Lecture 24: Stein's method for Poisson approximation: preliminaries

As we have seen in Lectures 5 and 7, for some statistics of an Erdős-Rényi random graph where there is threshold behaviour, there is a critical window where the statistic converges weakly to a Poisson distribution. For example for X_n the number of isolated vertices in $G_n \sim \mathrm{ER}(n, (\log(n)+a)/n)$, for fixed $a \in \mathbb{R}$, then

$$\mathscr{L}(X_n) \xrightarrow[n\to\infty]{} \mathrm{Po}(e^{-a}).$$

The way we have been showing such results is to apply the following lemma that is a "method of moments" for indicators (Exercise 2.11 from Roch).

Lemma 24.1. *For a triangular array of events $(A_{n,i})_{1\le i\le n<\infty}$, if there is $\mu \ge 0$ such that, for every $r \ge 1$,*

$$\sum_{\{i_1,\dots,i_r\}\subseteq\{1,\dots,n\}} \mathbb{P}\big(\cap_{j=1}^r A_{n,i_j}\big) \to \frac{\mu^r}{r!}, \tag{24.1}$$

then

$$\mathscr{L}\Big(\sum_{i=1}^n \mathbb{I}[A_{n,i}]\Big) \xrightarrow[n\to\infty]{} \mathrm{Po}(\mu).$$

To understand this result, consider the case where for each n, the distribution of $(A_{n,i_j})_{1\le j\le r}$ is the same for every $\{i_1,\dots,i_r\} \subseteq \{1,\dots,n\}$; this is the case, for example, if $A_{n,i}$ is the event that vertex i is isolated in $G_n \sim \mathrm{ER}(n,p_n)$. In this case, (24.1) is the same as

$$\binom{n}{r}\mathbb{P}\big(\cap_{j=1}^r A_{n,j}\big) \xrightarrow[n\to\infty]{} \frac{\mu^r}{r!} \iff n^r\mathbb{P}\big(\cap_{j=1}^r A_{n,j}\big) \xrightarrow[n\to\infty]{} \mu^r. \tag{24.2}$$

Since the case $r=1$ is $n\mathbb{P}(A_{n,1}) \to \mu$, the convergence on the right hand side of (24.2) is $(A_{n,i})_{i=1}^r$ are independent, in Lecture 8 we used coupling to show that

$$d_{\mathrm{TV}}\Big(\Big(\sum_{i=1}^n \mathbb{I}[A_{n,i}]\Big), \mathrm{Po}\big(n\mathbb{P}(A_{n,1})\big)\Big) \le n\mathbb{P}(A_{n,1})^2, \tag{24.3}$$

which tends to zero at rate $1/n$ if $n\mathbb{P}(A_{n,1}) \to \mu$. So the basic idea of the condition (24.1), is that it roughly implies that the $(A_{n,i_j})_{1\le j\le r}$ are "close" to independent for any $\{i_1,\dots,i_r\} \subseteq \{1,\dots,n\}$ and large n, and that only finite number of the $A_{n,i}$'s occur. The meaning of "close" is quantified by (24.1).

Our next topic is a better method than (24.1) to show Poisson convergence. The main advantages are 1) in many situations the calculations involved will be no more difficult than those for establishing (24.1) for only $r = 1, 2$, and 2) automatic quantitative bounds such as that of (24.3).

Tool: Stein's method for Poisson approximation. Stein's method for Poisson approximation is a general technique for quantifying the approximation of a non-negative integer-valued random variable W with mean μ by a Poisson distribution with the same mean. In particular, letting $Y_\mu \sim \mathrm{Po}(\mu)$, we want to bound

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) = \sup_{A\subseteq\mathbb{Z}_+} \big|\mathbb{P}(W\in A) - \mathbb{P}(Y_\mu \in A)\big|.$$

The main idea behind the method is to replace the characteristic/moment-generating/probability-generating function (or moments, as in Lemma 24.1), which are classically used to show convergence in distribution, by a *characterising operator*, as per the next lemma.

Lemma 24.2. *Let W be a non-negative integer-valued random variable with mean μ and define the operator $\mathcal{A}_\mu$ on functions $f : \mathbb{Z}_+ \to \mathbb{R}$ by*

$$\mathcal{A}_\mu f(k) = \mu f(k+1) - k f(k).$$

Then $W \sim \mathrm{Po}(\mu)$ if and only if for all bounded $f : \mathbb{Z}_+ \to \mathbb{R}$,

$$\mathbb{E}\big[\mathcal{A}_\mu f(W)\big] = 0.$$

The operator $\mathcal{A}_\mu$ is called a characterising operator *for the Poisson(μ) distribution.*

Before proving the lemma, we record the following result and an important consequence.

Lemma 24.3. *If $A \subseteq \mathbb{Z}_+$ and we let $Y_\mu \sim \mathrm{Po}(\mu)$ and $U_k := \{0, 1, \dots, k-1\}$, then the function $f_A : \mathbb{Z}_+ \to \mathbb{R}$ defined by*

$$f_A(k) = \mu^{-k} e^\mu (k-1)! \big[\mathbb{P}(Y_\mu \in A \cap U_k) - \mathbb{P}(Y_\mu \in A)\mathbb{P}(Y_\mu \in U_k)\big],$$

satisfies

$$\mathcal{A}_\mu f_A(k) = \mathbb{I}[k \in A] - \mathbb{P}(Y_\mu \in A). \tag{24.4}$$

The last lemma easily gives the following key result.

