STEIN-CHEN METHOD FOR POISSON APPROXIMATION
ST414 (TERM 2, 2013–2014)

PARTHA DEY
UNIVERSITY OF WARWICK

Content.
- Direct Poisson approximation for independent events.
- Description of the Stein-Chen method.
- Applications of the Stein-Chen method for classes of independent and dependent events.
- Examples.

Objectives.
- Understand the principles of the Stein-Chen method.
- Apply the Stein-Chen method to examples.

Prerequisites.
- Expectation, independence, conditional distributions, Bernoulli indicator random variables. Poisson random variables.
- The course ST111 (Probability A) covers all of them and is essential for this course.
- ST318 (Probability Theory) might also help with various proofs and ideas, but is not mandatory.

Reading.
- Poisson approximation: Barbour, Holst and Janson, 1992. This course is based on a subset of the book.
- Coupling and Poisson approximation: Janson, 1994. Survey paper.
- Lecture notes online after lecture.
- Example sheets at the end of the notes. Do these as you go along.

Contents

1. Introduction

Stein-Chen method (named after Charles Stein and Louis Chen) was developed to show that the probabilities of rare events can be approximated by Poisson probabilities. Just as common events can often be approximated by the Normal distribution, we will see that probabilities associated with rare events can often be modeled by the Poisson distribution. Stein-Chen method is a powerful, modern technique which extends classical Poisson approximation results such as Poisson's law of small numbers, even to cases with dependence between events. There is a large literature on Stein's method applied to general distribution, but in this course we will only look at Poisson approximation. First we will describe the classical result, and look at some motivating examples for why this might need extending[1].

1.1. **Laws of Large and small numbers.** For the purpose of comparison first recall the law of large numbers. Let $X_1, X_2, \ldots$ be independent, identically distributed (i.i.d.) random variables with mean $\mathbb{E}(X_1) = \mu$.

Theorem 1.1 (Weak Law of Large Numbers)**.** *As $n \to \infty$, for every $\varepsilon > 0$*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right) \to 0.$$

Moreover if the variance $\mathrm{Var}(X_i) = \sigma^2$ is finite:

Theorem 1.2 (Central Limit Theorem)**.** *As $n \to \infty$,*

$$\sqrt{n} \cdot \left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} \mathrm{N}(0, \sigma^2).$$

This explains why distributions that are approximately Normal are often observed. However, in the context of rare events, the Poisson distribution is often observed. Recall that, $\mathrm{Bin}(n,p)$ is the binomial distribution with parameters n and p, *i.e.,*

$$W \sim \mathrm{Bin}(n,p) \;\Leftrightarrow\; \mathbb{P}(W = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k \in \{0, 1, \ldots, n\}$. Similarly, $\mathrm{Poi}(\lambda)$ is the Poisson distribution with parameter λ, *i.e.,*

$$\widetilde{W} \sim \mathrm{Poi}(\lambda) \;\Leftrightarrow\; \mathbb{P}(\widetilde{W} = k) = e^{-\lambda}\lambda^k / k!$$

for $k \in \{0, 1, 2, \ldots\}$.

Theorem 1.3 (Poisson law of small numbers)**.** *Let $W \sim \mathrm{Bin}(n, \lambda/n)$, $\lambda > 0$. As $n \to \infty$, for any $k \in \{0, 1, 2, \ldots\}$,*

$$\mathbb{P}(W = k) \to e^{-\lambda}\frac{\lambda^k}{k!} = \mathbb{P}(\mathrm{Poi}(\lambda) = k).$$

1.2. **Applications.** Consider n independent trials each with success probability p. Let W be the number of successes $\sim \mathrm{Bin}(n,p)$. If n is large and $p = \lambda/n$ is small, then $W \approx \mathrm{Poi}(\lambda)$. First we look at a real life dataset.

(a) **Horse kick data by Bortkewitch.** Bortkewitch (1898) gathered 200 items of data, the number of people kicked to death by horses each year for 10 corps in the Prussian army over a 20 year period.

No. deaths k	Freq.	Poisson approx. $200 \times \mathbb{P}(\mathrm{Poi}(0.61) = k)$
0	109	108.67
1	65	66.29
2	22	20.22
3	3	4.11
4	1	0.63
5	0	0.08
6	0	0.01

Observations:

- Death by horse-kick in the Prussian army was a rare event. Median and mode are 0. The average is $\lambda = 0.61$.

[1]These course note is based on Prof. Wilfrid Kendall's course notes of the same name for 2007-08 spring term

- You might expect deaths per corps per year to be independent, or weakly dependent.
- The data set is famous for being such a good fit for the Poisson distribution. This suggest publication bias. Nonetheless the phenomenon is real.

(b) **Treize.** Treize is a simple card game studied by Montmort in 1713. Take a regular pack of 52 cards. For simplicity, assume the cards are labelled 1,2,. . . ,13; with four cards of each type.
- Draw 13 cards.
- Let $X_i = 1$ if the i-th card is of type i; otherwise 0.
- Is $W = \sum_{i=1}^{13} X_i$, the total number of matches, approximately Poi(1)?
- Independence fails. The (X_i) are positively related.

(c) **Occupancy problems.**
- Place m balls into n boxes.
- Place each ball into box i with probability p_i, $i = 1, \dots, n$; independently of the other balls.
- Let $X_i = 1$ if box i is empty; 0 otherwise.
- Is W approximately Poisson? Independence fails: the (X_i) are negatively related.

(d) **Nearest-neighbor statistics.**
- Independently choose points $Y_1, \dots, Y_n$ uniformly at random in the unit square $[0, 1]^2$.
- Let $|Y_i - Y_j|$ denote the distance between Y_i and Y_j with respect to toroidal boundary conditions.
- Let $X_{ij} = 1$ if $|Y_i - Y_j| < r$ (r small); 0 otherwise.
- Let $W = \sum_{i<j} X_{ij}$.
- Is W approximately Poisson? Independence fails: the (X_{ij}) are positively related and weakly dissociated.

By applying the techniques that we will prove in this course, we will show how Poisson approximation can be applied in each of these examples and give some quantitative description of the approximation. The idea is to prove Poisson approximation in a greater generality. The general idea being, the total number of successes in a large number of trials in which each individual success is rare and the trials are "weakly dependent", is approximately Poisson.

Notation. We will write $X \sim \mathrm{Ber}(p)$ if

$$\mathbb{P}(X = i) = \begin{cases} p & i = 1, \\ 1 - p & i = 0. \end{cases}$$

Let $X_1, \dots, X_n \sim \mathrm{Ber}(p)$. Then $W = \sum_{i=1}^{n} X_i \sim \mathrm{Bin}(n, p)$. We will use $\mathbb{E}(f(W); A)$ for $\mathbb{E}(f(W)\mathbb{1}_A)$ and will reserve $\widetilde{W}$ as a random variable having $\mathrm{Poi}(\lambda)$ distribution.

2. Poisson approximation for Binomial distribution

We will now prove the Poisson law of small numbers (Theorem 1.3), *i.e.,* if $W \sim \mathrm{Bin}(n, \lambda/n)$ with $\lambda > 0$, then as $n \to \infty$,

$$\mathbb{P}(W = k) \to e^{-\lambda}\frac{\lambda^k}{k!} = \mathbb{P}(\mathrm{Poi}(\lambda) = k).$$

Proof. It is an exercise to show that:

$$\exp(-p/(1-p)) \leqslant 1 - p \leqslant \exp(-p) \quad \text{forall} \quad p \in (0,1). \tag{1}$$

Thus

$$\begin{aligned}\mathbb{P}(W = k) &= \binom{n}{k}(\lambda/n)^k(1-\lambda/n)^{n-k}\\ &= \frac{n(n-1)\cdots(n-k+1)}{k!}(\lambda/n)^k(1-\lambda/n)^n(1-\lambda/n)^{-k}\\ &= (1-\lambda/n)^n \times \frac{\lambda^k}{k!} \times \frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n} \times (1-\lambda/n)^{-k}\end{aligned}$$

For fixed k, as $n \to \infty$,

$$\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n} \to 1, \quad (1-\lambda/n)^k \to 1.$$

Using (1),

$$\exp\left(-\frac{\lambda/n}{1-\lambda/n}\right) \leqslant 1 - \frac{\lambda}{n} \leqslant \exp(-\lambda/n),$$

so

$$\exp\left(-\frac{\lambda}{1-\lambda/n}\right) \leqslant \left(1-\frac{\lambda}{n}\right)^n \leqslant \exp(-\lambda).$$

By the sandwich principle $\left(1-\frac{\lambda}{n}\right)^n \to \exp(-\lambda)$ as $n \to \infty$. ∎

By adapting the argument above, we can produce a slightly more general result. Let $(p_n : n \geqslant 1)$ denote a sequence of positive numbers.

