Lecture 25: Stein's method for Poisson approximation

We are in the middle of developing Stein's method for Poisson approximation. From last lecture, we had the following result.

Lemma 24.3. *If $A \subseteq \mathbb{Z}_+$ and we let $Y_\mu \sim \mathrm{Po}(\mu)$ and $U_k := \{0, 1, \ldots, k-1\}$, 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

$$\mathcal{A}_\mu f_A(k) := \mu f_A(k+1) - k f_A(k) = \mathbb{I}[k \in A] - \mathbb{P}(Y_\mu \in A).$$

We used this lemma to bound the total variation distance between a sum of independent indicators and a Poisson distribution, assuming the functions f_A are such that $|f_A(k+1) - f_A(k)|$ is uniformly bounded. We now prove this assumption. For a function f with domain D, denote the supremum norm $\|f\| := \sup_{k \in D} |f(k)|$, and, for $f : \mathbb{Z}_+ \to \mathbb{R}$, denote the first difference $\Delta f(k) := f(k+1) - f(k)$.

Lemma 25.2. *If $A \subseteq \mathbb{Z}_+$ and f_A is as in Lemma 24.3, then*

$$\|\Delta f_A\| \leq \frac{1 - e^{-\mu}}{\mu} \leq \min\big\{1, \mu^{-1}\big\}.$$

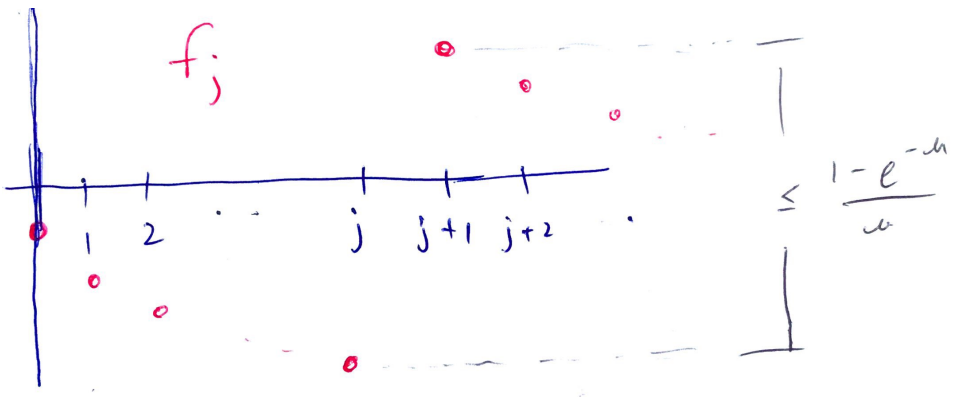
Proof. After rewriting

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

straightforward considerations imply that for $j \geq 1$, $f_j := f_{\{j\}}$ satisfies

- $f_j(k) \leq 0$ for $k \leq j$ and $f_j(k) \geq 0$ for $k > j$,
- $\Delta f_j(k) \leq 0$ for $k \neq j$, and $\Delta f_j(j) \geq 0$,
- $\Delta f_j(j) \leq \min\big\{j^{-1}, (1 - e^{-\mu})/\mu\big\}$.

And also $\Delta f_0(k) < 0$. The following picture illustrates for $j > 0$.



Since

$$\Delta f_A(k) = \sum_{j \in A} \Delta f_j(k)$$

is a sum of terms which are all negative except for at most one, we find

$$\Delta f_A(k) \leq \frac{1 - e^{-\mu}}{\mu}. \tag{25.1}$$

Since $f_{A^c} = -f_A$, (25.1) yields the second assertion. □

We can now adapt the argument from last lecture for sums of independent indicators to prove our main result.

