Math/Stat 235, Fall 2005 (J. Gravner).

NOTES ON THE CHEN-STEIN METHOD

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, X and Y, or X and μ_Y, etc.)

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

Proposition 1.

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

Proof. Denote the RHS 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 $a = \sum_{k\in A}(P(X = k) - P(Y = k))_+$, $b = \sum_{k\in A}(P(X = k) - P(Y = k))_-$. 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 get the "$\ge$" one, take $A = \{k : P(X \in A) > P(Y \in A)\}$. $\square$

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 2. *For any* $\alpha, \lambda > 0$,

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

In fact, we also have the upper bound $\alpha/\sqrt{\lambda + \alpha}$. For the proof, see the book "Poisson Approximation," by A. D. Barbour, Lars Holst, and Svante Janson, on which these notes are based. We will not use this inequality.

Proof. Recall that $P_{\lambda+\alpha}$ is the law of the indpendent sum of two Poissons, with laws P_λ and P_α. Trivially, P_λ is the independent sum of a P_λ r.v. and a random variable with law $\delta_0 = 1_{\{0\}}$. Therefore

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

2. The key estimate.

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

$$1_{\{k\in A\}} - P_\lambda(A) = \lambda f_A(k+1) - k f_A(k), \quad f_A(0) = 0, \tag{1}$$

where P_λ is the Poisson probability. Note that f_A is uniquely determined by (1).

In other words, if L is the operator on functions $f : \mathbb{Z} \to \mathcal{R}$, given by $Lf(k) = \lambda f(k+1) - kf(k)$, $k \geq 0$, and $g_A(k) = 1_{\{k\in A\}} - P_\lambda(A)$, then f_A is the unique function that solves $Lf = g_A$, $f(0) = 0$. Note right away that L is linear, and

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

These properties imply

$$\begin{aligned} f_{A\cup B} &= f_A + f_B \quad \text{ if } A \cap B = \emptyset, \\ f_{A^c} &= -f_A. \end{aligned} \tag{2}$$

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

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

If fact f_A can be computed – by induction we get

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

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

$$\begin{aligned} f_A(k+1) &= \frac{k!}{\lambda^{k+1}} e^\lambda [P_\lambda(A \cap U_k) - P_\lambda(A) P_\lambda(U_k)] \\ &= \frac{k!}{\lambda^{k+1}} e^\lambda [P_\lambda(A \cap U_k) - P_\lambda(A \cap U_k) P_\lambda(U_k) + P_\lambda(A \cap U_k) P_\lambda(U_k) - P_\lambda(A) P_\lambda(U_k)] \\ &= \frac{k!}{\lambda^{k+1}} e^\lambda [P_\lambda(A \cap U_k) P_\lambda(U_k^c) + P_\lambda(A \cap U_k^c) P_\lambda(U_k)]. \end{aligned} \tag{3}$$

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 (3) that, for every fixed k, $f_{A'_n}(k+1) \to 0$ as $n \to \infty$. Therefore, by the first line of (2),

$$f_{A_n}(k+1) \to f_A(k+1) \quad \text{ as } n \to \infty, \tag{4}$$

pointwise in k.

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

Proof. What we need to demonstrate is that

$$f_A(k+1) - f_A(k) \leq \lambda^{-1}(1 - e^{-\lambda}), \tag{5}$$

uniformly in A and $k \geq 1$. Indeed, by (5) and the second line of (2),

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

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

To prove (5) we may, by (4), assume that A is finite. In this case, we have, by (2),

$$f_A = \sum_{j \in A} f_j \tag{6}$$

where f_j is the abbreviation for $f_{\{j\}}$. By the first line of (3),

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

If $k \geq j$, then

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

which is positive and decreasing in k. If $k < j$, then

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

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 (3),

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

For $j = 0$, we merely observe that $f_0(k+1) - f_0(k) \leq 0$ for $k \geq 1$. Thus we have, by (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 (5) and ends the proof. $\square$

The essence of the Chen-Stein method is that an estimate

$$E[\lambda f_A(W+1) - W f_A(W)] \leq \alpha, \tag{7}$$

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:

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

To get (7) 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. Let $p_i = E(I_i)$, $W = \sum_{i \in \Gamma} I_i$, $W_i = W - I_i$, and $\lambda = EW = \sum_{i \in \Gamma} p_i$.