Theorem 2.1. *If $np_n \to \lambda$ as $n \to \infty$, for any fixed non-negative integer k*

$$\mathbb{P}(\mathrm{Bin}(n, p_n) = k) \to \mathbb{P}(\mathrm{Poi}(\lambda) = k) \ \text{as } n \to \infty.$$

These results suggest that $\mathrm{Poi}(np)$ is a good approximation for $\mathrm{Bin}(n,p)$ for large n and small p. How good is the approximation? Let

$$A_{n,p,k} = \frac{\mathbb{P}(\mathrm{Bin}(n,p) = k)}{\mathbb{P}(\mathrm{Poi}(np) = k)} = \begin{cases}(1-1/n)\dots(1-(k-1)/n)\exp(np)(1-p)^{n-k} & \text{if } 0 \leqslant k \leqslant n\\ 0 & \text{if } k > n.\end{cases}$$

Then

$$\begin{aligned}A_{n,p,k} &\geqslant \exp\left(-\sum_{j=1}^{k-1}\frac{j/n}{1-j/n} + np - (n-k)\frac{p}{1-p}\right)\\ &= \exp\left(-\sum_{j=1}^{k-1}\frac{j}{n-j} - \frac{(np-k)p}{1-p}\right) \geqslant \exp\left(-\frac{k(k-1)}{2(n-k+1)} + \frac{(k-np)p}{1-p}\right).\end{aligned}$$

and

$$A_{n,p,k} \leqslant \exp\left(-\sum_{j=1}^{k-1}\frac{j}{n} + np - (n-k)p\right) \leqslant \exp\left(-\frac{k(k-1)}{2n} + kp\right).$$

Theorem 2.2. *If $0 \leqslant k \leqslant n$,*

$$\begin{aligned}\mathbb{P}(\mathrm{Poi}(np) = k)\exp\left(-\frac{k(k-1)}{2(n-k+1)} + \frac{(k-np)p}{1-p}\right)&\\ \leqslant \mathbb{P}(\mathrm{Bin}(n,p) = k)&\\ \leqslant \mathbb{P}(\mathrm{Poi}(np) = k)\exp\left(-\frac{k(k-1)}{2n} + kp\right)&.\end{aligned}$$

Note that the bounds are not the best possible and only works for sum of independent Bernoullis. We need a good way to measure the distance between $\mathrm{Bin}(n,p)$ and $\mathrm{Poi}(np)$ over the whole of $\mathbb{Z}_+$?

3. Total-variation distance and Coupling

We have obtained bounds for $\mathrm{Bin}(n,p)$ probabilities in terms of $\mathrm{Poi}(np)$ probabilities. In what follows however, it will be useful to define a single measure of how apart two distributions are.

Definition 3.1. *Let P and Q denote two probability measures on $\mathbb{Z}_+$. The total-variation distance between P and Q is defined as*

$$d_{\mathrm{TV}}(P,Q) = \sup_{A\subseteq\mathbb{Z}_+}\Big(P(A)-Q(A)\Big)$$
$$= \frac{1}{2}\sum_{k=0}^{\infty}\big|P(k)-Q(k)\big|.$$

The second form is obviously symmetric. It is an exercise to show that the two forms are equivalent and that therefore

$$\sup_{A\subseteq\mathbb{Z}_+}(P(A)-Q(A)) = d_{\mathrm{TV}}(P,Q) = d_{\mathrm{TV}}(Q,P) = \sup_{A\subseteq\mathbb{Z}_+}(Q(A)-P(A))$$

Convergence of probabilities implies total variation distance converging to zero (exercise). Theorem 2.2 can be used to bound above $d_{\mathrm{TV}}(\mathrm{Bin}(n,p),\mathrm{Poi}(np))$. The following table compares the resulting bound with the bound we will obtain using the Stein-Chen method.

n	$d_{\mathrm{TV}}(\mathrm{Bin}(n,5/n),\mathrm{Poi}(5))$	Direct calculation bound	Stein-Chen bound
10	0.172	1.345	0.5
20	0.071	0.478	0.25
30	0.045	0.291	0.16666667
40	0.033	0.209	0.125
50	0.026	0.163	0.1

3.1. **Coupling.** Given two distributions P and Q, a coupling between P and Q is a joint distribution $(X,Y)\sim\mu$ such that marginally $X\sim P$ and $Y\sim Q$. We will use the word "coupling between X and Y" to mean "coupling between the distribution of X and the distribution of Y.

Consider the results of two unbiased coin flips: $X,Y\sim\mathrm{Ber}(1/2)$. How can they be coupled together? For any $\alpha\in[0,1]$, consider the following joint probabilities:

X\Y	0	1
0	$\frac{1}{2}\alpha$	$\frac{1}{2}(1-\alpha)$
1	$\frac{1}{2}(1-\alpha)$	$\frac{1}{2}\alpha$

- If $\alpha=1/2$, the random variables are independent.
- If $\alpha>1/2$, the random variables are positively correlated.
- If $\alpha<1/2$, the random variables are negatively correlated.

Note that $d_{\mathrm{TV}}(X,Y)=0$: X and Y can be coupled perfectly. This corresponds to the case $\alpha=1$. In fact we have the following result:

Lemma 3.2. *Let P,Q be two distributions on $\mathbb{Z}_+$. We have*

$$d_{\mathrm{TV}}(P,Q) = \inf_{(X,Y)\sim\mu}\mathbb{P}(X\neq Y)$$

where the infimum is over all coupling μ of P and Q.

Proof. It is easier to show that

$$P(A)-Q(A)\leqslant\mathbb{P}(X\neq Y)$$

for any set A and any coupling $(X,Y)\sim\mu$ of P and Q, by noting that

$$P(A)=\mathbb{P}(X\in A)=\mathbb{P}(X\in A, Y\notin A)+\mathbb{P}(X\in A, Y\in A)$$
$$\leqslant\mathbb{P}(X\neq Y)+\mathbb{P}(Y\in A)=\mathbb{P}(X\neq Y)+Q(A).$$

Thus we have

$$d_{\mathrm{TV}}(P,Q)\leqslant\inf_{(X,Y)\sim\mu}\mathbb{P}(X\neq Y).$$

To show the equality we have construct a coupling that achieves the lower bound. Thus we try to construct a joint distribution for X,Y that maximizes the probability that they are equal. Clearly, the best we can do is, for each value $k\in\mathbb{Z}_+$, to make $X=Y=k$ with probability $\min\{P(k),Q(k)\}$ and redistribute the remaining mass. Let $A_+ = \{k : P(k) > Q(k)\}$ and $A_- = \{k : P(k) < Q(k)\}$. Note that $A_+ \cap A_- = \emptyset$ and

$$\sum_{k \in A_+} (P(k) - Q(k)) = \sum_{k \in A_-} (Q(k) - P(k)) = \frac{1}{2} \sum_k |P(k) - Q(k)|.$$

Define

$$\theta := \frac{1}{2} \sum_k |P(k) - Q(k)|.$$

We define the joint distribution as follows:

$$\mathbb{P}(X = k, Y = l) = \begin{cases} \min\{P(k), Q(k)\} & \text{if } k = l \\ \frac{1}{\theta}(P(k) - Q(k))(Q(l) - P(l)) & \text{if } k \in A_+, l \in A_- \\ 0 & \text{otherwise.} \end{cases}$$

Note that for $k \notin A_+$ we have $\mathbb{P}(X = k) = \min\{P(k), Q(k)\} = P(k)$ and for $k \in A_+$

$$\mathbb{P}(X = k) = Q(k) + P(k) - Q(k) = P(k).$$

Similarly, $\mathbb{P}(Y = l) = Q(l)$ for all l. Thus (X, Y) is a coupling of P and Q. Moreover,

$$\mathbb{P}(X \neq Y) = \theta = \frac{1}{2} \sum_k |P(k) - Q(k)|.$$

This completes the proof. ∎

Thus, finding total variation distance is equivalent to finding the best coupling in terms of maximizing the "agreement probability". We will come back to coupling later in dealing with Stein-Chen method.

