CLT and Poisson Convergence

8.1 Central Limit Theorems

Definition 8.1 (**Normal distribution**). *$X \sim N(0,1)$ if density function of X is,*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}$$

and the distribution function is $\Phi(x) = \int_{-\infty}^{x} \phi(t)\mathrm{d}t$. $X \sim N(\mu, \sigma^2)$ if $\frac{X-\mu}{\sigma} \sim N(0,1)$.

Proposition 8.2. *Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$. If $X_1 \perp X_2$, then*

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

In particular if $X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$, then $X_1 + \ldots + X_n \sim N(n\mu, n\sigma^2)$ or,

$$\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1).$$

Theorem 8.3 (**Basic Central Limit Theorem**). *If $X_1, \ldots, X_n$ are i.i.d. random variables with $\mathbb{E}(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$, then,*

$$\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z \sim N(0,1).$$

Theorem 8.4 (**Lindeberg's Central Limit Theorem**). *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}(X_i) = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$. Define $s_n^2 := \sum_{i=1}^n \sigma_i^2$. Then,*

$$\frac{1}{s_n} \sum_{i=1}^{n} (X_i - \mu_i) \xrightarrow{(d)} N(0,1),$$

if $s_n \to \infty$ and $\forall \varepsilon > 0$,

$$\frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\left(|X_i - \mu_i|^2 \mathbb{1}_{|X_i - \mu_i| \geqslant \varepsilon s_n} \right) \to 0.$$

8.1.1 Triangular Arrays

Roughly speaking, a sum of many small independent random variables will be approximately normally distributed. To formulate such a limit theorem, we must consider a sequence of sums of more and more, smaller and smaller random variables. Therefore, throughout this section we shall study the sequence of sums

$$S = \sum_j X_{ij}$$

obtained by summing the rows of a **triangular array** of random variables

$$\begin{array}{l} X_{11}, X_{12}, \ldots, X_{1n_1} \\ X_{21}, X_{22}, \ldots\ldots, X_{2n_2} \\ X_{31}, X_{32}, \ldots\ldots\ldots, X_{3n_3} \\ \vdots \quad \vdots \quad \vdots \quad \vdots \end{array}$$

It will be assumed throughout that the triangular arrays we consider satisfy three **Triangular Array Conditions**[1] (here i ranges over $\{1, 2, \ldots\}$, and j ranges over $\{1, 2, \ldots, n_i\}$):

1. For each i, the n_i random variables $X_{i1}, X_{i2}, \ldots, X_{in_i}$ in the ith row are mutually independent.
2. $\mathbb{E}\, X_{ij} = 0$ for all i, j, and
3. $\sum_j \mathbb{E}\, X_{ij}^2 = 1$ for all i.

We have some remarks for these conditions:

- It is **not** assumed that random variables in each row are identically distributed.
- It is **not** assumed that different rows are independent. In fact, a common application of triangular arrays is sums $X_1 + X_2 + \cdots + X_n$ obtained from a sequence of independent random variables $X_1, X_2, \ldots$.
- It will usually be the case that $n_i \to \infty$ as $i \to \infty$. And according to the nature of our problem, we should have the variables in each row tend to be smaller and smaller as i increases. Both of these two conditions are implied by the Lindeberg Condition which we will discuss below.

8.1.2 The Lindeberg Condition and Some Consequences

Theorem 8.5 (Lindeberg's Theorem). *Suppose that in addition to the Triangular Array Conditions, the triangular array satisfies Lindeberg's condition:*

$$\forall \varepsilon > 0, \quad \lim_{i \to \infty} \sum_{j=1}^{n_i} \mathbb{E}[X_{ij}^2 \mathbb{1}(|X_{ij}| > \varepsilon)] = 0. \tag{8.1}$$

Then $S_i \xrightarrow{(d)} \mathrm{N}(0, 1)$.

The Lindeberg condition makes precise the sense in which the random variables must be smaller and smaller. It says that for arbitrarily small $\varepsilon > 0$, the contribution to the total row variance from the terms with absolute value greater than ε becomes negligible as you go down the rows. We see this as follows:

$$\begin{aligned} X_{ij}^2 &\leqslant \varepsilon^2 + X_{ij}^2 \mathbb{1}(|X_{ij}| > \varepsilon) \\ \mathbb{E}\, X_{ij}^2 &\leqslant \varepsilon^2 + \mathbb{E}\, X_{ij}^2 \mathbb{1}(|X_{ij}| > \varepsilon) \leqslant \varepsilon^2 + \sum_{k \geqslant 1} \mathbb{E}\, X_{ik}^2 \mathbb{1}(|X_{ik}| > \varepsilon). \end{aligned}$$

[1] This is not standard terminology, but is used here as a simple referent for these conditions.

