15.1 Stein's Method

Recall the following set-up:

- A finite index set $|I|< \infty$,
- Binary random variables $\{X_i\}_{i\in I}$,
- $p_i = P\{X_i = 1\}$,
- $W = \sum_{i\in I} X_i$,
- $\lambda = \mathbb{E}W = \sum_{i\in I} p_i$.
- A dependency graph for the random variables $\{X_i\}$.

We stated this theorem last lecture:

Theorem 15.1 *Let $\{X_i\}_{i\in I}$ be a collection of 0/1-valued random variables, and let (I, E) be a dependency graph for $\{X_i\}_{i\in I}$. Let $p_{ij} = P(X_i = X_j = 1)$, and let P_W denote the probability distribution of $W = \sum X_i$. Then*

$$\|P_W - \mathit{Poisson}(\lambda)\|_{TV} \leq \min(3, \lambda^{-1}) \left[\sum_{i\in I} \sum_{j\in N_i\setminus\{i\}} p_{ij} + \sum_{i\in I} \sum_{j\in N_i} p_i p_j \right].$$

This lecture will be spent proving Theorem 15.1.

We first present and prove a fact from analysis which we will use in the proof:

Lemma 15.2 *Let $P_\lambda(A)$ denote the probability that the value of a Poisson(λ) random variable falls in A. For every $A \subseteq \mathbb{N}$, there is a unique function $f : \mathbb{N} \to \mathbb{R}$ such that*

$$\lambda f(w + 1) - wf(w) = \delta_A(w) - P_\lambda(A)$$

for $w = 0, 1, 2, \ldots$. Moreover, $|f(w)|\leq 1.25$ and $|f(w + 1) - f(w)|\leq \min(3, \lambda^{-1})$.

Proof: Set $f(0) = 0$. Notice that the given equation dictates that

$$f(w + 1) = \frac{1}{\lambda}[wf(w) + \delta_A(0) - P_\lambda(A)],$$

so the equation determines $f(w)$ for all w once $f(0)$ is set. It remains to show the bounds on $|f(w)|$ and $|f(w + 1) - f(w)|$.

Multiplying both sides of the given recurrence by $\frac{\lambda^w}{w!}$:

$$\frac{\lambda^{w+1}}{w!}f(w+1) - \frac{\lambda^w}{(w-1)!}f(w) = \frac{\lambda^w}{w!}[\delta_A(w) - P_\lambda(A)].$$

Summing it up for $w = 0, \ldots, k-1$, the LHS is a telescoping series, and so we obtain

$$\frac{\lambda^k}{(k-1)!}f(k) = \sum_{w=0}^{k-1} \frac{\lambda^w}{w!}[\delta_A(w) - P_\lambda(A)]. \tag{15.1}$$

Since

$$\begin{aligned}\sum_{w=0}^{\infty} \frac{\lambda^w}{w!}[\delta_A(w) - P_\lambda(A)] &= \sum_{w\in A} \frac{\lambda^w}{w!} - P_\lambda(A)\sum_{w=0}^{\infty}\frac{\lambda^w}{w!} \\ &= e^\lambda P_\lambda(A) - P_\lambda(A)\cdot e^\lambda \\ &= 0,\end{aligned}$$

we also have

$$\frac{\lambda^k}{(k-1)!}f(k) = \sum_{w=k}^{\infty} \frac{\lambda^w}{w!}[\delta_A(w) - P_\lambda(A)]. \tag{15.2}$$

Because $|\delta_A(w) - P_\lambda(A)| \leq 1$, we can bound the RHS of Equation 15.1:

$$\begin{aligned}|f(k)| &= \frac{(k-1)!}{\lambda^k}\left|\sum_{w=0}^{k-1}\frac{\lambda^w}{w!}[\delta_A(w) - P_\lambda(A)]\right| \\ &\leq \frac{1}{\lambda}\sum_{w=0}^{k-1}\frac{(k-1)!}{\lambda^{k-1-w}w!} \\ &= \frac{1}{\lambda}\sum_{w=0}^{k-1}\frac{(k-1)!}{\lambda^w(k-1-w)!} \\ &\leq \frac{1}{\lambda}\sum_{w=0}^{\infty}\left(\frac{k-1}{\lambda}\right)^w \\ &= \frac{1}{\lambda}\frac{1}{1-\frac{k-1}{\lambda}} \\ &= \frac{1}{\lambda-(k-1)}.\end{aligned}$$