4. Sums of general independent Bernoulli random variables

We will now consider the case when $X_i \sim \mathrm{Ber}(p_i), i = 1, 2, \ldots, n$ are independent and

$$W := \sum_{i=1}^{n} X_i \text{ with } \lambda := \mathbb{E}(W) = \sum_{i=1}^{n} p_i.$$

The following calculation is extremely ugly and goes to show that we need a better method. (**This calculation is not needed for the subsequent developement.**)

Let $\sum'$ denote $\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}$. Then

$$\mathbb{P}(W=k) = \sum{}' \left(\prod_{j=1}^{k} p_{i_j}\right)\left(\prod_{j\notin\{i_1,\ldots,i_k\}}(1-p_j)\right) = \prod_{i=1}^{n}(1-p_i)\sum{}'\left(\prod_{j=1}^{k}\frac{p_{i_j}}{1-p_{i_j}}\right).$$

Thus

$$\mathbb{P}(\mathrm{Poi}(\lambda)=k) - \mathbb{P}(W=k) = \frac{\lambda^k}{k!}\left(e^{-\lambda} - \prod_{i=1}^{n}(1-p_i)\right) \tag{2}$$

$$+\prod_{i=1}^{n}(1-p_i)\left(\frac{\lambda^k}{k!} - \sum{}'\left(\prod_{j=1}^{k}p_{i_j}\right)\right) \tag{3}$$

$$+\prod_{i=1}^{n}(1-p_i)\sum{}'\left(\prod_{j=1}^{k}p_{i_j} - \prod_{j=1}^{k}\frac{p_{i_j}}{1-p_{i_j}}\right). \tag{4}$$

Looking first at (2):

$$\begin{aligned} 0 \leqslant e^{-\lambda} - \prod_{i=1}^{n}(1-p_i) &= e^{-\lambda}\left(1 - \prod_i (1-p_i)e^{p_i}\right) \\ &\leqslant e^{-\lambda}\sum_i (1-(1-p_i)e^{p_i}) \\ &\leqslant e^{-\lambda}\sum_i (1-(1-p_i^2)) = e^{-\lambda}\sum_i p_i^2 \leqslant \lambda e^{-\lambda}\max_i p_i. \end{aligned}$$

Hence

$$0 \leqslant (2) \leqslant \mathbb{P}(\mathrm{Poi}(\lambda)=k)\lambda \max_i p_i.$$

Secondly, writing $\lambda^k = \sum_{i_1=1}^{n}\cdots\sum_{i_k=1}^{n} p_{i_k}$ we see that

$$0 \leqslant \lambda^k - \sum{}'\left(\prod_{j=1}^{k}p_{i,j}\right) \leqslant \binom{k}{2}\sum_{i=1}^{n}p_i^2\left(\sum_{i=1}^{n}p_i\right)^{k-2} = \binom{k}{2}\lambda^{k-2}\sum_{i=1}^{n}p_i^2 \leqslant \binom{k}{2}\lambda^{k-1}\max_i p_i$$

so

$$0 \leqslant (3) \leqslant \mathbb{P}(\mathrm{Poi}(\lambda)=k)\binom{k}{2}\lambda^{-1}\max_i p_i.$$

Thirdly,

$$\begin{aligned} 0 \leqslant \sum{}'\left(\prod_{j=1}^{k}\frac{p_{i_j}}{1-p_{i_j}} - \prod_{j=1}^{k}p_{i_j}\right) &\leqslant \sum{}'\left(\prod_{j=1}^{k}\frac{p_{i_j}}{1-p_{i_j}}\right)\left(1-\prod_{j=1}^{k}(1-p_{i_j})\right) \\ &\leqslant \sum{}'\left(\prod_{j=1}^{k}\frac{p_{i_j}}{1-p_{i_j}}\right)\left(\sum_{j=1}^{k}p_{i_j}\right) \leqslant \sum{}'\left(\prod_{j=1}^{k}\frac{p_{i_j}}{1-p_{i_j}}\right)k\max_i p_i \end{aligned}$$

so

$$-k\mathbb{P}(W=k)\max_i p_i \leqslant (4) \leqslant 0.$$

Collecting the above gives:

Theorem 4.1. *If $W = X_1 + \cdots + X_n$, $X_i \sim \mathrm{Ber}(p_i)$ and $\lambda = \sum_i p_i$,*

$$|\mathbb{P}(W=k) - \mathbb{P}(\mathrm{Poi}(\lambda)=k)| \leqslant \max\left\{\left(\lambda + \frac{k(k-1)}{2\lambda}\right)\mathbb{P}(\mathrm{Poi}(\lambda)=k), k\mathbb{P}(W=k)\right\}\cdot\max_i p_i.$$

Corollary 4.2. *We have*

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) \leqslant \frac{5}{4}\lambda \max_i p_i.$$

Proof. Note that

$$\mathbb{E}\binom{\mathrm{Poi}(\lambda)}{k} = \lambda^k/k!.$$

Summing over k, we have

$$\sum_{k=0}^{\infty} |\mathbb{P}(W=k) - \mathbb{P}(\mathrm{Poi}(\lambda)=k)| \leqslant \left(\lambda + \frac{1}{2\lambda}\mathbb{E}(\widetilde{W}(\widetilde{W}-1)) + \mathbb{E}(W)\right) \cdot \max_i p_i = \frac{5}{2}\lambda \max_i p_i$$

where $\widetilde{W} \sim \mathrm{Poi}(\lambda)$. Recall that

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) = \frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}(W=k) - \mathbb{P}(\mathrm{Poi}(\lambda)=k)|.$$

The result follows. ■

Thus W is close to $\mathrm{Poi}(\lambda)$ in total-variation if $\lambda \max_i p_i$ is small. This holds if, for example $\max_i p_i \approx n^{-1/2}$. For independent summands, (4.1) can be improved upon using complex analysis. We will head in a different direction, because we want to be able to deal with dependent summands as to produce sharper results.

5. The Stein-Chen method

First we will explain the intuition behind Stein-Chen method and then we will go into the rigorous estimates and details in the next Section. To show that a random variable W is close to a Poisson distribution with mean λ we need to show that

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) := \sup_{A \subseteq \mathbb{Z}_+} (\mathbb{P}(W \in A) - \mathbb{P}(\mathrm{Poi}(\lambda) \in A))$$

is small. It is impossible to exactly evaluate the probability and bound the differences in probability directly, except in the independent case. The main idea is to write $\mathbb{P}(W \in A) - \mathbb{P}(\mathrm{Poi}(\lambda) \in A)$ as expectation of some other function which is easy to bound directly. First we will show that $\mathrm{Poi}(\lambda)$ distribution is **characterized** by the following operator (called the Stein characterizing operator)

$$\mathcal{A}_\lambda g(n) = \lambda g(n+1) - ng(n), n \geqslant 0$$

for $g : \mathbb{Z}_+ \to \mathbb{R}$, in the sense that a r.v. X is $\mathrm{Poi}(\lambda)$ iff $\mathbb{E}(\mathcal{A}_\lambda g(X)) = 0$ for all $g : \mathbb{Z}_+ \to \mathbb{R}$ bounded. This is the content of Lemma 5.1.

Lemma 5.1 (Stein characterizing equation for Poisson distribution)**.** *A non-negative integer valued random variable X is $\mathrm{Poi}(\lambda)$ if and only if*

$$\mathbb{E}(\lambda g(X+1) - Xg(X)) = 0$$

for every bounded function $g : \mathbb{Z}_+ \to \mathbb{R}$.

This suggest that if we want to show that a random variable W is approximately $\mathrm{Poi}(\lambda)$, we should show that $\mathbb{E}(\lambda g(W+1) - Wg(W)) \approx 0$ for bounded functions g. Our next goal is to **solve the equation** $\mathcal{A}_\lambda g = f$ for a given bounded function f with $\mathbb{E}(f(\widetilde{W})) = 0, \widetilde{W} \sim \mathrm{Poi}(\lambda)$. In Lemma 5.2 we will show that there is indeed a unique bounded solution.

Lemma 5.2. *If $f : \mathbb{Z}_+ \to \mathbb{R}$ satisfies $\mathbb{E}(f(\widetilde{W})) = 0$ where $\widetilde{W} \sim \mathrm{Poi}(\lambda)$ and is bounded then there is a unique bounded function g such that*

$$f(n) = \lambda g(n+1) - ng(n) \text{ for all } n \geqslant 0.$$

Moreover, g is given by $g(0) = 0$ and

$$g(n) = \frac{(n-1)!}{\lambda^n} \sum_{k=0}^{n-1} f(k) \frac{\lambda^k}{k!} = \frac{\mathbb{E}(f(\widetilde{W}); \widetilde{W} < n)}{n\mathbb{P}(\widetilde{W} = n)} \text{ for } n \geqslant 1.$$

The rest of the Stein-Chen method involves looking, for $A \subseteq \mathbb{Z}_+$ and $\lambda > 0$, at functions $g_{A,\lambda}$ defined as follows.