Theorem 25.3. *Let $(X_i)_{i=1}^n$ be a collection of Bernoulli random variables with $\mathbb{P}(X_i = 1) = p_i$, and for each $i = 1, \dots, n$, let $(X_j^{(i)})_{j\neq i}$ be defined on the same space as $(X_i)_{i=1}^n$ such that*

$$\mathscr{L}\big((X_j^{(i)})_{j\neq i}\big) = \mathscr{L}\big((X_j)_{j\neq i} | X_i = 1\big).$$

Then for $W = \sum_{i=1}^n X_i$ and $\mu := \mathbb{E}[W] = \sum_{i=1}^n p_i$, we have

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) \leq \min\{1, \mu^{-1}\} \sum_{i=1}^n p_i \mathbb{E}\big|X_i + \sum_{j\neq i}(X_j - X_j^{(i)})\big|.$$

A quick remark before the proof: in the case that the X_i are independent, then we can set $(X_j^{(i)})_{j\neq i} = (X_j)_{j\neq i}$, and then we recover the bound from the previous lecture.

Proof. For $A \subseteq \mathbb{Z}_+$, let f_A be as in Lemma 24.3. Then for $Y_\mu \sim \mathrm{Po}(\mu)$, basic manipulations and the definition of $(X_j^{(i)})_{j\neq i}$ imply

$$\begin{aligned}
\mathbb{P}(W \in A) - \mathbb{P}(Y_\mu \in A) &= \mathbb{E}[\mu f(W+1) - W f(W)] \\
&= \sum_{i=1}^n p_i \mathbb{E}\Big[f\Big(\sum_{i=1}^n X_i + 1\Big)\Big] - \sum_{i=1}^n \mathbb{E}\Big[X_i f\Big(\sum_{j\neq i} X_j + X_i\Big)\Big] \\
&= \sum_{i=1}^n p_i \Big\{\mathbb{E}\Big[f\Big(\sum_{i=1}^n X_i + 1\Big)\Big] - \mathbb{E}\Big[f\Big(\sum_{j\neq i} X_j + 1\Big) | X_i = 1\Big]\Big\} \qquad (25.2) \\
&= \sum_{i=1}^n p_i \Big\{\mathbb{E}\Big[f\Big(\sum_{i=1}^n X_i + 1\Big) - f\Big(\sum_{j\neq i} X_j^{(i)} + 1\Big)\Big]\Big\}.
\end{aligned}$$

Now note that for any $i \leq j$, we can write

$$f(j) - f(i) = \sum_{k=i}^{j-1} \Delta f(k),$$

so that for any i, j,

$$|f(i) - f(j)| \leq \|\Delta f\| |i - j|.$$

Applying the triangle inequality, this last display, and the bounds on $\|\Delta f\|$ from Lemma 25.2 to (25.2) yields

$$\begin{aligned}
\big|\mathbb{P}(W \in A) - \mathbb{P}(Y_\mu \in A)\big| &\leq \sum_{i=1}^n p_i \mathbb{E}\Big|f\Big(\sum_{i=1}^n X_i + 1\Big) - f\Big(\sum_{j\neq i} X_j^{(i)} + 1\Big)\Big| \\
&\leq \|\Delta f\| \sum_{i=1}^n p_i \mathbb{E}\big|X_i + \sum_{j\neq i}(X_j - X_j^{(i)})\big| \\
&\leq \min\{1, \mu^{-1}\} \sum_{i=1}^n p_i \mathbb{E}\big|X_i + \sum_{j\neq i}(X_j - X_j^{(i)})\big|,
\end{aligned}$$

and taking the supremum over $A \subseteq \mathbb{Z}_+$ gives the theorem. □

As a first simple example of how to use this result, consider the case of isolated vertices. We showed in Lecture 7 using the method of moments, that if $X = X_n$ is the number of isolated vertices in $G_n \sim \mathrm{ER}(n, (s + \log(n))/n)$, where $s \in \mathbb{R}$ is fixed, then $\mathscr{L}(X_n) \xrightarrow[n\to\infty]{} \mathrm{Po}(e^{-s})$. We use Theorem 25.3 to show the following, which greatly improves on this convergence result.