This last inequality is true for all j, so we have:

$$\max_j \mathbb{E}\, X_{ij}^2 \leqslant \varepsilon^2 + \sum_j \mathbb{E}\, X_{ij}^2 \mathbb{1}(|X_{ij}| > \varepsilon) \tag{8.2}$$

The Lindeberg condition says that, as $i \to \infty$, the summation on the RHS of (8.2) tends to zero. Since (8.2) holds for all $\varepsilon > 0$, we get

$$\lim_{i\to\infty} \max_j \mathbb{E}\, X_{ij}^2 = 0, \tag{8.3}$$

which implies $n_i \to \infty$ as $i \to \infty$, since we assume in Triangular Array Condition that $\sum_j \mathbb{E}\, X_{ij}^2 = 1$ for all i. Another consequence follows from (8.3) and Chebyshev's inequality: since we have

$$\mathbb{P}(|X_{ij}| > \varepsilon) \leqslant \frac{\mathbb{E}\, X_{ij}^2}{\varepsilon^2} \text{ for all } \varepsilon > 0,$$

taking the maximum over j and $i \to \infty$, we get that $X_{ij} \xrightarrow{\mathbb{P}} 0$, uniformly in j:

$$\forall \varepsilon > 0, \quad \lim_{i\to\infty} \max_j \mathbb{P}(|X_{ij}| > \varepsilon) = 0. \tag{8.4}$$

An array with property (8.4) is said to be **Uniformly Asymptotically Negligible (UAN)**, and there is a striking converse to Lindeberg's Theorem:

Theorem 8.6 (Feller's Theorem)**.** *If a triangular array satisfies the Triangular Array Conditions and is UAN, then* $S_i \xrightarrow{(d)} \mathrm{N}(0,1)$ *(if and) only if Lindeberg's condition* (8.1) *holds.*

Proof. See Billingsley, Theorem 27.4, or Kallenberg, 5.12. ■

Proof of Basic Central Limit Theorem. We show that the Lindeberg condition holds by taking

$$X_{nj} = \frac{X_j - \mu}{\sqrt{n\sigma^2}}, \quad j = 1, 2, \ldots, n.$$

Then

$$\begin{aligned}\sum_{j=1}^n \mathbb{E}[X_{nj}^2 \mathbb{1}(|X_{nj}| > \varepsilon)] &= \frac{n}{n\sigma^2}\, \mathbb{E}[(X_1-\mu)^2 \mathbb{1}(|X_1 - \mu| > \varepsilon\sigma\sqrt{n})] \\ &= \frac{1}{\sigma^2}\, \mathbb{E}[(X_1-\mu)^2 \mathbb{1}(|X_1 - \mu| > \varepsilon\sigma\sqrt{n})]\end{aligned}$$

which converges to 0 by DCT. ■

8.1.3 Rate of Convergence for CLT

Theorem 8.7 (Berry-Esseen Theorem)**.** *Let* $X_1, X_2, \ldots$ *be i.i.d. r.v.s with* $\mathbb{E}(X_1) = 0, \mathrm{Var}(X_1) = \sigma^2$ *and* $\mathbb{E}\,|X_1|^3 < \infty$*. Define* $S_n := (X_1 + X_2 + \cdots + X_n)/\sqrt{n\sigma^2}$*. Then*

$$\sup_{x\in\mathbb{R}} |\,\mathbb{P}(S_n \leqslant x) - \Phi(x)| \leqslant \frac{3\,\mathbb{E}\,|X_1|^3}{\sqrt{n}\sigma^3}.$$

8.2 Preliminaries to the proof of Lindeberg's Theorem

There are several ways of proving Central Limit Theorems:

1. Use characteristic or moment generating functions or some distributional transform, or
2. Use moment method to show that the k-th moment converges to the k-th moment of standard Normal for all $k \geqslant 1$, or
3. Use Fixed Point method (*e.g.,* maximizing entropy given fixed mean and variance, zero bias transformation etc.) or
4. Replacement or exchange techniques.
5. Stein's method (to be discussed later in the course).