Definition 5.3 (The Stein approximating equation)**.** *Let $A \subset \mathbb{Z}_+$ and $\lambda > 0$. The function $g_{A,\lambda} : \mathbb{Z}_+ \to \mathbb{R}$ is defined as the unique bounded function satisfying*

$$\lambda g_{A,\lambda}(n+1) - ng_{A,\lambda}(n) = \mathbb{1}_{n \in A} - \mathbb{P}(\mathrm{Poi}(\lambda) \in A).$$

By Lemma 5.2 the function $g_{A,\lambda}$ exists and is unique. We will also prove that,

Theorem 5.4. *For all $A \subseteq \mathbb{Z}_+$,*

$$\sup_{n \geqslant 1} |g_{A,\lambda}(n+1) - g_{A,\lambda}(n)| \leqslant \min \left\{1, \frac{1}{\lambda}\right\}.$$

Now we have

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) := \sup_{A \subseteq \mathbb{Z}_+} \mathbb{E}(\mathcal{A}_\lambda g_{A,\lambda}(W))$$

and bounding the expectation (the main part of Stein-Chen method) will give the required bound. Let us now explain the last part in the case when

$$W = \sum_{i \in I} X_i, \quad X_i \sim \mathrm{Ber}(p_i), \quad \lambda = \sum_i p_i = \mathbb{E}(W), \quad \widetilde{W} \sim \mathrm{Poi}(\lambda).$$

Notice that the definition of $g_{A,\lambda}$ depends on A and λ but not on the distribution of W. We are trying to show that W is approximately Poisson. Putting W into the approximating equation,

$$\lambda g_{A,\lambda}(W+1) - Wg_{A,\lambda}(W) = \mathbb{1}_{W \in A} - \mathbb{P}(\widetilde{W} \in A).$$

Taking expecations,

$$\mathbb{E}(\lambda g_{A,\lambda}(W+1) - Wg_{A,\lambda}(W)) = \mathbb{P}(W \in A) - \mathbb{P}(\widetilde{W} \in A).$$

We therefore want to bound

$$\sup_{A\subseteq\mathbb{Z}_+} |\mathbb{E}(\lambda g_{A,\lambda}(W+1) - W g_{A,\lambda}(W))|.$$

5.1. **The independent case.** For starters, assume that the X_i are independent Bernoulli r.v.s. Let

$$V_i = W - X_i = \sum_{j\neq i} X_j.$$

Then

$$\mathbb{E}(X_i g_{A,\lambda}(W)) = \mathbb{E}(X_i g_{A,\lambda}(V_i + X_i)) = p_i \mathbb{E}(g_{A,\lambda}(V_i+1)).$$

Therefore

$$\begin{aligned}\mathbb{E}(\lambda g_{A,\lambda}(W+1) - W g_{A,\lambda}(W)) &= \sum_i \mathbb{E}(p_i g_{A,\lambda}(W+1) - X_i g_{A,\lambda}(W)) \\ &= \sum_i p_i \mathbb{E}(g_{A,\lambda}(W+1) - g_{A,\lambda}(V_i+1)).\end{aligned}$$

Note that for integers m, n, we have

$$|f(m) - f(n)| \leqslant \sup_k |f(k+1) - f(k)| \cdot |m-n|.$$

Hence

$$\begin{aligned}|\mathbb{P}(W\in A) - \mathbb{P}(\widetilde{W}\in A)| &= |\mathbb{E}(\lambda g_{A,\lambda}(W+1) - W g_{A,\lambda}(W))| \\ &\leqslant \sup_{n\geqslant 1} |g_{A,\lambda}(n+1) - g_{A,\lambda}(n)| \cdot \sum_i p_i \mathbb{E}|W - V_i| \\ &= \min\{1, 1/\lambda\} \cdot \sum_i p_i^2.\end{aligned}$$

Thus

$$\begin{aligned}d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) &\leqslant \min\{1, 1/\lambda\} \cdot \sum_i p_i^2 \\ &\leqslant \min\{\lambda, 1\} \cdot \max_i p_i.\end{aligned}$$

Compare this with the (4.2) bound of $\frac{5}{4}\lambda \max_i p_i$.

Example 5.5 (The binomial distribution)**.** *Applying this bound to the Binomial distribution gives*

$$d_{\mathrm{TV}}(\mathrm{Bin}(n,p), \mathrm{Poi}(np)) \leqslant \min\{np, 1\}p = \min\{p, np^2\}.$$

Example 5.6. $W = \sum_{i=1}^n X_i$ *with* $X_i \sim \mathrm{Ber}(i/n^2)$*. Then with* $\lambda = \binom{n+1}{2}/n^2$,

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) = \sum_{i=1}^n i^2/n^4 = \frac{n(n+1/2)(n+1)}{3n^4} = \frac{1}{3n}\left(1+\frac{1}{2n}\right)\left(1+\frac{1}{n}\right).$$

If the X_i's are not independent, the definition of the $V_i, i\in I$ has to be defined in terms of a coupling (see Section 3.1). Moreover, the coupling should be chosen to minimize $\mathbb{E}|W - V_i|$. Thus to apply the Stein-Chen method we will need to be able to find couplings that are as close as possible. Consider $X\sim \mathrm{Bin}(100,p)$, $Y \sim \mathrm{Bin}(101, p)$. A simple coupling is to set

$$Y := X + Z, \quad Z\sim \mathrm{Ber}(p), \quad X \perp Z.$$

Then $\mathbb{E}|X - Y| = \mathbb{E}(Z) = p$.

Definition 5.7 (Stein-Chen coupling)**.** *For $i\in I$, let (U_i, V_i) denote coupled random variables such that*

$$U_i \sim W, \quad 1 + V_i \sim W \mid X_i = 1.$$

U_i and V_i must be defined on the same probability space, but V_i and V_j $(i\neq j)$ don't need to be. The **main result of the course** is:

Theorem 5.8. *Let*

$$W = \sum_i X_i, \quad X_i\sim \mathrm{Ber}(p_i), \quad \lambda = \sum_i p_i = \mathbb{E}(W), \quad \widetilde{W}\sim \mathrm{Poi}(\lambda).$$

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

$$|\mathbb{P}(W\in A) - \mathbb{P}(\widetilde{W}\in A)| \leqslant \min\{1, 1/\lambda\} \sum_{i=1}^n p_i \mathbb{E}|U_i - V_i|.$$

The "quality" of the couplings (U_i, V_i) will determine the strength of the bound. Note that for independent summands we have $U_i = W$ and $V_i = \sum_{j\neq i} X_j \sim W - 1 \mid X_i = 1$.

Proof of Theorem 5.8. By the definition of $g_{A,\lambda}$,

$$\mathbb{P}(W \in A) - \mathbb{P}(\widetilde{W} \in A) = \mathbb{E}(\lambda g_{A,\lambda}(W+1) - W g_{A,\lambda}(W))$$

Since $\lambda = \sum_i p_i$,

$$\mathbb{E}(\lambda g_{A,\lambda}(W+1)) = \sum_i p_i \mathbb{E}(g_{A,\lambda}(W+1)) = \sum_i p_i \mathbb{E}(g_{A,\lambda}(U_i+1)).$$

Since $W = \sum_i X_i$,

$$\begin{aligned}
\mathbb{E}(W g_{A,\lambda}(W)) &= \sum_i \mathbb{E}(X_i g_{A,\lambda}(W)) \\
&= \sum_i p_i \mathbb{E}(X_i g_{A,\lambda}(W)|X_i = 1) + (1-p_i)\mathbb{E}(X_i g_{A,\lambda}(W)|X_i = 0) \\
&= \sum_i p_i \mathbb{E}(g_{A,\lambda}(W)|X_i = 1) = \sum_i p_i \mathbb{E}(g_{A,\lambda}(1+V_i)).
\end{aligned}$$

Thus

$$\begin{aligned}
|\mathbb{P}(W \in A) - \mathbb{P}(\widetilde{W} \in A)| &= \left| \sum_i p_i \mathbb{E}(g_{A,\lambda}(U_i+1) - g_{A,\lambda}(V_i+1)) \right| \\
&\leqslant \sup_{n\geqslant 1} |g_{A,\lambda}(n+1) - g_{A,\lambda}(n)| \times \sum_{i=1}^{n} p_i \mathbb{E}|U_i - V_i|.
\end{aligned}$$

Theorem 5.4 now completes the proof. ∎

6. Estimates and Proofs

Proof of Lemma 5.1. Suppose X is Poi(λ):

$$\begin{aligned}\mathbb{E}(\lambda g(X+1) - Xg(X)) &= \sum_{n=0}^{\infty} \lambda g(n+1) e^{-\lambda} \frac{\lambda^n}{n!} - \sum_{n=0}^{\infty} n g(n) e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \sum_{n=1}^{\infty} g(n) e^{-\lambda} \frac{\lambda^n}{(n-1)!} - \sum_{n=1}^{\infty} g(n) e^{-\lambda} \frac{\lambda^n}{(n-1)!}.\end{aligned}$$

Conversely, consider $g(x) = \mathbb{1}_{x=k}$. Then

$$\mathbb{E}(\lambda g(X+1) - Xg(X)) = \lambda \mathbb{P}(X = k-1) - k\mathbb{P}(X = k) = 0.$$

Hence for every $k \geqslant 1$,

$$\frac{\mathbb{P}(X=k)}{\mathbb{P}(X=k-1)} = \lambda/k = \frac{\mathbb{P}(\text{Poi}(\lambda)=k)}{\mathbb{P}(\text{Poi}(\lambda)=k-1)}.$$

■

Proof of Lemma 5.2. Let $g(0) = 0$; g is then defined inductively in terms of f by

$$g(n+1) = \frac{f(n) + ng(n)}{\lambda}.$$

Thus

$$g(1) = \frac{f(0)}{\lambda}, \quad g(2) = \frac{f(1)}{\lambda} + \frac{1 \cdot f(0)}{\lambda^2}, \quad g(3) = \frac{f(2)}{\lambda} + \frac{2f(1)}{\lambda^2} + \frac{2 \cdot 1 \cdot f(0)}{\lambda^3}, \quad \cdots$$

and in general, by induction (as $\mathbb{E}(f(\widetilde{W})) = 0$)

$$g(n) = \frac{(n-1)!}{\lambda^n} \sum_{k=0}^{n-1} f(k) \frac{\lambda^k}{k!} = -\frac{(n-1)!}{\lambda^n} \sum_{k=n}^{\infty} f(k) \frac{\lambda^k}{k!}$$

for $n \geqslant 1$. If $n > \lambda$,

$$\left| \frac{(n-1)!}{\lambda^n} \sum_{k=n}^{\infty} f(k) \frac{\lambda^k}{k!} \right| \leqslant \frac{\max_k |f(k)|}{n-\lambda}.$$

Thus g is bounded. The uniqueness of g is left as an exercise. ■

6.1. Solution to Stein's approximating equation. (Proof of 5.4)

To use the bound

$$|\mathbb{P}(W \in A) - \mathbb{P}(\widetilde{W} \in A)| \leqslant \sup_{n \geqslant 1} |g_{A,\lambda}(n+1) - g_{A,\lambda}(n)| \sum_i p_i \mathbb{E}|U_i - V_i|$$

we will solve Stein's approximating equation.

