Lecture 2

The Chen Stein method

Recall that given $A \subset \mathbb{N}$ and $\lambda \in (0, \infty)$, Stein's equation 1.1 looks for a function $g_A : \mathbb{N} \to \mathbb{R}$ so that $g_A(0) = 0$ and

$$\lambda g_A(k+1) - k g_A(k) = \chi_A(k) - \mathbb{E}\left[\chi_A(Z_\lambda)\right] \text{ for all } k \in \mathbb{N}. \tag{2.1}$$

2.1 Bounds on Stein's equation

Furthermore, recall the following bound on partial sums in Newton's binomial theorem:

Lemma 1.9. *Let $r, s \in \mathbb{N}$ so that $2r \leq s$. Then:*

$$\sum_{i=0}^{r} \binom{s}{i} \leq \frac{s-r+1}{s-2r+1} \cdot \binom{s}{r}.$$

We now have the following bound on $||g_A||$:

Proposition 1.10. *Let $A \subset \mathbb{N}$. Then*

$$||g_A|| \leq 1.$$

Proof. To simplify matters, we define a new function $f : \mathbb{N} \to \mathbb{R}$ by

$$f(k) = \chi_A(k) - \mathbb{E}\left[\chi_A(Z_\lambda)\right]$$

for all $k \in \mathbb{N}$. Note that by definition

$$\mathbb{E}\left[f(Z_\lambda)\right] = 0.$$

Set $g_A(0) = 0$. From (1.1) we obtain that for all $k \in \mathbb{N}$:

$$g_A(k+1) = \frac{1}{\lambda} f(k) + \frac{k}{\lambda} g_A(k).$$

Hence

$$g_A(k+1) = \frac{1}{\lambda} \sum_{j=0}^{k} \frac{k(k-1)\cdots(k-j+1)}{\lambda^j} f(k-j) = \frac{k!}{\lambda^{k+1}} \sum_{i=0}^{k} \frac{\lambda^i}{i!} f(i).$$

Thus

$$g_A(k+1) = \frac{1}{\lambda \cdot \mathbb{P}\left[Z_\lambda = k\right]} \sum_{j=0}^{k} \mathbb{P}\left[Z_\lambda = j\right] f(j),$$

where $Z_\lambda : \Omega \to \mathbb{N}$ is a Poisson variable with mean λ. Filling in definition of f we get

$$\chi_{[0,k]}(Z_\lambda) f(Z_\lambda) = \chi_{A \cap [0,k]}(Z_\lambda) - \chi_{[0,k]}(Z_\lambda) \mathbb{P}\left[Z_\lambda \in A\right].$$

To shorten notation, let us write:

$$p_\lambda(B) = \mathbb{P}\left[Z_\lambda \in B\right]$$

for all $B \subset \mathbb{N}$ and $U_k = [0,k] \cap \mathbb{N}$. We get

$$\begin{aligned}
\mathbb{E}\left[\chi_{U_k}(Z_\lambda) f(Z_\lambda)\right] &= \mathbb{E}\left[\chi_{A \cap U_k}(Z_\lambda)\right] - \mathbb{E}\left[\chi_{U_k}(Z_\lambda)\right] p_\lambda(A) \\
&= p_\lambda(A \cap U_k) - p_\lambda(U_k) p_\lambda(A) \\
&= p_\lambda(A \cap U_k) \cdot p_\lambda(\mathbb{N} \setminus U_k) - p_\lambda(A \setminus U_k) \cdot p_\lambda(U_k).
\end{aligned}$$

So we obtain

$$g_A(k+1) = \frac{p_\lambda(A \cap U_k) \cdot p_\lambda(\mathbb{N} \setminus U_k) - p_\lambda(A \setminus U_k) \cdot p_\lambda(U_k)}{\lambda \cdot p_\lambda(k)}.$$

Hence

$$
\begin{aligned}
|g_A(k+1)| &\leq \frac{\max\{p_\lambda(A\cap U_k)\cdot p_\lambda(\mathbb{N}\setminus U_k), p_\lambda(A\setminus U_k)\cdot p_\lambda(U_k)\}}{\lambda\cdot p_\lambda(k)} \\
&\leq \frac{p_\lambda(U_k)\cdot p_\lambda(\mathbb{N}\setminus U_k)}{\lambda\cdot p_\lambda(k)}.
\end{aligned}
$$

Filling in the Poisson probabilities, we obtain:

$$
\begin{aligned}
|g_A(k+1)| &\leq \frac{k!\cdot e^{-\lambda}}{\lambda^{k+1}}\cdot\sum_{i=0}^{k}\frac{\lambda^i}{i!}\sum_{j=k+1}^{\infty}\frac{\lambda^j}{j!} \\
&= k!\cdot e^{-\lambda}\cdot\sum_{i=0}^{k}\frac{\lambda^i}{i!}\sum_{j=0}^{\infty}\frac{\lambda^j}{(j+k+1)!}
\end{aligned}
$$