Assume first that these are *independent* indicators. Then W_i is independent of I_i, so that

$$\begin{aligned} E[\lambda f_A(W+1) - W f_A(W)] &= \sum_{i \in \Gamma} [p_i E f_A(W+1) - E(I_i f_A(W))] \\ &= \sum_{i \in \Gamma} [p_i E f_A(W+1) - E(I_i f_A(W_i+1))] \\ &= \sum_{i \in \Gamma} [p_i E f_A(W+1) - p_i E(f_A(W_i+1))] \\ &= \sum_{i \in \Gamma} p_i^2 E[f_A(W+1) - f_A(W_i+1)|I_i = 1], \end{aligned}$$

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 $\Delta f_A \cdot \sum_{i \in \Gamma} p_i^2$. This implies the following theorem, originally due to L. Le Cam.

Theorem 1. *If I_i are independent*

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

.

The first generalization of Theorem 1 is in the direction of *local* dependence. Assume that each indicator I_i has a set of indices Γ_i so that $i \notin \Gamma_i$ and so that $i \neq j \notin \Gamma_i$ implies I_j is independent of I_i.

Theorem 2.

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

.

Proof. Let $Z_i = \sum_{j \in \Gamma_i} I_j$ and $Y_i = W - I_i - Z_i$. The point is that Y_i are independent I_i. From here on the proof proceeds on familiar grounds

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

Now by telescoping

$$f_A(W+1) - f_A(Y_i+1) \le \Delta f_A \cdot (Z_i + I_i),$$
$$|f_A(Y_i+Z_i+1) - f_A(Y_i+1)| \le \Delta f_A \cdot Z_i,$$

and so

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

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

The second approach is *coupling*. The basic version requires that, *for a fixed* i, I_j and J_{ji} are constructed on the same probability space so that the following equality in distribution between the two vectors holds

$$(J_{ji})_{j\ne i} \stackrel{d}{=} (I_j)_{j\ne i} \,|\, I_i = 1.$$

For the method to work, we expect J_{ji} not to be very far from I_j, otherwise any coupling (say, the independent one) would do.

Theorem 3. *Under any coupling as above*

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

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

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

Now,

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

which is equivalent to the claim. □

If a coupling exists so that $J_{ji} \ge I_j$ (resp. $J_{ji} \le I_j$) for all i and $j \ne i$, then I_i are *positively* (resp. *negatively*) *related*.

Note that positively related indicators are positively correlated:

$$E(I_j) \le E(J_{ji}) = E(I_j|I_i = 1) = E(I_iI_j)/E(I_i).$$

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

In the positively related case,

$$p_iE|J_{ji} - I_j| = p_iE(J_{ji} - I_j) = E(I_iI_j) - p_ip_j,$$

while in the negatively related case,

$$piE|J_{ji} - I_j| = p_iE(I_j - J_{ji}) = p_ip_j - E(I_iI_j).$$

Corollary 1.
(1) In the positively related case

$$\begin{aligned} d_{TV}(W, P_\lambda) &\le \min(1, \lambda^{-1}) \left[2\sum_{i\in\Gamma} p_i^2 + \sum_{i,j,i\neq j} E(I_iI_j) - \lambda^2 \right] \\ &= \min(1, \lambda^{-1}) \left[2\sum_{i\in\Gamma} p_i^2 + \operatorname{Var} W - \lambda \right]. \end{aligned}$$

(2) In the negatively related case

$$\begin{aligned} d_{TV}(W, P_\lambda) &\le \min(1, \lambda^{-1}) \left[\lambda^2 - \sum_{i,j,i\neq j} E(I_iI_j) \right] \\ &= \min(1, \lambda^{-1}) \left[\lambda - \operatorname{Var} W \right]. \end{aligned}$$

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 Barbour et al.). 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!

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

Corollary 2. *Assume that Γ_i^n are sets of indices such that $i \notin \Gamma_i^n$ and such that $i \neq j \notin \Gamma_i^n$ implies $J_{ji} \ge I_j$. Then*

$$\begin{aligned} &d_{TV}(W, P_\lambda) \\ &\le \min(1, \lambda^{-1}) \left[2\sum_{i\in\Gamma} p_i^2 - \lambda^2 + \sum_{i,j,i\neq j\notin\Gamma_i^n} E(I_iI_j) + \sum_{i,j,j\in\Gamma_i^n} (2p_ip_j + E(I_iI_j)) \right]. \end{aligned}$$

Proof. For $j \in \Gamma_i^n$ simply estimate $p_iE|J_{ji} - I_j| \le p_iE(J_{ji} + I_j) = E(I_iI_j) + p_ip_j$ to get

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

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

4. Examples.

Example 1: Records. Here, I_i, $i = 1, \ldots, n$ are independent with $p_i = 1/i$. Then $\lambda = \lambda_n = 1 + \cdots + \frac{1}{n}$ and by Theorem 1

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

which is of some worth by itself, but we also get the CLT (assuming Z is a r.v. with $\mu_Z = P_\lambda$, and N a standard normal r.v.)