Lemma 6.1. *The solution to Stein's approximating equation can be written*

$$g_{A,\lambda}(n) = \frac{\mathbb{P}(\widetilde{W} \in A, \widetilde{W} < n) - \mathbb{P}(\widetilde{W} \in A)\mathbb{P}(\widetilde{W} < n)}{n\mathbb{P}(\widetilde{W} = n)} \quad n \geqslant 1.$$

Proof. The proof follows from Lemma 5.2 by taking $f(n) = \mathbb{1}_{n \in A} - \mathbb{P}(\widetilde{W} \in A)$. ■

Remark 6.2. *Note that* $g_{A,\lambda}(n) = \sum_{j \in A} g_{\{j\},\lambda}(n)$ *with*

$$g_{\{j\},\lambda}(n) = \frac{\mathbb{P}(\widetilde{W} = j)(1_{j<n} - \mathbb{P}(\widetilde{W} < n))}{n\mathbb{P}(\widetilde{W} = n)}$$

and

$$\sum_{j \in \mathbb{Z}_+} g_{\{j\},\lambda}(n) = g_{\mathbb{Z}_+,\lambda}(n) = 0.$$

Lemma 6.3. $|g_{A,\lambda}(n+1) - g_{A,\lambda}(n)| \leqslant g_{\{n\},\lambda}(n+1) - g_{\{n\},\lambda}(n)$.

Proof. Recall that

$$g_{\{j\},\lambda}(n) = \frac{\mathbb{P}(\widetilde{W}=j)(1_{j<n} - \mathbb{P}(\widetilde{W}<n))}{n\mathbb{P}(\widetilde{W}=n)}.$$

If $n \leqslant j$,

$$\begin{aligned} g_{\{j\},\lambda}(n) = \frac{-\mathbb{P}(\widetilde{W}=j)\mathbb{P}(\widetilde{W}<n)}{n\mathbb{P}(\widetilde{W}=n)} &= -\mathbb{P}(\widetilde{W}=j)\sum_{k=0}^{n-1}\lambda^{k-n}\frac{(n-1)!}{k!} \\ &= -\mathbb{P}(\widetilde{W}=j)\sum_{l=0}^{n-1}\frac{(n-1)\times\dots(n-l)}{\lambda^{l+1}} \end{aligned}$$

is negative and decreasing in n. If $n > j$,

$$\begin{aligned} g_{\{j\},\lambda}(n) = \frac{\mathbb{P}(\widetilde{W}=j)\mathbb{P}(\widetilde{W}\geqslant n)}{n\mathbb{P}(\widetilde{W}=n)} &= \mathbb{P}(\widetilde{W}=j)\sum_{k=n}^{\infty}\lambda^{k-n}\frac{(n-1)!}{k!} \\ &= \mathbb{P}(\widetilde{W}=j)\sum_{l=0}^{\infty}\frac{\lambda^l}{n(n+1)\times\dots(n+l)} \end{aligned}$$

is positive and decreasing in n. Plotting $g_{\{j\},\lambda}(n)$ against n: we see that $g_{\{j\},\lambda}(n+1) - g_{\{j\},\lambda}(n) > 0$ if

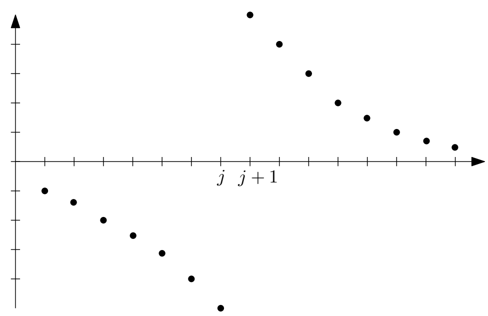


FIGURE 1. Plot of $g_{\{j\},\lambda}(\cdot)$

and only if $n = j$. As

$$\sum_{j=0}^{\infty}\Big(g_{\{j\},\lambda}(n+1) - g_{\{j\},\lambda}(n)\Big) = 0$$

the maximum of $|g_{A,\lambda}(n+1) - g_{A,\lambda}(n)|$ is obtained when $A = \{n\}$ or $A = \{n\}^c$. ∎

Lemma 6.4. $g_{\{n\},\lambda}(n+1) - g_{\{n\},\lambda}(n) \leqslant \frac{1}{\lambda}(1-e^{-\lambda})$.

Proof. Note that

$$\begin{aligned} g_{\{n\},\lambda}(n+1) - g_{\{n\},\lambda}(n) &= \frac{\mathbb{P}(\widetilde{W}=n)\mathbb{P}(\widetilde{W}>n)}{(n+1)\mathbb{P}(\widetilde{W}=n+1)} - \frac{-\mathbb{P}(\widetilde{W}=n)\mathbb{P}(\widetilde{W}<n)}{n\mathbb{P}(\widetilde{W}=n)} \\ &= \frac{1}{\lambda}\sum_{r=n+1}^{\infty} e^{-\lambda}\frac{\lambda^r}{r!} + \frac{1}{n}\sum_{r=0}^{n-1} e^{-\lambda}\frac{\lambda^r}{r!} \\ &= \frac{e^{-\lambda}}{\lambda}\left(\sum_{r=n+1}^{\infty}\frac{\lambda^r}{r!} + \sum_{r=1}^{n}\frac{\lambda^r}{r!}\frac{r}{n}\right) \leqslant \frac{e^{-\lambda}}{\lambda}\left(\sum_{r=0}^{\infty}\frac{\lambda^r}{r!} - 1\right) \leqslant \frac{1-e^{-\lambda}}{\lambda}. \end{aligned}$$

∎

Lemma 6.5. $\frac{1}{\lambda}(1-e^{-\lambda}) \leqslant \min\{1, 1/\lambda\}$.

Proof. Clearly, $1 - e^{-\lambda} < 1$. We also have $1 - \lambda < e^{-\lambda}$ or $1 - e^{-\lambda} < \lambda$. Thus $1 - e^{-\lambda} \leqslant \min(1,\lambda)$ or $\frac{1}{\lambda}(1-e^{-\lambda}) \leqslant \min\{1, 1/\lambda\}$. ∎

7. Dissociated random variables

Consider a sum of Bernoulli random variables with independence between some but not all of the random variables,

$$W = \sum_{i\in I} X_i, \quad X_i \sim \mathrm{Ber}(p_i), \quad \lambda = \sum_{i\in I} p_i.$$

We say that the indicator variables $(X_i : i \in I)$ are *dissociated* with respect to a collection of neighborhoods $(\mathcal{N}_i : i \in I)$ if for each $i \in I$,

$$X_i \text{ is independent of } \{X_j : j \notin \mathcal{N}_i \cup \{i\}\}.$$

We will assume that $i \notin \mathcal{N}_i$ for all i.

Theorem 7.1.