Now we reorder the terms and get:

$$
\begin{aligned}
|g_A(k+1)| &\leq k!\cdot e^{-\lambda}\cdot\sum_{n=0}^{\infty}\lambda^n\sum_{i=0}^{\min\{n,k\}}\frac{1}{i!(n+k+1-i)!} \\
&= k!\cdot e^{-\lambda}\cdot\sum_{n=0}^{\infty}\frac{\lambda^n}{(n+k+1)!}\sum_{i=0}^{\min\{n,k\}}\binom{n+k+1}{i}.
\end{aligned}
$$

Note that $2\cdot\min\{n,k\} < n+k+1$ for all $n,k\in\mathbb{N}$, so Lemma 1.9 applies. Hence we get:

$$
|g_A(k+1)| \leq k!\cdot e^{-\lambda}\cdot\sum_{n=0}^{\infty}\frac{\lambda^n\cdot\binom{n+k+1}{\min\{n,k\}}}{(n+k+1)!}\frac{n+k+2-\min\{n,k\}}{n+k+2-2\cdot\min\{n,k\}}.
$$

A straightforward computation shows that

$$
\frac{\binom{n+k+1}{n}}{(n+k+1)!} = \frac{1}{n!(k+1)!} \quad\text{and}\quad \frac{\binom{n+k+1}{k}}{(n+k+1)!} = \frac{1}{(n+1)!k!}.
$$

This implies that

$$
\begin{aligned}
|g_A(k+1)| &\leq e^{-\lambda}\cdot\left(\sum_{n=0}^{k}\frac{\lambda^n}{n!}\frac{k+2}{(k+1)(k-n+2)}\right. \\
&\qquad\left.+\sum_{n=k+1}^{\infty}\frac{\lambda^n}{n!}\frac{n+2}{(n+1)(n-k+2)}\right) \\
&\leq e^{-\lambda}\cdot e^{\lambda} \\
&= 1.
\end{aligned}
$$

□

2.2 Approximation theorems

We are now ready to put all the above together into concrete approximation theorems. The type of random variables we will be considering later on are counting variables. In particular, they will be variables that are obtained as the sum of (not necessarily mutually independent) Bernoulli variables.

To this end, let $\mathcal{I}$ be a set and let $X_i : \Omega \to \mathbb{N}$ be a Bernoulli variable with

$$\mathbb{E}\left[X_i\right] = \mathbb{P}\left[X_i = 1\right] = p_i$$

for all $i \in \mathcal{I}$. We will be interested in approximating the random variable

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

with a Poisson variable. To this end, set

$$p_{ij} = \mathbb{E}\left[X_i X_j\right]$$

and, given $i \in \mathcal{I}$, define

$$\mathcal{D}_i = \left\{ j \in \mathcal{I};\ j \neq i,\ X_i \text{ and } X_j \text{ not independent} \right\}.$$

Given this data, we define the following three quantities

$$B_1 = \sum_{i \in \mathcal{I}} p_i^2, \quad B_2 = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{D}_i} p_i p_j \text{ and } B_3 = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{D}_i} p_{ij}.$$

The first approximation theorem we state is the following:

Theorem 2.1. *Let W be as above such that $\lambda = \mathbb{E}\left[W\right] \in (0, \infty)$. Furthermore, let $Z_\lambda : \Omega \to \mathbb{N}$ be Poisson distributed random variable with mean λ. Then*

$$\mathrm{d}_{\mathrm{TV}}\left(W, Z_\lambda\right) \leq 2 \cdot (B_1 + B_2 + B_3).$$

Proof. Recall from Theorem 1.8 that we need to bound $\mathbb{E}\left[\lambda g_A(W+1) - W g_A(W)\right]$ for sets $A \subset \mathbb{N}$. Our first step will be to find a convenient decomposition of this quantity. First note that

$$\begin{aligned}
\mathbb{E}\left[\lambda g_A(W+1) - W g_A(W)\right] &= \sum_{i\in\mathcal{I}} \mathbb{E}\left[\mathbb{E}\left[X_i\right] g_A(W+1) - X_i g_A(W)\right] \\
&= \sum_{i\in\mathcal{I}} \mathbb{E}\left[X_i\right] \mathbb{E}\left[g_A(W+1)\right] - \mathbb{E}\left[X_i g_A(W)\right]
\end{aligned}$$

