. \tag{2.7}
$$

If $k \geq j$, then, by (2.7),

$$
f_j(k+1) = \frac{1}{\lambda} P_\lambda(j) \sum_{i=1}^{\infty} \frac{\lambda^i}{(i+k)(i-1+k) \cdots (1+k)},
$$

which is positive and decreasing in k. If $k < j$, then, by (2.7),

$$
f_j(k+1) = -\frac{1}{\lambda} P_\lambda(j) \left(1 + \frac{k}{\lambda} + \frac{k(k-1)}{\lambda^2} + \cdots + \frac{k!}{\lambda^k}\right),
$$

which is negative and decreasing in k. The only $k \geq 1$ for which $f_j(k+1) - f_j(k) \geq 0$ then is $k = j$. For $j \geq 1$, by the third line of (2.3),

$$
\begin{aligned}
f_j(j+1) - f_j(j) &= \frac{j!}{\lambda^{j+1}} e^{\lambda} P_\lambda(j) P_\lambda(U_j^c) + \frac{(j-1)!}{\lambda^j} e^{\lambda} P_\lambda(j) P_\lambda(U_{j-1}) \\
&= \frac{1}{\lambda} \sum_{i=j+1}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} + \frac{1}{j} \sum_{i=0}^{j-1} \frac{\lambda^i}{i!} e^{-\lambda} \\
&= \frac{e^{-\lambda}}{\lambda} \left(\sum_{i=j+1}^{\infty} \frac{\lambda^i}{i!} + \sum_{i=1}^{j} \frac{\lambda^i}{i!} \cdot \frac{i}{j} \right) \\
&\leq \frac{e^{-\lambda}}{\lambda} \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} = \lambda^{-1}(1 - e^{-\lambda}).
\end{aligned}
\tag{2.8}
$$

For $j = 0$, we use that $f_0(k+1)$ is decreasing for all $k \geq 0$, so that $f_0(k+1) - f_0(k) \leq 0$ for $k \geq 1$. Thus we have, by (2.6), for every A and $k \geq 1$,

$$
f_A(k+1) - f_A(k) = \sum_{j \in A} (f_j(k+1) - f_j(k)) \leq f_k(k+1) - f_k(k) \leq \lambda^{-1}(1 - e^{-\lambda}),
$$

which proves (2.5) and ends the proof. □

It is easy to show that, if W is a Poisson random variable with $EW = \lambda$, and f is any function on $\mathbb{Z}_+$ that grows at most exponentially (that is, $|f(x)| \leq \exp(Cx)$ for some constant C), $E[\lambda f(W+1) - W f(W)] = 0$. It is reasonable to expect that, if a random variable W with values in $\mathbb{Z}_+$ is such that this equality approximately holds for a class of functions f, then W is approximately Poisson. Indeed, the essence of Chen-Stein method is that an estimate

$$
E[\lambda f_A(W+1) - W f_A(W)] \leq \alpha, \tag{2.9}
$$

where α does not depend on A, immediately implies (as we can apply it to $f_{A^c} = -f_A$) the same bound for the absolute value and hence for the total variation distance from P_λ: by (2.1), $1_{\{W \in A\}} - P_\lambda(A) = \lambda f_A(W+1) - W f_A(W)$ and then

$$
d_{\mathrm{TV}}(W, P_\lambda) = \sup_A |P(W \in A\} - P_\lambda(A)| = \sup_A |E[\lambda f_A(W+1) - W f_A(W)]| \leq \alpha.
$$

To get (2.9) using the Lemma, one needs to produce Δf_A as a factor in an upper bound for $E[\lambda f_A(W+1) - W f_A(W)]$. This can be done in many cases when W is a sum of mildly dependent indicators.

3 The theorems

Suppose that I_i, $i \in \Gamma$ are indicators, where Γ is a finite index set (not necessarily a set of integers). We will use the following notation throughout: $p_i = E(I_i)$, $W = \sum_{i \in \Gamma} I_i$, $W_i = W - I_i$, and $\lambda = EW = \sum_{i \in \Gamma} p_i$.

Our first theorem, on independent indicators, is originally due to L. Le Cam (who proved it by a different method).