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) \leqslant \min\{1, 1/\lambda\} \sum_{i\in I} \Big(p_i^2 + \sum_{j\in\mathcal{N}_i} \Big(p_i p_j + \mathbb{E}(X_i X_j)\Big)\Big).$$

Proof. Let

$$U_i = W = \sum_{i\in I} X_i$$

and

$$V_i = \sum_{j\notin\mathcal{N}_i, j\neq i} X_j + \sum_{j\in\mathcal{N}_i} Y_j^{(i)}, \quad Y_j^{(i)} \sim X_j \mid X_i = 1$$

Then

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda) \leqslant \min\{1, 1/\lambda\} \sum_i p_i \mathbb{E}|U_i - V_i|.$$

with

$$\begin{aligned}
\mathbb{E}|U_i - V_i| &= \mathbb{E}\Big|X_i + \sum_{j\in\mathcal{N}_i} (X_j - Y_j^{(i)})\Big| \\
&\leqslant \mathbb{E}|X_i| + \sum_{j\in\mathcal{N}_i} \Big(\mathbb{E}|X_j| + \mathbb{E}|Y_j^{(i)}|\Big) \\
&= p_i + \sum_{j\in\mathcal{N}_i} \Big(p_j + \mathbb{E}(X_j \mid X_i = 1)\Big) = p_i + \sum_{j\in\mathcal{N}_i} \Big(p_j + \frac{\mathbb{E}(X_i X_j)}{p_i}\Big).
\end{aligned}$$

Theorem 5.8 now completes the proof. ∎

Example 7.2 (A Birthday problem)**.** *Suppose that there are 73 students taking a course, and each student has 10 friends. What is the probability that two friends share a birthday?*

Take the students birthdays to be independent and identically distributed on the set of days of the year (excluding February 29th),

$$Z_i \sim \mathit{Uniform}\{1, 2, \dots, 365\}.$$

The number of pairs of friends is

$$W = \sum_{\{i,j\}\ \mathit{friends}} X_{ij}, \quad X_{ij} = \begin{cases} 1 & \mathit{if}\ Z_i = Z_j, \\ 0 & \mathit{otherwise.} \end{cases}$$

For each pair of friends ij, $X_{ij} \sim \mathrm{Ber}(p_{ij})$ with $p_{ij} = 1/365$. The number of pairs of friends is $73 \times 10/2 = 365$ so the expected value of W is $\lambda = 365/365 = 1$.

The variables (X_{ij}) are dissociated with respect to the neighborhoods

$$\mathcal{N}_{ij} = \{kl : |\{i,j\} \cap \{k,l\}| = 1\}, \quad |\mathcal{N}_{ij}| = 9 + 9.$$

Thus

$$\begin{aligned}
d_{\mathrm{TV}}(W, \mathrm{Poi}(1)) &\leqslant \sum_{ij\ \textbf{friends}} \Big(p_{ij}^2 + \sum_{kl\in\mathcal{N}_{ij}} \Big(p_{ij} p_{kl} + \mathbb{E}(X_{ij} X_{kl})\Big)\Big) \\
&\leqslant 365 \times \Big(\frac{1}{365^2} + 18\Big(\frac{1}{365^2} + \frac{1}{365^2}\Big)\Big) = \frac{37}{365}.
\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{P}(W \geqslant 1) &\in \Big[\mathbb{P}(\mathrm{Poi}(1) \geqslant 1) - \frac{37}{365}, \mathbb{P}(\mathrm{Poi}(1) \geqslant 1) + \frac{37}{365}\Big] \\ &= \Big[1 - e^{-1} - \frac{37}{365}, 1 - e^{-1} + \frac{37}{365}\Big].\end{aligned}$$

Example 7.3 (Counting number of wedges in Erdös-Rènyi random graphs)**.** *An Erdös-Rényi random graph is formed on N vertices. Each unordered pair $\{i,j\}$ of vertices is* connected *with probability p, independently of all the other pairs.*

A wedge *(or path of length 2) is a tuple $(i,\{j,k\})$ where i,j,k are distinct and each of the edges $\{i,j\}$ and $\{i,k\}$ is connected. There are $N\binom{N-1}{2}$ many such tuples and each tuple forms an wedge (i.e. both connections are present) with probability p^2. Let W denote the number of wedges contained in the random graph. Clearly,*

$$\lambda := \mathbb{E}(W) = N\binom{N-1}{2}p^2.$$

Is W approximately $\mathrm{Poi}(\lambda)$*?*

Let $(Y_{ij} : 1 \leqslant i < j \leqslant N)$ be i.i.d. $\mathrm{Ber}(p)$ *random variables. The r.v. Y_{ij} is the indicator of whether the edge $\{i,j\}$ is connected. We will use ij for the edge $\{i,j\}$. We can write*

$$W = \sum_{i=1}^{N} \sum_{\substack{j<k,\\ j,k\neq i}} X_{i,jk}$$

where $X_{i,jk} = Y_{ij}Y_{ik} \sim \mathrm{Ber}(p^2)$. Here $p_{i,jk} = p^2$.

If we define

$$\mathcal{N}_{i,jk} = \{(a,\{b,c\}) : |\{ij,ik\} \cap \{ab,ac\}| = 1\}$$

as the set of tuples that share a common edge with the tuple (i,jk), then $X_{i,jk}$ is independent of all wedges not indexed by $\mathcal{N}_{i,jk} \cup \{(i,jk)\}$. Moreover, we have

$$|\mathcal{N}_{i,jk}| = 2(N-3) + 2(N-2) = 2(2N-5)$$

where the first counts tuples that also have the same center vertex i and the second counts tuples that have center vertex different from i. Thus

$$\begin{aligned}d_{\mathrm{TV}}(W, \mathrm{Poi}(1)) &\leqslant \min\{1, 1/\lambda\} \sum_{(i,jk)} \Big(p_{i,jk}^2 + \sum_{(a,bc)\in\mathcal{N}_{i,jk}} \Big(p_{i,jk}p_{a,bc} + \mathbb{E}(X_{i,jk}X_{a,bc})\Big)\Big) \\ &= \min\{1, 1/\lambda\} N\binom{N}{2}(p^4 + 2(2N-5)(p^4+p^3)) \\ &\leqslant \min\{\lambda, 1\}(p^2 + 8Np) \\ &= \min\{N^3p^2/2, 1\}(p + 8N)p.\end{aligned}$$

The final term converges to zero iff $Np \to 0$ as $N \to \infty$. Thus $W \approx \mathrm{Poi}(\lambda)$ when $Np \to 0$ as $N \to \infty$.

8. Positive-association

A collection $(X_i : i \in I)$ of Bernoulli random variables is *positively related* if for every $i \in I$, we can construct $(Y_j^{(i)} : j \neq i)$ coupled with X_i such that

$$(Y_j^{(i)} : j \neq i) \sim (X_j : j \neq i) \mid X_i = 1 \quad \text{and} \quad \forall j \neq i,\ Y_j^{(i)} \geqslant X_j.$$

Let

$$W = \sum_{i \in I} X_i, \quad X_i \sim \text{Ber}(p_i), \quad \lambda = \sum_{i \in I} p_i.$$

Theorem 8.1.

$$d_{\text{TV}}(W, \text{Poi}(\lambda)) \leqslant \min\{1, 1/\lambda\} \left(\text{Var}(W) - \lambda + 2 \sum_i p_i^2 \right).$$

Proof. Let

$$U_i = W, \quad V_i = \sum_{j \neq i} Y_j^{(i)} \sim W - 1 \mid X_i = 1.$$

Then

$$d_{\text{TV}}(W, \text{Poi}(\lambda) \leqslant \min\{1, 1/\lambda\} \sum_i p_i \mathbb{E}|U_i - V_i|$$

with

$$\begin{aligned} p_i \mathbb{E}|U_i - V_i| = p_i \mathbb{E}\Big| X_i + \sum_{j \neq i} X_j - Y_j^{(i)} \Big| &\leqslant p_i \mathbb{E}|X_i| + \sum_{j \neq i} p_i \mathbb{E}(Y_j^{(i)} - X_j) \\ &\leqslant \mathbb{E}(X_i)^2 + \sum_{j \neq i} \Big[\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \Big] \\ &\leqslant \sum_j \Big(\mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) \Big) - \mathbb{E}(X_i^2) + 2\mathbb{E}(X_i)^2 \end{aligned}$$

so

$$\sum_i p_i \mathbb{E}|U_i - V_i| = \text{Var}(W) - \lambda + 2 \sum_i p_i^2.$$

This completes the proof. ∎

Example 8.2 (Neighboring '1's on a circle)**.** *Let $Z_1, \ldots, Z_n$ denote a collection of* Ber(p) *random variables arranged in a circle. Think of Z_i as indicating that tube stations on London's circle line are out of service. How many neighboring pairs of stations are both out of service?*

Let $X_i, i = 1, \ldots, n$ denote the event $\{Z_i = Z_{i+1} = 1\}$ (indices modulo n). The X_i's are positively related, as conditional distribution of $(Z_1, \ldots, Z_n)$ given $X_i = 1$ is the same as the distribution of the vector $(Z_1, \ldots, Z_{i-1}, 1, 1, Z_{i+2}, \ldots, Z_n)$ and the indicators that two consecutive stations are out of service is bigger for $(Z_1, \ldots, Z_{i-1}, 1, 1, Z_{i+2}, \ldots, Z_n)$ than $(Z_1, \ldots, Z_n)$. Let W count the number of pairs of out-of-service stations,

$$W = \sum_{i=1}^n X_i.$$

The variance of W is

$$\begin{aligned} \text{Var}(W) &= \sum_{i=1}^n \Big(\mathbb{E}(X_i^2) + \sum_{j \neq i} \mathbb{E}(X_i X_j) \Big) - (\mathbb{E}W)^2 \\ &= \lambda + \sum_i \sum_{j \neq i} \mathbb{E}(X_i X_j) - n^2 p^4 \\ &= \lambda + n \left(2p^3 + (n-3)p^4 \right) - n^2 p^4 = \lambda + 2np^3 - 3np^4. \end{aligned}$$