We introduce two preliminaries to the proof.

Lemma 8.8. *If $X \sim \mathrm{N}(0, \sigma^2)$, $Y \sim \mathrm{N}(0, \tau^2)$ are independent, then $X + Y \sim \mathrm{N}(0, \sigma^2 + \tau^2)$.*

Lemma 8.9. *$S_i \xrightarrow{(d)} Z$ if and only if $\lim_{i\to\infty} \mathbb{E}\, f(S_x) = \mathbb{E}\, f(Z)$ for all $f \in \mathbf{C}_b^3(\mathbb{R})$, the set of functions from $\mathbb{R}$ to $\mathbb{R}$ with three bounded, continuous derivatives.*

Proof. See Durrett, Theorem 2.2, and use that $\mathbf{C}_b^3(\mathbb{R})$ is dense in $\mathbf{C}_b(\mathbb{R})$. ■

8.3 Proof of Lindeberg's Theorem

Proof. First, we will work with a fixed row. To simplify things we will avoid writing the subscript i, so that X_{ij} will be denoted by X_j and n_i will be denoted by n.

Let $X_1, X_2, \ldots, X_n$ be independent random variables, not necessarily identically distributed. Suppose $\mathbb{E}\, X_j = 0$ and let $\sigma_j^2 = \mathbb{E}(X_j^2) < \infty$. Then for $S = \sum_{j=1}^n X_j$ we have $1 = \operatorname{Var} S = \sum_{j=1}^n \sigma_j^2$. Note that,

1. If $\forall j$, $X_j \sim \mathrm{N}(0, \sigma_j^2)$, then $S \sim \mathrm{N}(0, 1)$.
2. Given independent random variables $X_1, X_2, \ldots, X_n$ with arbitrary distributions, we can always construct a new sequence $Z_1, Z_2, \ldots, Z_n$ of **Normal** random variables with matching means and variances so that all of Z_i and X_i are mutually independent. This may involve changing the basic probability space, but that does not matter because the distribution of S is determined by the joint distribution of $(X_1, X_2, \ldots, X_n)$, which remains the same.

Let

$$\begin{aligned} S := W_0 &:= X_1 + X_2 + X_3 + \ldots + X_n, \\ W_1 &:= Z_1 + X_2 + X_3 + \ldots + X_n, \\ W_2 &:= Z_1 + Z_2 + X_3 + \ldots + X_n, \\ &\vdots \qquad\quad \vdots \qquad\quad \vdots \\ T := W_n &:= Z_1 + Z_2 + Z_3 + \ldots + Z_n, \end{aligned}$$

We want to show that S is "close" in distribution to T, i.e., that $\mathbb{E}\, f(S)$ is close to $\mathbb{E}\, f(T)$ for all $f \in \mathbf{C}_b^3(\mathbb{R})$ with uniform bound K on f and its first three derivatives: $|f^{(i)}|,\ i = 1, 2, 3$.

By the triangle inequality,

$$|\,\mathbb{E}\, f(S) - \mathbb{E}\, f(T)| \leqslant \sum_{j=1}^{n} |\,\mathbb{E}\, f(W_j) - \mathbb{E}\, f(W_{j-1})|. \tag{8.5}$$

Let $W_j^- = Z_1 + Z_2 + \cdots + Z_{j-1} + X_{j+1} + \cdots + X_n$ be the sum of the common terms in W_{j-1} and W_j. Then

$$W_{j-1} = W_j^- + X_j \text{ and } W_j = W_j^- + Z_j.$$

Note that by construction, W_j^- and X_j are independent, as are W_j^- and Z_j. We need to compare $\mathbb{E}\, f(W_j^- + X_j)$ and $\mathbb{E}\, f(W_j^- + Z_j)$. By the Taylor series expansion up to the third term,

$$f(W_j^- + X_j) = f(W_j^-) + X_j f^{(1)}(W_j^-) + \frac{X_j^2}{2!} f^{(2)}(W_j^-) + \mathrm{Err}_f(W_j^-, X_j)$$
$$f(W_j^- + Z_j) = f(W_j^-) + Z_j f^{(1)}(W_j^-) + \frac{Z_j^2}{2!} f^{(2)}(W_j^-) + \mathrm{Err}_f(W_j^-, Z_j)$$

where

$$\mathrm{Err}_f(a, x) := f(a + x) - f(a) - x f'(a) - \frac{x^2}{2} f''(a)$$

satisfies

$$|\mathrm{Err}_f(a, x)| \leqslant \min\left\{ \frac{|x|^3}{6} \cdot ||f^{(3)}||_\infty, x^2 \cdot ||f^{(2)}||_\infty \right\} \leqslant K x^2 \min\{1, |x|\}.$$

Here we used the minimum bound, so that upper bound stays integrable for $x = X_j$.