Theorem 3.1. *If I_i are independent, then*

$$d_{\mathrm{TV}}(W, P_\lambda) \le \min(1, \lambda^{-1}) \sum_{i\in\Gamma} p_i^2.$$

Proof. As I_i are independent, W_i is independent of I_i, and then

$$\begin{aligned}
E[\lambda f_A(W+1) - W f_A(W)] &= \sum_{i\in\Gamma} [p_i E f_A(W+1) - E(I_i f_A(W))] \\
&= \sum_{i\in\Gamma} [p_i E f_A(W+1) - E(I_i f_A(W_i+1))] \\
&= \sum_{i\in\Gamma} [p_i E f_A(W+1) - p_i E(f_A(W_i+1))] \\
&= \sum_{i\in\Gamma} p_i E\left((f_A(W+1) - f_A(W_i+1)) I_i\right),
\end{aligned}$$

the last line because $W+1 = W_i+1$ on $\{I_i = 0\}$. On $\{I_i = 1\}$, however, $W+1 = (W_i+1)+1$, therefore the above expression is bounded above by $\Delta f_A \cdot \sum_{i\in\Gamma} p_i^2$. Lemma 2.1 finishes the proof. □

The first generalization of Theorem 1 is in the direction of *local* dependence. Assume that each indicator I_i has a set of indices $\Gamma_i \subset \Gamma$ so that $i \notin \Gamma_i$ and so that the vector $(I_j : j \notin \Gamma_i \cup \{i\})$ is independent of I_i. Thus Γ_i is the "neighborhood of dependence" for i.

Theorem 3.2.

$$d_{\mathrm{TV}}(W, P_\lambda) \le \min(1, \lambda^{-1}) \left[\sum_{i\in\Gamma} p_i^2 + \sum_{i\in\Gamma, j\in\Gamma_i} (p_i p_j + E(I_i I_j)) \right].$$

Proof. Let $Z_i = \sum_{j\in\Gamma_i} I_j$ and $Y_i = W - I_i - Z_i$. Thus $W = I_i + Z_i + Y_i$, and Y_i is independent of I_i. This time we write

$$\begin{aligned}
&E[\lambda f_A(W+1) - W f_A(W)] \\
&= \sum_{i\in\Gamma} [p_i E f_A(W+1) - E(I_i f_A(W_i+1))] \\
&= \sum_{i\in\Gamma} [p_i E(f_A(W+1) - f_A(Y_i+1)) - E(I_i(f_A(Y_i+Z_i+1) - f_A(Y_i+1)))],
\end{aligned}$$

where the last inequality holds because $E(I_i f_A(Y_i+1)) = p_i E f_A(Y_i+1)$. Now, by telescoping,

$$\begin{aligned}
f_A(W+1) - f_A(Y_i+1) &\le \Delta f_A \cdot (Z_i + I_i), \\
|f_A(Y_i+Z_i+1) - f_A(Y_i+1)| &\le \Delta f_A \cdot Z_i,
\end{aligned}$$

and so

$$\begin{aligned}
E[\lambda f_A(W+1) - W f_A(W)] &\le \Delta f_A \cdot \sum_{i\in\Gamma} [p_i(EZ_i + p_i) + E(I_i Z_i)] \\
&\le \min(1, \lambda^{-1}) \sum_{i\in\Gamma} [p_i^2 + p_i E Z_i + E(I_i Z_i)],
\end{aligned}$$

by Lemma 2.1, and this is equivalent to the claim. □

The second approach is *coupling*. To describe how this works, we *fix an* $i \in \Gamma$. (Usually, there is enough symmetry so that all i play the same role, but that is not necessary.) Then we require that I_j and some random variables J_{ji}, $j \neq i$ are constructed on the same probability space so that the following equality in distribution between the two vectors holds:

$$(J_{ji})_{j\neq i} \stackrel{d}{=} (I_j)_{j\neq i} \mid I_i = 1. \tag{3.1}$$

For the method to be successful, the new random variables J_{ji} should not to be very far from I_j, otherwise any coupling (say, the independent one) would do.