Thus

$$\begin{aligned} d_{\text{TV}}(W, \text{Poi}(np^2)) &= \min\{1, 1/(np^2)\} \big[\text{Var}(W) - \lambda + 2 \sum_i p_i^2 \big] \\ &= \min\{1, 1/(np^2)\} \big[2np^3 - 3np^4 + 2np^4 \big] = \min\{np^2, 1\}(2-p)p. \end{aligned}$$

Example 8.3 (Treize)**.** *Take a regular pack of 52 cards. For simplicity, assume the cards are labelled 1,2,. . . ,13; with four cards of each type. Draw 13 cards.*

- *Let* $X_i = 1$ *if the i-th card is of type* i*; otherwise* 0*.*
- *Let* $W = \sum_{i=1}^{13} X_i$*.* $\lambda = \mathbb{E}(W) = 1$*. Is* W *approximately* $\mathrm{Poi}(1)$*?*

The X_i*'s are positively related. Construct* $U_i = X_1 + X_2 + \cdots + X_{13}$*, and* $1 + V_i = Y_1^{(i)} + Y_2^{(i)} + \cdots + Y_{13}^{(i)}$ *as follows: the* $Y_j^{(i)}$ *are obtained by sequentially drawing cards after having previously arranged for the i-th draw to have face value i (by secretly reserving one of the four cards of face value i for draw i). Now the* X_i *are obtained by swapping that secretly reserved card with a card drawn at random. Then* $Y_j^{(i)} \geqslant X_j$ *(since the swap may destroy a match and will not create any). Consequently we have a positively related situation, and noting that* $\lambda = 1$*, the Stein-Chen method yields*

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(1)) \leqslant \min\{1, 1\} \left(\mathrm{Var}(W) - 1 + 2 \sum_{i=1}^{13} \frac{1}{13^2} \right) = \frac{21}{221}.$$

since

$$\mathrm{Var}(W) = \sum_{i=1}^{n} \Big(\mathbb{E}(X_i^2) + \sum_{j \neq i} \mathbb{E}(X_i X_j) \Big) - (\mathbb{E}W)^2 = \lambda + \sum_i \sum_{j \neq i} \mathbb{E}(X_i X_j) - \lambda^2$$

$$= 13 \cdot (13 - 1) \cdot \left(\frac{1}{13} \times \frac{4}{51} \right) = \frac{16}{17}.$$

9. Negative-association

A collection $(X_i : i \in I)$ of Bernoulli random variables is *negatively related* if for every $i \in I$, we can construct $(Y_j^{(i)} : j \neq i)$ coupled with X_i such that

$$(Y_j^{(i)} : j \neq i) \sim (X_j : j \neq i) \mid X_i = 1 \quad \text{and} \quad \forall j \neq i,\ Y_j^{(i)} \leqslant X_j.$$

Let

$$W = \sum_{i \in i} X_i, \quad X_i \sim \mathrm{Ber}(p_i), \quad \lambda = \sum_{i \in I} p_i.$$

Theorem 9.1.

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) \leqslant \min\{1, 1/\lambda\}\Big(\lambda - \mathrm{Var}(W)\Big).$$

Proof. Let

$$U_i = W, \quad V_i = \sum_{j \neq i} Y_j^{(i)} \sim W - 1 \mid X_i = 1.$$

Then

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda) \leqslant \min\{1, 1/\lambda\} \sum_i p_i \mathbb{E}|U_i - V_i|$$

with

$$\begin{aligned} p_i\mathbb{E}|U_i - V_i| = p_i\mathbb{E}\Big|X_i + \sum_{j \neq i} X_j - Y_j^{(i)}\Big| &\leqslant p_i\mathbb{E}|X_i| + \sum_{j \neq i} p_i\mathbb{E}(X_j - Y_j^{(i)}) \\ &\leqslant \mathbb{E}(X_i)^2 + \sum_{j \neq i}\Big[\mathbb{E}(X_i)\mathbb{E}(X_j) - \mathbb{E}(X_iX_j)\Big] \\ &\leqslant \mathbb{E}(X_i^2) - \sum_j \Big(\mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\Big) \end{aligned}$$

so

$$\sum_i p_i\mathbb{E}|U_i - V_i| = \lambda - \mathrm{Var}(W).$$

■

Example 9.2 (The occupancy problem)**.** *Take m balls and n empty boxes. Place each ball in a box uniformly at random into one the boxes, with the choice made independently of the other balls. Let*

$$X_i = \mathbb{1}_{\textit{box } i \textit{ is empty}} \sim \mathrm{Ber}((1 - 1/n)^m), \quad W = \sum_i X_i.$$

W is the total number of empty boxes; $\lambda = \mathbb{E}(W) = n(1 - 1/n)^m$. The variables are negatively related: construct V_i from U_i by emptying box i and redistributing the balls it contained amongst the other boxes. We have

$$\mathrm{Var}(W) = \sum_{i=1}^n \mathbb{E}(X_i^2) + \sum_{i \neq j} \mathbb{E}(X_iX_j) - \mathbb{E}(W)^2 = \lambda + n(n-1)(1 - 2/n)^m - \lambda^2.$$

Thus we have

$$\begin{aligned} d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) &\leqslant \min(1, 1/\lambda)(n^2(1 - 1/n)^{2m} - n(n-1)(1 - 2/n)^m) \\ &\leqslant \min(\lambda, \lambda^2)\Big(1 - \Big(1 - \frac{1}{n}\Big)\Big(1 + \frac{1}{n(n-2)}\Big)^{-m}\Big) \end{aligned}$$

Now we have $1 + 1/n(n-2) \leqslant \exp(1/n(n-2))$ or $\Big(1 + \frac{1}{n(n-2)}\Big)^{-m} \geqslant e^{-m/n(n-2)}$. Also $1 - 1/n \geqslant e^{-1/(n-1)} \geqslant e^{-1/(n-2)}$ and $\lambda \leqslant ne^{-m/n}$. Thus

$$d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda)) \leqslant \min(\lambda, \lambda^2)\Big(1 - e^{-(m+n)/n(n-2)}\Big) \leqslant ne^{-m/n} \cdot \frac{m+n}{n(n-2)} \leqslant \frac{m+n}{n-2}e^{-m/n}.$$

Thus W is approximately $\mathrm{Poi}(\lambda)$ *whenever $m/n \to \infty$ as $n \to \infty$.*

Exercise sheet 1

(1) Show that for probability measures P, Q on $\mathbb{Z}_+$

$$\sup_{A \subseteq \mathbb{Z}_+} (P(A) - Q(A)) = \frac{1}{2} \sum_{k=0}^{\infty} |P(k) - Q(k)|.$$

(2) Describe an algorithm that uses a supply of Uniform$[0,1]$ random variables to sample two random variables X and Y such that

(1) X has distribution P,
(2) Y has distribution Q, and
(3) $\mathbb{P}(X = Y)$ is maximised.

(3) Show that $1 + x \leqslant e^x$ for $x \in \mathbb{R}$. Hint: consider the derivative of the function $(e^x - x - 1)$.

(4) Show that $\exp(-p/(1-p)) \leqslant 1 - p \leqslant \exp(-p)$ for $p \in (0,1)$. Hint: question 2 can be used to give both bounds.

(5) Show (without looking at your notes) that if $0 \leqslant k \leqslant n$,

$$\mathbb{P}(\mathrm{Poi}(np) = k) \exp\left(-\frac{\binom{k}{2}}{n-k+1} + \frac{(k-np)p}{1-p}\right) \leqslant \mathbb{P}(\mathrm{Bin}(n,p) = k)$$
$$\leqslant \mathbb{P}(\mathrm{Poi}(np) = k) \exp\left(-\frac{\binom{k}{2}}{n} + kp\right).$$

Use these inequalities to write an expression (an infinite sum) that bounds above $d_{\mathrm{TV}}(\mathrm{Bin}(n,p), \mathrm{Poi}(np))$.

(6) Show that if $\lambda = \sum_{i=1}^{n} p_i$,

$$\lambda^k - k! \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \left(\prod_{j=1}^{k} p_{i_j}\right) \leqslant \binom{k}{2} \sum_{i=1}^{n} p_i^2 \lambda^{k-2}.$$

(7) Show that

$$1 - \prod_{i=1}^{n} (1 - p_i) \leqslant \sum_{i=1}^{n} p_i.$$

(8) If $\widetilde{W} \sim \mathrm{Poi}(\lambda)$, show that

$$\lambda \mathbb{P}(\widetilde{W} = n) = (n+1) \mathbb{P}(\widetilde{W} = n+1).$$

(9) Show that if $\widetilde{W} \sim \mathrm{Poi}(\lambda)$ and $n \geqslant 1$, then

$$\frac{\mathbb{P}(\widetilde{W} < n)}{n \mathbb{P}(\widetilde{W} = n)} \leqslant \frac{n+1}{n(n+1-\lambda)} \text{ if } n > \lambda - 1.$$

(10) Show that if $\widetilde{W} \sim \mathrm{Poi}(\lambda)$ and $n \geqslant 1$, then

$$\frac{\mathbb{P}(\widetilde{W} < n)}{n \mathbb{P}(\widetilde{W} = n)} \leqslant \frac{1}{\lambda + 1 - n} \text{ if } n < \lambda + 1.$$

Exercise sheet 2

(1) Prove that if W is Poisson(λ), then $\mathbb{E}(\lambda g(W+1) - Wg(W)) = 0$ for every bounded function $g : \mathbb{Z}_+ \to \mathbb{R}$.