Lemma 24.4. *Let $W \geq 0$ be an integer-valued random variable with mean μ, $Y_\mu \sim \mathrm{Po}(\mu)$, and f_A as in Lemma 24.3. Then*

$$\big|\mathbb{P}(W \in A) - \mathbb{P}(Y_\mu \in A)\big| = \big|\mathbb{E}\mathcal{A}_\mu f_A(W)\big| = \big|\mathbb{E}\big[\mu f_A(W+1) - W f_A(W)\big]\big|. \tag{24.5}$$

Before proceeding further, we make some remarks.

Remark 24.5. We will show that we can bound the right hand side of (24.5) uniformly in A, so that by taking the supremum over all subsets A of $\mathbb{Z}_+$ on the left hand side, we can use Lemma 24.4 to get an upper bound on $d_{\mathrm{TV}}(\mathscr{L}(W), \mathrm{Po}(\mu))$. The key difference between the left and right hand side of (24.5) is that there is no Y_μ on the right hand side. Thus we can bound it using only properties of W, rather than having to "find" something close to Y_μ in the structure of W, as we did to show (24.3).

Remark 24.6. While Lemma 24.2 is not strictly necessary to get to Lemma 24.4, it shows that (24.5) may give good bounds: the left hand side of (24.5) is zero for all A if and only if $W \sim \mathrm{Po}(\mu)$, and Lemma 24.2 implies that the right hand side has the same property. Otherwise put, we could try to implement the implicit strategy of the last two lemmas of solving (24.4) for an operator $\mathcal{A}_\mu$ that is not characterising and then taking expectation, but the upper bound would be bad as it may not be zero for $W \sim \mathrm{Po}(\mu)$.

Remark 24.7. Lemma 24.3 looks a bit strange because it is stated "backwards". In light of the previous two remarks, if we were working forward from the characterising operator to try to arrive at (24.5), we would want to solve (24.4). The statement of the lemma just cuts to the chase and says what the good solution is. The proof of the lemma is a straightforward exercise and thus is omitted.

Proof of Lemma 24.2. If $W \sim \mathrm{Po}(\mu)$, then

$$\begin{aligned}\mu\mathbb{E}[f(W+1)] &= \sum_{k\geq 0} e^{-\mu}\frac{\mu^{k+1}}{k!} f(k+1) \\ &= \sum_{k\geq 0} (k+1) f(k+1) e^{-\mu}\frac{\mu^{k+1}}{(k+1)!}\end{aligned}$$

$$= \sum_{k\geq 1} k f(k) e^{-\mu} \frac{\mu^k}{k!}$$
$$= \mathbb{E}[W f(W)],$$

as desired. If $\mu\mathbb{E}[f(W+1)] = \mathbb{E}[Wf(W)]$, for all bounded f, then we apply the identity to $f(k) = \mathbb{I}[k = j]$ for fixed $j \in \mathbb{Z}_+$ to find

$$\mu\mathbb{P}(W = j-1) = j\mathbb{P}(W = j) \implies \mathbb{P}(W = j) = (\mu/j)\mathbb{P}(W = j-1).$$

Iterating the latter equality, we find

$$\mathbb{P}(W = j) = \frac{\mu^j}{j!}\mathbb{P}(W = 0),$$

and since the probabilities have to sum to one, $\mathbb{P}(W = 0) = e^{-\mu}$ and $W \sim \mathrm{Po}(\mu)$. □

If you haven't seen any of this before, it is probably not clear how to work on the right hand side of (24.5). So let's do a simple example, where $W = \sum_{i=1}^n X_i$ for independent $X_i \sim \mathrm{Be}(p_i)$. Then for $f = f_A$ as above,

$$\mathbb{E}[Wf(W)] = \sum_{i=1}^n \mathbb{E}\Big[X_i f\Big(\sum_{j\neq i} X_j + X_i\Big)\Big] = \sum_{i=1}^n p_i \mathbb{E}\Big[f\Big(\sum_{j\neq i} X_j + 1\Big)\Big],$$

where the last equality is by taking the expectation with respect to X_i. We also have

$$\mu\mathbb{E}[f(W+1)] = \sum_{i=1}^n p_i \mathbb{E}\Big[f\Big(\sum_{j\neq i} X_j + X_i + 1\Big)\Big],$$

and so, again by taking expectation with respect to X_i, we find

$$\begin{aligned}\mathbb{E}[\mu f(W+1) - Wf(W)] &= \sum_{i=1}^n p_i \mathbb{E}\Big[f\Big(\sum_{j\neq i} X_j + X_i + 1\Big) - f\Big(\sum_{j\neq i} X_j + 1\Big)\Big], \\ &= \sum_{i=1}^n p_i^2 \mathbb{E}\Big[f\Big(\sum_{j\neq i} X_j + 2\Big) - f\Big(\sum_{j\neq i} X_j + 1\Big)\Big].\end{aligned}$$

Since we are thinking of $\sum_{j\neq i} X_j$ as being close to Poisson, there is no hope of using asymptotics or high probability properties to bound the expectation in the sum, so instead our hope is that we can uniformly bound $|f_A(k+1) - f_A(k)|$. In fact, we shall see in next lecture that

$$|f_A(k+1) - f_A(k)| \leq \frac{1-e^{-\mu}}{\mu} \leq \min\{1, \mu^{-1}\},$$

and so the work above with (24.5) implies that

$$\big|\mathbb{P}(W \in A) - \mathbb{P}(Y_\mu \in A)\big| \leq \min\{1, \mu^{-1}\} \sum_{i=1}^n p_i^2.$$

This is an improvement of the bound we obtained by direct coupling in Lecture 8 (similar to (24.3)) by the factor in front of the sum, which is already an advantage of the new method. But we will see another improvement is that very similar calculations can be performed and usefully applied for more general sums of indicators, for example, when looking at the number of isolated vertices, or the number of triangles in an Erdős-Rényi random graph.