Theorem 3.3. *Under any coupling which satisfies* (3.1),

$$d_{\mathrm{TV}}(W, P_\lambda) \leq \min(1, \lambda^{-1}) \left[\sum_{i\in\Gamma} p_i^2 + \sum_{i,j,j\neq i} p_i E|J_{ji} - I_j| \right].$$

Proof. Let $V_i = \sum_{j\neq i} J_{ji}$. Then

$$V_i + 1 \stackrel{d}{=} W \mid I_i = 1.$$

Now,

$$\begin{aligned}
&E[\lambda f_A(W+1) - W f_A(W)] \\
&= \sum_{i\in\Gamma} [p_i E f_A(W+1) - E(I_i f_A(W))] \\
&= \sum_{i\in\Gamma} p_i [E f_A(W+1) - E(f_A(W) \mid I_i = 1)] \\
&= \sum_{i\in\Gamma} p_i [E f_A(W+1) - E(f_A(V_i + 1))] \\
&\leq \Delta f_A \cdot \sum_{i\in\Gamma} p_i E|W - V_i| \\
&\leq \min(1, \lambda^{-1}) \left[\sum_{i\in\Gamma} p_i E(I_i + \sum_{j,j\neq i} |I_j - J_{ji}|) \right],
\end{aligned}$$

which is equivalent to the claim. □

Often, the event $\{I_i = 1\}$ helps the events $\{I_j = 1\}$ to happen, or perhaps it hinders them. Either circumstance simplifies Theorem 3.3 quite a bit. A formal definition is as follows. If a coupling exists so that $J_{ji} \geq I_j$ (resp. $J_{ji} \leq I_j$) for all i and $j \neq i$, then I_i are *positively* (resp. *negatively*) *related*.

Note that positively related indicators are positively correlated:

$$P(I_j = 1) = E(I_j) \leq E(J_{ji}) = P(I_j = 1 \mid I_i = 1) = P(I_i = 1, I_j = 1)/P(I_i = 1).$$

The opposite implication does not hold, as positive relatedness is about more than pairs of indicators.

Corollary 3.4. *(1) In the positively related case*

$$d_{\mathrm{TV}}(W, P_\lambda) \le \min(1, \lambda^{-1}) \left[2 \sum_{i \in \Gamma} p_i^2 + \sum_{i,j,i \ne j} E(I_i I_j) - \lambda^2 \right]$$
$$= \min(1, \lambda^{-1}) \left[2 \sum_{i \in \Gamma} p_i^2 + \operatorname{Var} W - \lambda \right].$$

(2) In the negatively related case

$$d_{\mathrm{TV}}(W, P_\lambda) \le \min(1, \lambda^{-1}) \left[\lambda^2 - \sum_{i,j,i \ne j} E(I_i I_j) \right]$$
$$= \min(1, \lambda^{-1}) \left[\lambda - \operatorname{Var} W \right].$$

Note that the indicators J_{ji} do not explicitly appear in the Corollary. It is therefore enough to know that they exist without an explicit construction. Such existence theorems do exist for many cases (see [**BHJ**]). Note also that for negatively related indicators, for W to be close to a Poisson random variable, it is enough that EW be close to $\operatorname{Var} W$, something that almost looks too good to be true!

Proof of Corollary 3.4. In the positively related case,

$$p_i E|J_{ji} - I_j| = p_i E(J_{ji} - I_j) = E(I_i I_j) - p_i p_j,$$

while in the negatively related case,

$$p_i E|J_{ji} - I_j| = p_i E(I_j - J_{ji}) = p_i p_j - E(I_i I_j).$$

In either case,

$$E(W^2) = \sum_{i,j} E(I_i I_j) = \lambda + \sum_{i,j,i \ne j} E(I_i I_j)$$

and

$$(EW)^2 = \lambda^2 = \sum_{i,j} p_i p_j = \sum_i p_i^2 + \sum_{i,j,i \ne j} p_i p_j,$$

and both results follows after some algebra. □

The next corollary covers the case when positive relatedness is violated locally. A similar result of course holds for negative relatedness.