(2) Prove the converse to the previous exercise. Hint: Try $g(r) = 1$ if $r = n$ and $g(r) = 0$ if $r \neq n$.

(3) Show that if $\lambda > 0$ and $W \sim \mathrm{Poi}(\lambda)$ then

$$\mathbb{E}\left[\binom{W}{k}\right] = \frac{\lambda^k}{k!}, \quad k \in \mathbb{Z}_+$$

(4) Consider

$$W = \sum_{i \in I} X_i, \quad X_i \sim \mathrm{Ber}(p_i), \quad \lambda = \sum_i p_i = \mathbb{E}(W), \quad \widetilde{W} \sim \mathrm{Poi}(\lambda).$$

Define Stein couplings and Stein's estimating equation. Show that for $A \subset \mathbb{Z}_+$,

$$|\mathbb{P}(W \in A) - \mathbb{P}(\widetilde{W} \in A)| \leqslant \sup_{n \geqslant 1} |g_{A,\lambda}(n+1) - g_{A,\lambda}(n)| \sum_i p_i \mathbb{E}|U_i - V_i|.$$

(5) Use the bound

$$d_{\mathrm{TV}}(W, \widetilde{W}) \leqslant \min\left\{1, \frac{1}{\lambda}\right\} \sum_{i=1}^{n} p_i \mathbb{E}|U_i - V_i|.$$

to bound above $d_{\mathrm{TV}}(\mathrm{Bin}(n, 1/(2n)), \mathrm{Poi}(1/2))$ and $d_{\mathrm{TV}}(\mathrm{Bin}(n, 7/n), \mathrm{Poi}(7))$.

(6) There are $n - m$ careful people, each has an accident with probability p^2. There are m careless people, each has an accident with probability p. Incidence of accidents are independent. Then number of accidents is

$$W = \sum_{i=1}^{n} X_i, \quad X_i \sim \begin{cases} \mathrm{Ber}(p^2), & i = 1, \dots, n-m, \\ \mathrm{Ber}(p), & i = n-m+1, \dots, n. \end{cases}$$

Calculate $\lambda = \mathbb{E}(W)$ and an upper bound on $d_{\mathrm{TV}}(W, \mathrm{Poi}(\lambda))$.

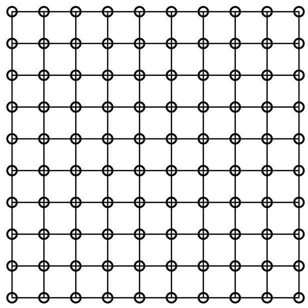
(7) A town is divided into $n \times n$ separate blocks based on a square grid. The town is surveyed after a mild earthquake, and blocks are marked if they contain any earthquake damage. Suppose that the probability of square (i, j) (for $i, j = 1, ..., n$) containing at least one building with earthquake damage is $ij/(4n^2)$. Supposing that different blocks are independent, bound $d_{\mathrm{TV}}(W, po(\lambda))$, where W is the number of blocks with earthquake damage, and

$$\lambda = \mathbb{E}(W) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{ij}{4n^2}.$$

Hint:

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \quad \text{and} \quad \sum_{i=1}^{n} i^2 = \frac{n(n+\frac{1}{2})(n+1)}{3}.$$

(1) Lighting Manhattan: Manhattan is built on a grid. Consider a 10×10 square grid. At each intersection



$\circ$ is a street light. Each of the 10^2 lights is *broken* with probability p; let independent random variables

$$Z_{ij} \sim \mathrm{Ber}(p), \quad i, j \in \{1, \ldots, 10\},$$

model which lights are broken. There are $180 = 2 \times 10 \times 9$ sections of roads (— and |) bounded on both sides by a lights. A section of road is *dark* if the lights at both ends are broken; each section of road is dark with probability p^2. Let W count the number of dark sections of road. Is W approximately Poisson?

(2) The changes of price of a stock are assumed to take the form $Z_i - 1/2$ were

$$Z_i \sim \mathrm{Ber}(1/2), i = 1, \ldots, 1009.$$

For $i \in \{1, \ldots, 1000\}$, let X_i indicate the event that starting on day i, the price decreased for five days, then increased for five days, i.e. $Z_i = \cdots = Z_{i+4} = 0$, $Z_{i+5} = \cdots = Z_{i+9} = 1$. Show that the X_i are negatively related. Is $W = \sum_i X_i$ approximately Poisson?

(3) (Example 7.3 revisited) An Erdös-Rènyi random graph $\mathcal{G}_n(p)$ has vertex set $1, \ldots, n$ and each edge $\{i, j\}$ is *open* (or *present*) with probability p, and *closed* (or *deleted*) with probability $1 - p$,

$$\mathscr{E} = \{\{i, j\} : X_{ij} = 1\}, \quad X_{ij} \sim \mathrm{Ber}(p).$$

A "path of length 2" is collection of 3 distinct vertices ijk (with $i < k$) such that E contains both $\{i, j\}$ and $\{j, k\}$. Let W count the number of paths of length 2 in the random graph. Find the mean and variance of W. Show how to use positive association of the indicator of wedges to find when W is approximately Poisson?

(4) Let W be the number of triangles in an Erdös-Rènyi random graph $\mathcal{G}_n(p)$, where a triangle is a set of three vertices $\{i, j, k\}$ such that each of the three edges $\{i, j\}, \{i, k\}, \{j, k\}$ is present. Find appropriate condition under which W is approximately Poisson distributed.

(5) Nearest-neighbor statistics: Independently choose points $Y_1, \ldots, Y_n$ uniformly at random in the unit square $[0, 1]^2$. Let $|Y_i - Y_j|$ denote the distance between Y_i and Y_j with respect to toroidal (periodic) boundary conditions. Let $r \in (0, 1)$ and let

$$X_{ij} = \begin{cases} 1 & \text{if } |Y_i - Y_j| < r, \\ 0 & \text{otherwise.} \end{cases}$$

Let $W = \sum_{i<j} X_{ij}$ denote the number of pairs at distance less than r. Is W approximately Poisson?

Exercise sheet 4

(1) An insurance company covers $n - m$ careful individuals and m careless individuals, where m is substantially smaller than n, and would like to estimate the number of claims these individuals are likely to make. Independently on any given day, each careful individual has probability p^2 of suffering an accident, while each careless individual has probability p of suffering an accident. Let W be the total number of individuals suffering an accident on a given day. Apply the Stein-Chen bounds in the case of independent Bernoulli random variables to estimate the total variation distance between W and a Poisson distribution of the same mean.

(2) Suppose X and Y are two variables defined on the same probability space, and that X is a Bernoulli random variable. Suppose further that Z is another random variable also defined on this probability space, having the distribution of Y conditional on $X = 1$. Show that

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Z).$$

(3) This is an adaptation of the DNA matching example. Given a target string of $n + k - 1$ independent binary symbols each equally likely to be 0 or 1, consider the distribution of the number W of matches of a fixed string of k binary symbols, chosen so that it is impossible for two different matches to overlap. Use the Stein-Chen approximation for negatively related summands to establish a Poisson approximation to W, together with error-estimate.

(4) In an investigation into suspected industrial espionage, suspicious fragmentary remains of an encrypted file are discovered on the hard disk of a company laptop. Investigators find that a binary string from the file, 40 bits long, exactly matches a substring of a highly sensitive encrypted company document which is 100000 bits in size. Use Stein-Chen Poisson approximation to estimate the probability of discovering one such match or more, if the laptop binary string is such that no two matches can overlap, and if the encrypted company document is modeled by 100000 independent bits independent of the laptop binary string, each equally likely to be 0 or 1 (this is a good model for strongly encrypted documents!).

(5) In a statistics lecture, there are 80 students. Each student is friends with 10 of the other students in attendance. Suppose that the birthdays of students are independently and uniformly distributed over the year (for simplicity, assume that there are exactly 400 days in each year). Apply the Stein-Chen approximation to show that there is a greater than 50% chance of there being at least one pair of friends in the lecture that share the same birthday.

(6) The Circle Line of the London Underground has $n \approx 30$ stations (arranged on a circle, of course!). Each station is closed for renovation independently with probability $p \approx 0.1$. How well does the Stein-Chen procedure approximate the distribution of W the number of pairs of stations in immediately neighboring locations that are both closed?