Similarly, we can bound the RHS of Equation 15.2:

$$\begin{aligned}
|f(k)| &\le \frac{(k-1)!}{\lambda^k}\sum_{w=k}^{\infty}\frac{\lambda^w}{w!} \\
&= \sum_{m=0}^{\infty}\frac{\lambda^m (k-1)!}{(k+m)!} \\
&\le \sum_{m=0}^{\infty}\frac{1}{k}\left(\frac{\lambda}{k+1}\right)^m \\
&= \frac{1}{k}\frac{1}{1-\frac{\lambda}{k+1}} \\
&= \frac{k+1}{k(k+1-\lambda)}.
\end{aligned}$$

Using the first bound for $k \le \lambda + \frac{1}{5}$ and the second bound for $k \ge \lambda + \frac{1}{5}$, we can conclude that that $|f(k)| \le 1.25$ for $k \ge 2$.

For $k = 1$, note that $f(1) = \frac{1}{\lambda}[\delta_A(0) - P_\lambda(A)]$. The RHS is largest when $A = \{0\}$, and smallest when $A = \{1, 2, \dots\}$. In either case,

$$|f(1)| \le \frac{1}{\lambda}(1 - e^{-\lambda}) \le 1.$$

To bound $|f(w+1) - f(w)|$, it is clear that

$$|f(w+1) - f(w)| \le |f(w+1)| + |f(w)| \le 2 \times 1.25 < 3.$$

As part of homework, we will show that $|f(w+1) - f(w)| \le \lambda$ for $\lambda \ge 1/3$. Putting the bounds together, we get $|f(w+1) - f(w)| \le \min(3, \lambda^{-1})$.

■

The key in all applications of Stein's method is the **Stein equation**:

Theorem 15.3 (Stein's Equation) *$Z \sim$ Poisson(λ) if and only if for every $f : \mathbb{N} \to \mathbb{R}$ bounded,*

$$\mathbb{E}\{\lambda f(Z+1) - Zf(Z)\} = 0. \tag{15.3}$$

Proof: Assume that $Z \sim$ Poisson(λ). Since Equation 15.3 is linear in f, it suffices to prove that Equation 15.3 holds for point delta functions, i.e. $f(j) = \delta_a(j)$ for some point a. In that case,

$$\begin{aligned}
\mathbb{E}\{\lambda f(Z+1) - Zf(Z)\} &= (\lambda - 0)P\{Z = a-1\} + (0-a)P\{Z = a\} \\
&= \lambda\frac{e^{-\lambda}\lambda^{a-1}}{(a-1)!} - a\frac{e^{-\lambda}\lambda^a}{a!} \\
&= 0.
\end{aligned}$$

In the opposite direction, assume that $\mathbb{E}\{\lambda f(W+1) - Wf(W)\} = 0$ for all bounded f. Pick any set A, and choose f as in Lemma 15.2, i.e. such that $\lambda f(w+1) - wf(w) = \delta_A(w) - P_\lambda(A)$. Then

$$\begin{aligned}
0 &= \mathbb{E}\{\lambda f(W+1) - Wf(W)\} \\
&= \mathbb{E}[\delta_A(W) - P_\lambda(A)] \\
&= P\{W \in A\} - P_\lambda(A),
\end{aligned}$$

i.e. $W \sim \text{Poisson}(\lambda)$. ■

We are now in a position to prove Theorem 15.1.

Proof:[of Theorem 15.1]

Take any $A \subseteq \mathbb{N}$, pick f as in Lemma 15.2. Then

$$\begin{aligned}
P_\lambda(A) - P\{W \in A\} &= \mathbb{E}(Wf(W) - \lambda f(W+1)) \\
&= \sum_{i \in I} \mathbb{E}\left[X_i f(W) - p_i f(W+1)\right] \\
&=: \Delta.
\end{aligned}$$

Set $W_i = W - X_i$, and $V_i = \sum_{j \in N_i^c} X_j$. By our assumptions about the dependency graph, X_i is independent of V_i. Also,

$$X_i f(W) = \begin{cases} 0 & \text{if } X_i = 0 \\ f(W_i + 1) & \text{if } X_i = 1 \end{cases} = X_i f(W_i + 1).$$