Corollary 3.5. *Assume that Γ_i^{unr} are sets of indices such that $i \notin \Gamma_i^{\mathrm{unr}}$ and such that $i \ne j \notin \Gamma_i^{\mathrm{unr}}$ implies $J_{ji} \ge I_j$. Then*

$$d_{\mathrm{TV}}(W, P_\lambda) \le \min(1, \lambda^{-1}) \left[2 \sum_{i \in \Gamma} p_i^2 - \lambda^2 + \sum_{i,j,i \ne j, j \notin \Gamma_i^{\mathrm{unr}}} E(I_i I_j) + \sum_{i,j,j \in \Gamma_i^{\mathrm{unr}}} (2 p_i p_j + E(I_i I_j)) \right].$$

Proof. For $j \in \Gamma_i^{\text{unr}}$ simply estimate $p_i E|J_{ji} - I_j| \le p_i E(J_{ji} + I_j) = E(I_i I_j) + p_i p_j$ to get

$$\begin{aligned}
&\sum_{i \in \Gamma} \left(p_i^2 + p_i \sum_{j \neq i} E|J_{ji} - I_j| \right) \\
&\le \sum_i p_i^2 + \sum_{i, i \neq j \notin \Gamma_i^n} (E(I_i I_j) - p_i p_j) + \sum_{i, i \neq j \in \Gamma_i^n} (E(I_i I_j) + p_i p_j) \\
&= \sum_i p_i^2 + \sum_{i, i \neq j \notin \Gamma_i^n} E(I_i I_j) - \sum_{i,j} p_i p_j + \sum_i p_i^2 + \sum_{i, i \neq j \in \Gamma_i^n} p_i p_j \\
&\qquad + \sum_{i, i \neq j \in \Gamma_i^n} (E(I_i I_j) + p_i p_j).
\end{aligned}$$

This finishes the proof, as $\sum_{i,j} p_i p_j = \lambda^2$. □ □

4 Examples

Example 4.1 (*Binomial distribution*)**.** It follows immediately from Theorem 3.1 that

$$d_{\text{TV}}(\text{Binomial}(n, p), P_{np}) \le \min(p, np^2).$$

Example 4.2 (*Records*)**.** Here, I_i, $i = 1, \ldots, n$, are independent with $p_i = 1/i$. Then $\lambda = \lambda_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}$ and, by Theorem 3.1,

$$d_{\text{TV}}(W, P_\lambda) \le \min(1, \lambda^{-1}) \sum_{i=1}^n p_i^2 \le \frac{\pi^2}{6\lambda} = \mathcal{O}\left(\frac{1}{\log n}\right).$$

Furthermore, we also get the following limit theorem, assuming Z is a r.v. with $\mu_Z = P_\lambda$, and N a standard Normal r.v.,

$$P\left(\frac{W - \lambda_n}{\sqrt{\lambda_n}} \le x\right) = P\left(\frac{Z - \lambda_n}{\sqrt{\lambda_n}} \le x\right) + \mathcal{O}\left(\frac{1}{\log n}\right) \to P(N \le x)$$

as $n \to \infty$, by the CLT for the Poisson law.

Example 4.3 (*Birthday Problem*)**.** Fix an integer $a \ge 2$ throughout. Sample, with replacement, k times (i.e., choose k people) from a set of n birthdays. Let Γ be the set of all subsets of size a of k people, I_i the indicator of the event that all members of i have the same birthday and $W = W_{n,k} = \sum_{i \in \Gamma} I_i$. Note that $|\Gamma| = \binom{k}{a}$. Then

$$\lambda = \lambda_n = EW = \binom{k}{a} n^{-a-1} = \frac{k^a}{a!\, n^{a-1}} + \mathcal{O}\left(\frac{k^{a-1}}{n^{a-1}}\right),$$

if k is large. Take $k = k_n = c \cdot n^{(a-1)/a}$. Thus $k^a / n^{a-1} = c^a$ and

$$\lambda = \frac{c^a}{a!} + \mathcal{O}\left(\frac{1}{n^{(a-1)/a}}\right).$$

Disjoint sets of people share birthdays independently, so we seek to apply Theorem 3.2, with $\Gamma_i = \{j : i \cap j \neq \emptyset\} \setminus \{i\}$. We have

$$\sum_i p_i^2 = \binom{k}{a} n^{-2(a-1)} = \mathcal{O}\left(\frac{k^a}{n^{2(a-1)}}\right) = \mathcal{O}\left(\frac{1}{n^{a-1}}\right),$$