We can take expectations in each of these identities and subtract the resulting equations. Using independence and the fact that X_j and Z_j agree on their first and second moments, we see that everything upto the second order cancels. Therefore,

$$\begin{aligned} |\,\mathbb{E}\, f(W_j) - \mathbb{E}\, f(W_{j-1})| &= |\,\mathbb{E}\, f(W_j^- + X_j) - \mathbb{E}\, f(W_j^- + Z_j)| \\ &\leqslant \mathbb{E}(|\mathrm{Err}_f(W_j^-, X_j)| + |\mathrm{Err}_f(W_j^-, Z_j)|) \\ &\leqslant K\, \mathbb{E}\left(X_j^2 \min\{1, |X_j|\} + |Z_j|^3\right). \end{aligned} \tag{8.6}$$

Let $c = \sqrt{8/\pi}$ be the third absolute moment of a standard normal random variable as,

$$c := 2 \int_0^\infty x^3 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 2 \cdot \frac{2}{\sqrt{2\pi}} < \infty$$

Therefore, $E|Z_j|^3 = c\sigma_j^3$. Moreover, we have $\min\{1, |x|\} \leqslant \varepsilon + \mathbb{1}_{|x|>\varepsilon}$ for all $\varepsilon > 0$. Applying this to (8.6), we get

$$|\,\mathbb{E}\, f(W_j) - \mathbb{E}\, f(W_{j-1})| \leqslant K\left(\varepsilon\, \mathbb{E}\, X_j^2 + \mathbb{E}\, X_j^2 \mathbb{1}_{|X_j|>\varepsilon} + c\sigma_j^3\right).$$

Combining, from (8.5), we get

$$|\,\mathbb{E}\, f(S) - \mathbb{E}\, f(T)| \leqslant K\left(\varepsilon + \sum_j \mathbb{E}\, X_j^2 \mathbb{1}_{|X_j|>\varepsilon} + c \sum_j \sigma_j^3\right) \tag{8.7}$$

So far we have only considered one row of the array, but (8.7) is in fact true for every row with K and c unchanged and $T \sim \mathrm{N}(0,1)$. Thus, for each i we have,

$$\begin{aligned} |\,\mathbb{E}\, f(S_i) - \mathbb{E}\, f(T)| &\leqslant K\big(\varepsilon + \sum_j \mathbb{E}\, X_{ij}^2 \mathbb{1}_{|X_{ij}|>\varepsilon} + c\sum_j \sigma_{ij}^3\big) \\ &\leqslant K\big(\varepsilon + \sum_j \mathbb{E}\, X_{ij}^2 \mathbb{1}_{|X_{ij}|>\varepsilon} + c\max_j \sigma_{ij}\big). \end{aligned} \tag{8.8}$$

Under Lindeberg's condition, the RHS of (8.8) goes to $K\varepsilon$ as $i \to \infty$. Since $\varepsilon > 0$ is arbitrary, the LHS converges to zero as $i \to \infty$. By Lemma 10.6, $S_i \xrightarrow{(d)} \mathrm{N}(0,1)$ as $i \to \infty$. ■

Here, we proved that for $f \in C_b^3(\mathbb{R}) : |f''|, |f'''| < \infty$, we have

$$\begin{aligned} |\mathbb{E}\, f(S_i) - \mathbb{E}\, f(T)| &= \left|\sum_{j=1}^n \mathbb{E}(\mathrm{Err}_f(W_{ij}^-, X_{ij}) - \mathrm{Err}_f(W_{ij}^-, Z_{ij}))\right| \\ &\leqslant \frac{1}{6}\sum_{j=1}^n \mathbb{E}(X_{ij}^2 \min\{6|f''|_\infty, |f'''|_\infty \cdot |X_{ij}|\} + c\cdot|f'''|_\infty \cdot \sigma_{ij}^3) \end{aligned}$$