Hence,

$$\begin{aligned}
\Delta &= \sum_i \mathbb{E}[(X_i - p_i) f(W_i + 1) + p_i (f(W_i + 1) - f(W + 1))] \\
&= \sum_i \mathbb{E}[(X_i - p_i) f(W_i + 1)] + \sum_i \mathbb{E}[p_i (f(W_i + 1) - f(W + 1))] \\
&= \sum_i \mathbb{E}[(X_i - p_i)(f(W_i + 1) - f(V_i + 1))] + \sum_i \mathbb{E}[p_i (f(W_i + 1) - f(W + 1))] \\
&=: (I) + (II).
\end{aligned}$$

If $X_i = 0$, we have $f(W_i+1) = f(W+1)$. If $X_i = 1$, Lemma 15.2 gives us the bound $|f(W_i+1) - f(W+1)| \leq \min(3, \lambda^{-1})$. Hence, $|f(W_i + 1) - f(W + 1)| \leq \min(3, \lambda^{-1}) X_i$, so we can bound (II):

$$|(II)| \leq \sum_{i \in I} p_i \mathbb{E}[\min(3, \lambda^{-1}) X_i] = \min(3, \lambda^{-1}) \sum_{i \in I} p_i^2.$$

To bound (I), if we set $N_i \setminus \{i\} = \{X_1', \dots X_m'\}$, then

$$\begin{aligned}
|f(W_i + 1) - f(V_i + 1)| &= \left| f\left(V_i + \sum_{j \in N_i \setminus \{i\}} X_i + 1 \right) - f\left(V_i + 1 \right) \right| \\
&= \left| f\left(V_i + \sum_{j=1}^{m} X_i' + 1 \right) - f\left(V_i + 1 \right) \right| \\
&\leq \sum_{j=1}^{m} \left| f\left(V_i + \sum_{k=1}^{j} X_i' + 1 \right) - f\left(V_i + \sum_{k=1}^{j-1} X_i' + 1 \right) \right| \\
&\leq \sum_{j=1}^{m} \min(3, \lambda^{-1}) X_j' \\
&= \min(3, \lambda^{-1}) \sum_{j \in N_i \setminus \{i\}} X_j,
\end{aligned}$$

so

$$
\begin{aligned}
|(I)| &\le \min(3,\lambda^{-1})\sum_{i\in I}\mathbb{E}\left[|X_i - p_i|\sum_{j\in N_i\setminus\{i\}} X_j\right] \\
&\le \min(3,\lambda^{-1})\sum_{i\in I}\sum_{j\in N_i\setminus\{i\}}(\mathbb{E}[X_iX_j] + p_i\mathbb{E}X_j) \\
&= \min(3,\lambda^{-1})\sum_{i\in I}\sum_{j\in N_i\setminus\{i\}}(p_{ij} + p_ip_j).
\end{aligned}
$$

Putting the bounds for (I) and (II) together, we get

$$
\begin{aligned}
P_\lambda(A) - P\{W\in A\} &\le \min(3,\lambda^{-1})\sum_{i\in I}p_i^2 + \min(3,\lambda^{-1})\sum_{i\in I}\sum_{j\in N_i\setminus\{i\}}(p_{ij}+p_ip_j) \\
&= \min(3,\lambda^{-1})\left[\sum_{i\in I}\sum_{j\in N_i\setminus\{i\}}p_{ij} + \sum_{i\in I}\sum_{j\in N_i}p_ip_j\right].
\end{aligned}
$$

■

Note:

There are 3 versions of Stein's Method for Poisson approximation:

1. Dependency graphs (Chen/Stein method, which we have done in class),
2. Method of exchangeable pairs,
3. Size biased coupling (Barbour's method).

15.2 Gaussian Heuristic

Let $\{X_i\}_{i\in I}$ be real-valued random variables. The Gaussian Heuristic can be stated as follows:

If I large and the X_i are "not too dependent" and "not too wild", then $W = \sum_{i\in I} X_i$ has an approximate $\mathcal{N}(\mu,\sigma^2)$ distribution, where $\mu = \mathbb{E}W$, $\sigma^2 = \text{Var } W$.

Example: Say we have an $n\times n$ grid. At each grid site (i,j) $(1\le i,j\le n)$, put a uniform random variable $U_{ij}\sim \text{Unif}(0,1)$, with the U_{ij}'s iid. Let W be the number of local maxima (i.e. bigger than its neighbors). If we let

$$
X_{ij} = \begin{cases} 1 & \text{if } U_{ij} \text{ is a local maxima,} \\ 0 & \text{otherwise,} \end{cases}
$$

then the Gaussian heuristic says taht W has an approximate normal distribution.