and

$$\sum_{i\in\Gamma, j\in\Gamma_i} p_i p_j = \binom{k}{a} \sum_{\ell=1}^{a-1} \binom{a}{\ell}\binom{k-a}{a-\ell} n^{-2(a-1)} = \mathcal{O}\left(\frac{k^a k^{a-1}}{n^{2(a-1)}}\right) = \mathcal{O}\left(\frac{1}{n^{(a-1)/a}}\right),$$

and

$$\begin{aligned}\sum_{i\in\Gamma, j\in\Gamma_i} E(I_i I_j) &= \binom{k}{a} \sum_{\ell=1}^{a-1} \binom{a}{\ell}\binom{k-a}{a-\ell} n^{-(2a-\ell-1)} \\ &= \mathcal{O}\left(k^a \sum_{\ell=1}^{a-1} k^{a-\ell} n^{-(2a-\ell-1)}\right) \\ &= \mathcal{O}\left(\sum_{\ell=1}^{a-1} n^{\ell/a-1}\right) = \mathcal{O}\left(\frac{1}{n^{1/a}}\right).\end{aligned}$$

So Theorem 3.2, together with Proposition 1.2, implies that

$$d_{\mathrm{TV}}(W, P_{c^a/a!}) = \mathcal{O}\left(\frac{1}{n^{1/a}}\right).$$

Now sample people indefinitely, and let $T = T_n$ be the time (measured in the number of people) when a people share a birthday. We have $P(T > k) = P(W = 0) \to \exp(-c^a/a!)$ as $n \to \infty$, and so $T/n^{(a-1)/a}$ converges in distribution to a nontrivial random variable.

Example 4.4 *(Runs)***.** Build a vector $(X_1, \ldots, X_n)$ in which each component is independently 1 with probability p and 0 with probability $1-p$. Declare $X_0 = 0$. Think of p as fixed and n as large. A *run* at i of size at least t is the pattern $0111\ldots1$, with t 1's, the leftmost of which is at i. The initial 0 is important — it is used for "declumping," i.e., to make sure long rungs are only counted once. How large should t be so that the number of such runs is approximately Poisson?

Let I_i indicate the event that there is a run of size at least t at i, $i = 1, \ldots, n-t+1$, and $W = W_{n,t} = \sum_i I_i$. So

$$EW = p^t + (n-t)(1-p)p^t = np^t(1-p) + (1 + t(1-p))p^t.$$

Take $t = t_n = -\log n/\log p + c$, where $c = c_n$ is bounded. (As t must be an integer, we cannot assume that c is a constant.) Then $p^t = p^c/n$ and

$$\lambda = EW = p^c(1-p) + \mathcal{O}\left(\frac{\log n}{n}\right).$$

For every $i = 1, \ldots, n-t+1$, let $\Gamma_i = [i-t, i+t] \cap [1, n]$. It is easy to check that, when $j \notin \Gamma_i$, I_j is independent of I_i. Also,

$$\sum_{i\in\Gamma, j\in\Gamma_i} E(I_i I_j) = 0$$

and

$$\sum_{i\in\Gamma, j\in\Gamma_i} p_i p_j \le n(2t+1)p^{2t} = \mathcal{O}\left(\frac{\log n}{n}\right).$$

Therefore,

$$d_{\mathrm{TV}}(W, P_{p^c(1-p)}) = \mathcal{O}\left(\frac{\log n}{n}\right).$$

It follows that $P(\text{no contiguous interval of 1's of size} \ge t) = P(W=0) = e^{p^c(1-p)} + \mathcal{O}\left(\frac{\log n}{n}\right)$.

Example 4.5 (*Isolated vertices in random graphs*)**.** Build a random graph on $\{1,\ldots,n\}$, with an (undirected) edge between each pair $\{i,j\}$ independently with probability p. The number of edges is thus Binomial with parameters $\binom{n}{2}$ and p. Let I_i indicate the event that the vertex i is isolated (not connected to any other vertex), so that $p_i = (1-p)^{n-1}$. Then $\lambda = \lambda_n = EW = n(1-p)^{n-1}$, and the question is what $p = p_n$ should be so that $P(W=0)$ converges neither to 0 nor to 1. If we take