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

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

Example 2: Birthday Problem. Fix an integer $a \geq 2$ 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

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

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

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

This is a local problem, so we seek to apply Theorem 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

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

and

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

So Theorem 2 (together with Proposition 2) implies

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

Example 3: 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 first of which is in i. The initial 0 is important – it is used for "declumping," as long runs occurs in clumps. How many such runs do we have?

Let I_i indicate the event that there is a run for 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. (Note that t must be an integer, so we cannot assume that c is a constant.) Then $p^t = p^c/n$ and

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

Also

$$\sum_{i,j\in\Gamma_i} E(I_iI_j) = 0$$

and

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

Therefore,

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

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: 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). Then $\lambda = \lambda_n = EW = n(1-p)^{n-1}$, and the question is how large should $p = p_n$ be so that W is not likely 0. 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 will not work. This however is one of the simplest coupling cases. 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 (8) is then clearly satisfied. (The event that $I_i = 0$ is exactly the event that the $n-1$ specific edges emanating from i are missing.) Moreover, $I_j \le J_{ji}$, so 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

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

This proves that

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

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 shows that no matter how p varies with n, $P(W=0$, graph not connected$) \to 0$ as $n \to \infty$. So this formula also gives us a probability estimate for connectedness of a random graph.

Another note is that one can play this game for other values of p and get useful estimates. 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

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

and consequently

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

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

Example 5: Fixed points in random permutations. Let $(\pi(i))_{i=1}^n$ be a random permutation, $I_i = 1_{\{\pi(i)=i\}}$ and

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

Now to check (8), imagine the random permutation as ordering of numbers $1,\dots,n$, and imagine it being constructed by first choosing the place for i (i.e., $\pi(i)$), then independently choosing the order of the other $n-1$ numbers. The final deterministic step then builds the ordering of n numbers. What we need to check to verify (8) is that the second case above (interchanging i with the number in place i) keeps the $n-1$ numbers in the uniform random order. Assume that $\pi(i)=j>i$. Then this operation cyclically permutes $j-i$ numbers in the $(n-1)$-ordering, which of course does not spoil uniformity. (In fact, *any* deterministic permutation applied to the $(n-1)$-ordering preserves uniformity, hence any *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

$$d_{TV}(W, P_1) = \mathcal{O}\left(\frac{1}{n}\right),$$

which looks good, but is in fact very far from a realistic estimate. It is relatively easy to do explicit calculations to show that in this case

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

so there is practically no difference between μ_W and P_1 for large n. Barbour et al. has an entire chapter on when the Chen-Stein method gives correct order of d_{TV}.

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

$$\sum_{i,j,j\in\Gamma_i^n} (2p_ip_j + E(I_iI_j)) = \mathcal{O}\left(n \cdot \frac{1}{n^2}\right) = \mathcal{O}\left(\frac{1}{n}\right),$$

and

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

Therefore, in this case we also have

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

Example 6: 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. The coupling in this case is

$$J_{ji} = \begin{cases} I_j, & \text{if } I_i = 1, \\ 1_{\{j \text{ missing after all existing } i \text{ are indep. exchanged for random coupon not } i\}}, & \text{otherwise.} \end{cases}$$

This is a negatively related case: $J_{ji} \leq I_j$.

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

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

and

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

It follows that

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

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

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

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 6: Hypergeometric distribution. Arrange m 1's and $N - m$ 0's at random to form a random N-vector. Let $\Gamma = \{1, \ldots, n\}$ and let I_i indicate the event that a 1 is in the position i. Then W has hypergeometric distribution

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

with

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

(This is a straightforward, but tedious computation.) Also I_i are negatively related with

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

Therefore

$$d_{TV}(W, P_\lambda) = \min(1, \lambda^{-1}) \frac{N}{N-1}\left(\frac{n}{N} + \frac{m}{N} - \frac{nm}{N} - \frac{1}{N}\right).$$

This works well if both n and m are $o(N)$.