In fact, we proved a stronger result than convergence in distribution. We proved that

$$\sup_{f\in C_b^3(\mathbb{R}):|f''|,|f'''|\leqslant K} |\mathbb{E}\, f(S_i) - \mathbb{E}\, f(T)| \leqslant \frac{K}{6}\sum_{j=1}^n (\mathbb{E}(X_{ij}^2 \min\{6, |X_{ij}|\}) + c\cdot\sigma_{ij}^3)$$

which goes to zero under Lindeberg's condition. If we have $\mathbb{E}\,|X_{ij}|^3 < \infty$, we get

$$\begin{aligned} \sup_{f\in C_b^3(\mathbb{R}):|f''|,|f'''|\leqslant K} |\mathbb{E}\, f(S_i) - \mathbb{E}\, f(T)| &\leqslant \frac{K}{6}\sum_{j=1}^n (\mathbb{E}\,|X_{ij}|^3 + c\cdot\sigma_{ij}^3) \\ &\leqslant \frac{K}{6}(1+c)\sum_{j=1}^n \mathbb{E}\,|X_{ij}|^3 \end{aligned}$$

using the fact that $\sigma_{ij}^3 = ||X_{ij}||_2^3 \leqslant ||X_{ij}||_3^3 = \mathbb{E}\,|X_{ij}|^3$. This is similar to the Berry-Esseen bound, but here the functions f are much smoother.

Example 8.10. ***(Least Square estimate in Linear Regression Model)*** *Let $Y_i = \beta x_i + \eta_i$ for $i = 1, 2, \ldots, n$ where $(x_i)_{i\geqslant 1}$ is a sequence of real numbers, η_i's are independent random variable with $\mathbb{E}(\eta_i) = 0$ and $\mathrm{Var}(\eta_i) = \sigma^2$ for all i. Here, β is unknown and $\sigma^2 > 0$ is known. The least square estimate of β is given by*

$$\beta_{LS} := \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

We have

$$\frac{\sqrt{\sum_{i=1}^n x_i^2}}{\sigma}\cdot(\beta_{LS} - \beta) = \frac{\sum_{i=1}^n x_i\eta_i}{\sqrt{\sum_{i=1}^n x_i^2}} \xrightarrow{(d)} N(0,1)$$

if for some $\varepsilon > 0$

$$\max_{i\geqslant 1} \mathbb{E}\,|\eta_i|^{2+\varepsilon} < \infty \text{ and } \frac{\max_{1\leqslant i\leqslant n} x_i^2}{\sum_{i=1}^n x_i^2} \to 0 \text{ as } n \to \infty.$$

8.4 Poisson Convergence and Law of Small Numbers

By Central Limit Theorem,

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{(d)} N(0,1)$$

as $n \to \infty$, where $X_1, X_2, \ldots$ are iid Berboulli(p) rvs and $S_n = X_1 + X_2 + \cdots + X_n, n \geqslant 1$. In general, if $p = p_n \to 0$ and $np_n(1-p_n) \to \infty$ or $\frac{1}{n} \ll p_n \ll 1$, the convergence still holds.

If $np_n \to 0$, we have

$$\mathbb{E}\, S_n^2 = (np_n)^2 + np_n(1-p_n) \to 0 \text{ and hence } S_n \xrightarrow{\mathbf{L^2}} 0.$$

What if $np_n \to \lambda > 0$?

Definition 8.11. *A r.v. X is Poisson(λ) distributed if*

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \text{ for } k = 0, 1, 2, \ldots.$$

Lemma 8.12. *If $np_n \to \lambda \in (0, \infty)$, then $S_n \xrightarrow{(d)}$ Poisson(λ).*

Proof. Suffices to show that, $\mathbb{P}(S_n = k) \to e^{-\lambda}\frac{\lambda^k}{k!}$ for all $k = 0, 1, 2, ...$ We can compute this quantity exactly: $\mathbb{P}(S_n = k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} = \frac{1}{k!}\prod_{i=0}^{k-1}\left(1 - \frac{i}{n}\right) \cdot (np)^k \left(1 - \frac{np}{n}\right)^{n-k} \to \frac{\lambda^k}{k!}e^{-\lambda}$. ■

Theorem 8.13 (Law of Small Numbers). *Let $\{X_{ij}, j = 1, 2, \ldots, n_i\}$ be independent, integer-valued random variables for each $i \geqslant 1$. Define $S_i := \sum_{j=1}^{n_i} X_{ij}$ for $i \geqslant 1$. Assume that*

(i) $\mathbb{P}(X_{ij} = 1) = p_{ij}$ *satisfies* $\sum_{j=1}^{n_i} p_{ij} \to \lambda \in (0, \infty)$,

(ii) $\sum_{j=1}^{n_i} \mathbb{P}(X_{ij} \notin \{0, 1\}) \to 0$ *and*

(iii) $\max_{1 \leqslant j \leqslant n_i} p_{ij} \to 0$

as $i \to \infty$. Then $S_i \xrightarrow{(d)}$ Poisson(λ).