$$p = \frac{\log n}{n} + \frac{c}{n},$$

then

$$\lambda = e^{-c} + \mathcal{O}\left(\frac{\log^2 n}{n}\right).$$

Clearly I_i and I_j are dependent for all i and j, so the local approach does not work. However, this is one of the simplest instances in which coupling works. In fact, J_{ji} can be defined on the original probability space: let J_{ji} indicate the event that j is isolated after all the edges (if any) emanating from i are removed. The conditional distribution property (3.1) is then clearly satisfied: the event that $I_i = 1$ is exactly the event that the $n-1$ specific edges emanating from i are missing. Clearly, $I_j \le J_{ji}$, so, by Corollary 3.4 (1), we need to estimate

$$\sum_i p_i^2 = n(1-p)^{2(n-1)} = \frac{\lambda^2}{n} = \mathcal{O}\left(\frac{1}{n}\right),$$

and

$$\sum_{i,j,i\neq j} E(I_i I_j) = n(n-1)(1-p)^{2n-3} = \lambda^2(1-p)^{-1} - \frac{\lambda^2}{n}(1-p)$$
$$= \lambda^2 + \mathcal{O}\left(p\lambda^2 + \frac{\lambda^2}{n}\right) = \lambda^2 + \mathcal{O}\left(\frac{\log n}{n}\right).$$

This proves that

$$d_{\mathrm{TV}}(W, P_{e^{-c}}) = \mathcal{O}\left(\frac{\log^2 n}{n}\right).$$

and thus that

$$P(W=0) = e^{-e^{-c}} + \mathcal{O}\left(\frac{\log^2 n}{n}\right).$$

A well known theorem for random graphs states that no matter how p varies with n, $P(W=0, \text{graph not connected}) \to 0$ as $n \to \infty$ (see [**JLR**]). So the last formula above also gives us a probability estimate for connectivity of a random graph.

This method also gives useful estimates for other values of p. For example, if $p = cn^{-1}\log n$, $c < 1$, then $\lambda = n^{1-c} + \mathcal{O}\left(n^{-c}\log^2 n\right)$,

$$\sum_i p_i^2 = \mathcal{O}\left(n^{1-2c}\right),$$

and

$$\sum_{i,j,i\neq j} E(I_i I_j) = \lambda^2 + \mathcal{O}\left(n^{1-2c}\log n\right).$$

It follows that

$$d_{\mathrm{TV}}(W, P_\lambda) = \mathcal{O}\left(\frac{1}{\lambda}\cdot n^{1-2c}\log n\right) = \mathcal{O}\left(\frac{\log n}{n^c}\right).$$

and consequently

$$d_{\mathrm{TV}}(W, P_{n^{1-c}}) = \mathcal{O}\left(\frac{\log^2 n}{n^c}\right).$$

It follows that $n^{-(1-c)/2}(W - n^{1-c}) \xrightarrow{d} N(0,1)$, by the CLT for Poisson.

Example 4.6 (*Fixed points in random permutations*)**.** Let $(\pi(i))_{i=1}^n$ be a random permutation, $I_i = 1_{\{\pi(i)=i\}}$ and

$$J_{ji} = \begin{cases} I_j & \text{if } \pi(i) = i, \\ 1_{\{j \text{ fixed after } i \text{ and } \pi(i) \text{ are interchanged}\}} & \text{otherwise.} \end{cases}$$

To be more precise and verify (3.1), we assume that $i = n$ (which we may, by symmetry) and that $(\Omega, \mathcal{F}, P)$ is any probability space on which we have a uniform random permutation σ of $\{1, \ldots, n-1\}$ and a uniform random variable $L \in \{1, \ldots, n\}$, so that σ and L are independent. Then we define a uniform random permutation π as follows: we let $\pi(n) = L$ and $\pi(\ell) = \sigma(\ell)$ if $\sigma(\ell) < L$ and $\pi(\ell) = \sigma(\ell) + 1$ otherwise. (That is, we keep the order imposed by σ.) Furthermore, we let π' be the permutation that agrees with π except for the interchange:

$$\pi'(\ell) = \begin{cases} \pi(\ell) & \text{if } \ell \neq n, \ell \neq \pi^{-1}(n), \\ n & \text{if } \ell = n, \\ L & \text{if } \ell = \pi^{-1}(n). \end{cases}$$

Finally, we define $I_j = 1_{\{\pi(j)=j\}}$ and $J_{ji} = 1_{\{\pi'(j)=j\}}$.