Proposition 25.4. *Fix $n \in \mathbb{N}$ and $p \in [0,1]$ and let W be the number of isolated vertices in $G \sim \mathrm{ER}(n,p)$. Then $\mu := \mathbb{E}[W] = n(1-p)^{n-1}$, and*

$$d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) \le \min\{\mu, \mu^2\}\left(\frac{p}{1-p} + \frac{1}{n}\right).$$

Before the proof we remark the ways in which the proposition improves over the convergence via the method of moments. First, in the case where there is convergence ($p = (s + \log(n))/n$), the proposition gives a rate of convergence of $\log(n)/n$ in the total variation metric. More importantly, the bound is explicit for every n and p, which quantifies the quality of approximation in any finite problem, but also shows that for n large, the bound can be small even if there is no convergence, in particular, when $\mu \to \infty$. For example, if $p = c_n/n$ with with $c_n \to \infty$ and $\limsup_n c_n/\log(n) < 1$.

Proof. Write $W = \sum_{i=1}^n X_i$, where X_i is the indicator that vertex i is isolated in G. Then $\mathbb{P}(X_i = 1) = (1-p)^{n-1}$ and so $\mathbb{E}[W] = n(1-p)^{n-1}$, which is the first assertion. For the second, in order to apply Theorem 25.3, for each i, we need to construct $(X_j^{(i)})_{j\neq i}$, distributed as $\mathscr{L}\big((X_j)_{j\neq i} | X_i = 1\big)$, which are closely coupled to $(X_j)_{j\neq i}$. In order to do this, we construct a graph $G^{(i)}$ from G such that $\mathscr{L}(G^{(i)}) = \mathscr{L}(G|X_i = 1)$, and then for each $j \neq i$, we let $X_j^{(i)}$ be the indicator that vertex j is isolated in $G^{(i)}$. Because of the independence of the edges in G, it is clear that $\mathscr{L}(G|X_i = 1)$ is the same as having vertex i isolated, and then putting edges between the remaining vertices independently and with probability p. Thus we define $G^{(i)}$ by erasing any edges from G that are attached to vertex i. Now, Theorem 25.3 implies

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) &\le \min\big\{1, \mu^{-1}\big\} \sum_{i=1}^n (1-p)^{n-1} \mathbb{E}\Big[|X_i| + \sum_{j\neq i} |X_j^{(i)} - X_j|\Big] \\ &\le \min\big\{1, \mu^{-1}\big\} n(1-p)^{n-1}\big((1-p)^{n-1} + (n-1)\mathbb{P}(X_2^{(1)} \neq X_2)\big), \end{aligned} \tag{25.3}$$

where the second inequality is because of symmetry. To compute $\mathbb{P}(X_2^{(1)} \neq X_2)$, note that the event occurs only if $X_2^{(1)} = 1$ and $X_2 = 0$; this is because we only remove edges in the construction of $G^{(1)}$ from G, and so it is not possible to "de-isolate" a vertex. Since we only remove edges from vertex 1 when generating $G^{(1)}$ from G, the event $\{X_2^{(1)} = 1, X_2 = 0\}$ is the same as "vertex 2 has degree one and attaches to vertex 1 in G", which occurs with probability $p(1-p)^{n-2}$. Plugging this in to (25.3) yields

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathscr{L}(W), \mathrm{Po}(\mu)\big) &\le \min\{\mu, 1\}\big((1-p)^{n-1} + (n-1)p(1-p)^{n-2}\big) \\ &\le \min\{\mu^2, \mu\}\left(\frac{1}{n} + \frac{(n-1)p}{n(1-p)}\right), \end{aligned} \tag{25.4}$$

and the bound now easily follows. □

Note that the proof only uses "second moment" information; that is, we only need to understand the relationship between pairs of vertices, rather than larger collections as with the method of moments. Another feature in this example that simplified the proof is that $X_j^{(i)} \ge X_j$, and we will see next time that in this case in general, the bound of the theorem can be more simply written.