There is a more general version of the LSN relaxing the independence assumption. The **Poisson Universality Class** contains sequences of rvs satisfying the three conditions above.

Proof Sketch: We'll show the total variation distance

$$d_{\mathrm{TV}}(S_i, N) := \sup_{A \in \mathcal{B}} |\,\mathbb{P}(S_i \in A) - \mathbb{P}(N \in A)| \to 0$$

where $N \sim \text{Poisson}(\lambda)$ (this is stronger than convergence in distribution).

Exercise 8.14. *(a) Prove that $d_{TV}(\cdot, \cdot)$ is a metric on the set of probability measures on Ω.*
(b) Prove that $d_{TV}(\mu, v) = \frac{1}{2}\sum_k |\mu\{k\} - \nu\{k\}|$ if μ, ν are discrete distributions on $\mathbb{Z}$.

Again, we fixed i and remove the subscript i for easy writing. Let

$$\widehat{X}_j = 1_{X_j = 1}, \quad S = \sum_{j=1}^{n} X_j \quad \text{and} \quad \widehat{S} = \sum_{j=1}^{n} \widehat{X}_j.$$

We have

$$d_{\mathrm{TV}}(S, N) \leqslant d_{\mathrm{TV}}(\widehat{S}, N) + d_{\mathrm{TV}}(S, \widehat{S})$$

and

$$d_{\mathrm{TV}}(S, \widehat{S}) \leqslant \mathbb{P}(S \neq \widehat{S}) \leqslant \mathbb{P}(\cup_{j=1}^{n} \{X_j \neq \widehat{X}_j\}) \leqslant \sum_{j=1}^{n} \mathbb{P}(X_j \notin \{0, 1\}).$$

We have $\mathbb{P}(\widehat{X}_j = 1) = p_j$ and $\mathbb{P}(\widehat{X} = 0) = 1 - p_j$. We proceed as follows

1. Sum of independent Poisson rvs is Poisson: $N_j \sim \mathrm{Poisson}(\lambda_j)$ for $j = 1, 2, \ldots, n$ and independent, then $\sum_{j=1}^{n} N_j \sim \mathrm{Poisson}(\sum_{j=1}^{n} \lambda_j)$.

2. Now assume that $\widehat{\lambda} := \sum_{j=1}^{n} p_j$. Let $Y_j \sim \mathrm{Poisson}(p_j)$, independent, for $i = 1, 2, \ldots, n$. Then $\sum_{j=1}^{n} Y_j \sim \mathrm{Poisson}(\widehat{\lambda})$.

3. We have

$$\begin{aligned} d_{\mathrm{TV}}(\widehat{X}_j, Y_j) &= \frac{1}{2} \sum_{k=0}^{\infty} |\, \mathbb{P}(\widehat{X}_j = k) - \mathbb{P}(Y_j = k)| \\ &= \frac{1}{2} \left(|1 - p_j - e^{-p_j}| + |p_j - p_j e^{-p_j}| + (1 - e^{-p_j} - p_j e^{-p_j}) \right) = p_j (1 - e^{-p_j}) \leqslant p_j^2. \end{aligned}$$

4. Finally, if suffices to show $d_{\mathrm{TV}}(X_1 + X_2, Y_1 + Y_2) \leqslant d(X_1, Y_1) + d(X_2, Y_2)$ when $X_1 \perp X_2$ and $Y_1 \perp Y_2$. We have,