To verify (3.1), we need to check that the interchange preserves uniform random order on $\{1, \ldots, n-1\}$. What precisely does it do? Let ζ_L be the permutation on $\{1, \ldots, n-1\}$ that maps the segment $(L, L+1, \ldots, n-1)$ cyclically into $(L+1, \ldots, n-1, L)$ and is identity otherwise. Then the interchange results in $\zeta_L \circ \sigma$. Note that ζ_L is a random permutation that is independent of σ. Composing σ with any deterministic permutation preserves the uniform order, and therefore composing with an independent random permutation also does.

The rest is easy. First, $J_{ji} \geq I_j$, we already know that $EW = \operatorname{Var} W = 1$, and $\sum_i p_i^2 = 1/n$. It follows that

$$d_{\mathrm{TV}}(W, P_1) \leq \frac{2}{n},$$

which looks good, but is in fact very far from a realistic estimate. It is not very difficult to do explicit calculations, which show that in this case

$$d_{\mathrm{TV}}(W, P_1) = \mathcal{O}\left(\frac{2^n}{n!}\right),$$

so there is practically no difference between μ_W and P_1 for large n. The book [**BHJ**] has an entire chapter on lower bounds for d_{TV}, and the comparison with the Chen-Stein upper bounds.

Another example in this vein are "approximate fixed points." Let I_i indicate the event that $|\pi(i) - i| \leq 1$. In this case

$$J_{ji} = \begin{cases} I_j, & \text{if } I_i = 1, \\ 1_{\{j \text{ fixed after } i \text{ and a random number among } \pi(i-1), \pi(i), \pi(i+1) \text{ are interchanged}\}}, & \text{otherwise.} \end{cases}$$

(Omit $\pi(i-1)$ above if $i = 1$ and $\pi(i+1)$ if $i = n$.) Checking (3.1) is very similar to the above case. Then $J_{ji} \geq I_j$ if $|j - i| \geq 3$. Also, we have $\lambda = EW = 3 + \mathcal{O}(1/n)$, $\sum_i p_i^2 = \mathcal{O}(1/n)$,

$$\sum_{i,j,j \in \Gamma_i^{\mathrm{unr}}} (2p_i p_j + E(I_i I_j)) = \mathcal{O}\left(n \cdot \frac{1}{n^2}\right) = \mathcal{O}\left(\frac{1}{n}\right),$$

and

$$\sum_{i,j,j \notin \Gamma_i^{\mathrm{unr}}} E(I_i I_j) = \sum_{i,j,|j-i| \geq 3} \frac{9}{n(n-1)} + \mathcal{O}\left(\frac{1}{n}\right) = 9 + \mathcal{O}\left(\frac{1}{n}\right).$$

Therefore, in this case we also have

$$d_{\mathrm{TV}}(W, P_3) = \mathcal{O}\left(\frac{1}{n}\right).$$

Example 4.7 (*Coupon collector*)**.** In this example we have k coupons, chosen independently at random from $\{1, \ldots, n\}$. Let I_i be the indicator of the event that i is missing from the collection. We will see that this is a negatively related case. Before we specify the coupling, we note that, conditional on $\{I_i = 1\}$, the k coupons are independent and uniform on $\{1, \ldots, n\} \setminus \{i\}$. We can achieve that from our original choice of k coupons by replacing every i by a random coupon in $\{1, \ldots, n\} \setminus \{i\}$.

To precisely construct the coupling, let $(\Omega, \mathcal{F}, P)$ be any probability space with with independent random variables $X_1, \ldots, X_k$ and $Y_1, \ldots, Y_k$, X_ℓ uniform on $\{1, \ldots, n\}$, and Y_ℓ uniform on $\{1, \ldots, n\} \setminus \{i\}$. Then the random variables Z_ℓ are obtained by the replacement described above:

$$Z_\ell = \begin{cases} X_\ell & \text{if } X_\ell \neq i, \\ Y_\ell & \text{if } X_\ell = i. \end{cases}$$

Finally, define $I_j = 1_{\{j \notin \{X_1, \ldots, X_k\}\}}$ and $J_{ji} = 1_{\{j \notin \{Z_1, \ldots, Z_k\}\}}$. The distributional requirement (3.1) is then satisfied (in fact, we have $(Z_\ell) \stackrel{d}{=} (Y_\ell)$), and clearly $J_{ji} \leq I_j$.

Take $k = n \log n + cn$. Then

$$\lambda = EW = n\left(1 - \frac{1}{n}\right)^k = ne^{-k^2/n}\left(1 + \mathcal{O}\left(\frac{k}{n^2}\right)\right) = e^{-c} + \mathcal{O}\left(\frac{\log n}{n}\right)$$

and

$$\sum_{i,j,j\neq i} E(I_iI_j) = n(n-1)\left(1-\frac{2}{n}\right)^k = n(n-1)e^{-2k/n}\left(1+\mathcal{O}\left(\frac{k}{n^2}\right)\right)$$

$$= \left(1-\frac{1}{n}\right)e^{-2c}\left(1+\mathcal{O}\left(\frac{\log n}{n}\right)\right) = e^{-2c}+\mathcal{O}\left(\frac{\log n}{n}\right).$$

It follows that

$$d_{\mathrm{TV}}(W, P_{e^{-c}}) = \mathcal{O}\left(\frac{\log n}{n}\right).$$

So in particular if T_n is the first time the collector has full collection,

$$P(T_n \le k) = P(W=0) = e^{-e^{-c}} + \mathcal{O}\left(\frac{\log n}{n}\right),$$

so $n^{-1}(T_n - n\log n)$ converges in distribution.

A very similar argument shows that with I_i indicating the event that the number of representatives of i is at most 1, then the correct scaling for Poisson limit is $k = n\log n + n\log\log n + cn$.

Example 4.8 (*Hypergeometric distribution*)**.** Assume N, n, m are positive integers, with $m \le N$ and $n \le N$. Arrange m 1's and $N-m$ 0's at random to form a random N-vector X, and let W be the number of 1s among the first n positions. Then W is the sum, over $i \in \Gamma = \{1,\dots,n\}$, of the indicators $I_i = 1_{\{X_i=1\}}$. Moreover, W has hypergeometric distribution,

$$P(W=j) = \frac{\binom{m}{j}\binom{N-m}{n-j}}{\binom{N}{n}}$$

with

$$\lambda = EW = \frac{nm}{N}, \quad \operatorname{Var} W = \frac{mn(N-n)(N-m)}{N^2(N-1)}.$$

(The variance computation is straightforward, but tedious.) We can see that I_i are negatively related by defining

$$J_{ji} = \begin{cases} I_j & \text{if } I_i = 1, \\ 1_{\{\text{1 at position } j \text{ after a randomly chosen 1 has been switched to 0}\}} & \text{otherwise.} \end{cases}$$

That is, define the vector Y as follows: choose a random position L, chosen uniformly among all positions ℓ such that $X_\ell = 1$, then switch X_i and X_L to get

$$Y_\ell = \begin{cases} X_\ell & \text{if } \ell \neq i, \ell \neq L, \\ X_i & \text{if } \ell = L, \\ X_L & \text{if } \ell = i, \end{cases}$$

and then let $I_j = 1_{\{X_j=1\}}$ and $J_{ji} = 1_{\{Y_j=1\}}$. This ensures (3.1), and the negative relation $J_{ji} \le I_j$ clearly holds. Therefore,

$$d_{\mathrm{TV}}(W, P_\lambda) \le \min(1, \lambda^{-1}) \cdot (\lambda - \operatorname{Var} W)$$

$$\le \min(\lambda, 1) \cdot \frac{N}{N-1}\left(\frac{n}{N} + \frac{m}{N} - \frac{nm}{N^2} - \frac{1}{N}\right),$$

and thus this distance approaches 0 as $N \to \infty$ if n and m are both $o(N)$.

References

[**BHJ**] A. D. Barbour, L. Holst, S. Janson, “Poisson Approximation,” Clarendon Press, 1992.

[**JLR**] S. Janson, T. Łuczak, A. Ruciński, “Random Graphs,” Wiley, 2000.

[**Ross**] N. Ross, *Fundamentals of Stein’s method*, Probability Surveys 8 (2001), 210—293.