$$\begin{aligned} &\sum_{k} |\, \mathbb{P}(X_1 + X_2 = k) - \mathbb{P}(Y_1 + Y_2 = k)| \\ &= \sum_{k} |\sum_{\ell} (\mathbb{P}(X_1 = \ell)\, \mathbb{P}(X_2 = k - \ell) - \mathbb{P}(Y_1 = \ell)\, \mathbb{P}(Y_2 = k - \ell))| \\ &\leqslant \sum_{k,\ell} |(\mathbb{P}(X_1 = \ell) - \mathbb{P}(Y_1 = \ell)) \cdot \mathbb{P}(X_2 = k - \ell) + \mathbb{P}(Y_1 = \ell) \cdot (\mathbb{P}(X_2 = k - \ell) - \mathbb{P}(Y_2 = k - \ell)| \\ &\leqslant \sum_{\ell} |\, \mathbb{P}(X_1 = \ell) - \mathbb{P}(Y_1 = \ell)| + \sum_{\ell} |\, \mathbb{P}(X_2 = \ell) - \mathbb{P}(Y_2 = \ell)| \end{aligned}$$

   So

$$d_{\mathrm{TV}} \left(\sum_{j=1}^{n} \widehat{X}_j, \sum_{j=1}^{n} Y_j \right) \leqslant \sum_{j=1}^{n} d_{\mathrm{TV}}(\widehat{X}_j, Y_j) \leqslant \sum_{j=1}^{n} p_j^2.$$

5. We have $d_{\mathrm{TV}}(\mathrm{Poisson}(\widehat{\lambda}), \mathrm{Poisson}(\lambda)) \leqslant \mathbb{P}(\mathrm{Poisson}(|\widehat{\lambda} - \lambda|) > 0) \leqslant |\widehat{\lambda} - \lambda|$.

Combining we get

$$d_{\mathrm{TV}}(S, N) \leqslant \sum_{j=1}^{n} \mathbb{P}(X_j \notin \{0, 1\}) + \sum_{j=1}^{n} p_j^2 + \left| \sum_{j=1}^{n} p_j - \lambda \right|.$$

Finally putting back i and using the three assumptions we have the result. ■

Remark 8.15. *It follows from the previous analysis that, if $X_j \sim$ Bernoulli(p_j) for $j = 1, 2, \ldots, n$ and are independent, then*

$$d_{TV}\Big(\sum_j X_j, Poisson(\lambda)\Big) \leqslant \sum_j p_j^2.$$

where $\lambda := \sum_{j=1}^n p_j$. However, the bound is not optimal. One can prove using Stein-Chen method that

$$d_{TV}\Big(\sum_j X_j, Poisson(\lambda)\Big) \leqslant \frac{\sum_j p_j^2}{\max\{1, \lambda\}} \leqslant \max_j p_j.$$

Moreover, if X_j's are positively correlated, we have

$$d_{TV}\Big(\sum_j X_j, Poisson(\lambda)\Big) \leqslant \frac{\sum_j p_j^2 + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)}{\max\{1, \lambda\}}$$

and if X_j's are negatively correlated, we have

$$d_{TV}\Big(\sum_j X_j, Poisson(\lambda)\Big) \leqslant \frac{\sum_j p_j^2 - \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)}{\max\{1, \lambda\}}$$

In some example the rate of decay is much faster than the one provided by the above Theorem.

Example 8.16. *Consider a random uniform permutation π and let $W = \sum_{i=1}^n \mathbb{1}_{\pi_i = i}$ be the number of fixed points in π. Here, $X_i = \mathbb{1}_{\pi_i = i}, i = 1, 2, \ldots, n$ are Bernoulli$(1/n)$ rvs. Hence $\mathbb{E}(W) = 1$. Note that, $\mathbb{E}(X_i X_j) - \mathbb{E}\, X_i\, \mathbb{E}\, X_j = 1/n(n-1) - 1/n^2 = 1/n^2(n-1) > 0$ for $i < j$. Thus, we have*

$$d_{TV}\Big(\sum_j X_j, Poisson(1)\Big) \leqslant 1/n + \sum_{i\neq j} 1/n^2(n-1) = 2/n.$$

However, using inclusion exclusion principle, we have

$$\mathbb{P}(W = k) = \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}, k = 0, 1, \ldots, n.$$

Thus,

$$d_{TV}\Big(\sum_j X_j, Poisson(1)\Big) = \frac{1}{2} \sum_{k=0}^{n} \frac{1}{k!} \sum_{i=n-k+1}^{\infty} \frac{(-1)^{i-n+k-1}}{i!} + \frac{1}{2} \sum_{k=n+1}^{\infty} \frac{e^{-1}}{k!} = \frac{2^n}{(n+1)!}(1 + O(1/n)).$$