Given $i \in \mathcal{I}$, define two new random variables $S_i, T_i : \Omega \to \mathbb{N}$ by

$$S_i = \sum_{j\in\mathcal{I}\setminus(\mathcal{D}_i\cup\{i\})} X_j \ \text{ and } \ T_i = \sum_{j\in\mathcal{D}_i} X_j$$

and observe that

$$W = S_i + T_i + X_i$$

for all $i \in \mathcal{I}$. Notice that

$$\mathbb{E}\left[X_i g_A(W)\right] = \mathbb{E}\left[X_i g_A(S_i + T_i + X_i)\right] = \mathbb{E}\left[X_i g_A(S_i + T_i + 1)\right].$$

Moreover

$$\begin{aligned}
\mathbb{E}\left[X_i g_A(W)\right] &= \mathbb{E}\left[X_i\right] \mathbb{E}\left[g_A(S_i+1)\right] + \mathbb{E}\left[(X_i - \mathbb{E}\left[X_i\right]) \cdot g_A(S_i+1)\right] \\
&\quad + \mathbb{E}\left[X_i(g_A(S_i + T_i + 1) - g_A(S_i+1))\right].
\end{aligned}$$

Independence of X_i and S_i implies that

$$\mathbb{E}\left[(X_i - \mathbb{E}\left[X_i\right]) \cdot g_A(S_i+1)\right] = \mathbb{E}\left[X_i - \mathbb{E}\left[X_i\right]\right] \cdot \mathbb{E}\left[g_A(S_i+1)\right] = 0.$$

So we obtain

$$\begin{aligned}
\mathbb{E}\left[\lambda g_A(W+1) - W g_A(W)\right] &= \sum_{i\in\mathcal{I}} \quad \mathbb{E}\left[X_i\right] \mathbb{E}\left[g_A(W+1)\right] - \mathbb{E}\left[X_i\right] \mathbb{E}\left[g_A(S_i+1)\right] \\
&\qquad - \mathbb{E}\left[X_i(g_A(S_i + T_i + 1) - g_A(S_i+1))\right] \\
&= \sum_{i\in\mathcal{I}} \quad \mathbb{E}\left[X_i\right] \mathbb{E}\left[g_A(T_i + S_i + X_i + 1) - g_A(S_i+1)\right] \\
&\qquad - \mathbb{E}\left[X_i(g_A(S_i + T_i + 1) - g_A(S_i+1))\right].
\end{aligned}$$

Now we will apply the fact that $||g_A|| \leq 1$ (Proposition 1.10). Given any $j, k \in \mathbb{N}$, we have that

$$|g_A(k+j) - g(k)| \leq 2 \cdot \chi_{\mathbb{N}_{>0}}(j) \leq 2j.$$

As such

$$|\mathbb{E}\left[g_A(T_i + S_i + X_i + 1) - g_A(S_i + 1)\right]| \leq 2 \cdot \mathbb{E}\left[T_i + X_i\right]$$

and

$$|\mathbb{E}\left[X_i(g_A(S_i + T_i + 1) - g_A(S_i + 1))\right]| \leq 2 \cdot \mathbb{E}\left[X_i T_i\right].$$

Plugging this into the inequality above, we obtain

$$\begin{aligned} \mathbb{E}\left[\lambda g_A(W+1) - W g_A(W)\right] &\leq 2 \cdot \sum_{i \in \mathcal{I}} \mathbb{E}\left[X_i\right] \mathbb{E}\left[T_i + X_i\right] + \mathbb{E}\left[X_i T_i\right] \\ &= 2 \cdot (B_1 + B_2 + B_3). \end{aligned}$$

□

Finally, we need a following multivariate version of Theorem 2.1. First we generalize our set up. Again, we let $\mathcal{I}$ be a set and $X_i : \Omega \to \mathbb{N}$ a Bernoulli variable dor all $i \in \mathcal{I}$. Now suppose that $d \in \mathbb{N}$ and

$$\mathcal{I} = \mathcal{I}_1 \sqcup \ldots \sqcup \mathcal{I}_d.$$

We will now be interested in approximating the random variable $W : \Omega \to \mathbb{N}^d$, coordinate-wise defined by

$$W_k = \sum_{i \in \mathcal{I}_k} X_i$$

with a Poisson variable. Again we set

$$p_i = \mathbb{E}\left[X_i\right], \quad p_{ij} = \mathbb{E}\left[X_i X_j\right]$$

and, given $i \in \mathcal{I}$, define

$$\mathcal{D}_i = \left\{j \in \mathcal{I};\ j \neq i,\ X_i \text{ and } X_j \text{ not independent}\right\}.$$

Note that this set may intersect with multiple of the sets $\mathcal{I}_k$.

We will also use three similar quantities to those before:

$$B_{1,k} = \sum_{i \in \mathcal{I}_k} p_i^2, \quad B_{2,k} = \sum_{i \in \mathcal{I}_k} \sum_{j \in \mathcal{D}_i} p_i p_j \quad \text{and} \quad B_{3,k} = \sum_{i \in \mathcal{I}_k} \sum_{j \in \mathcal{D}_i} p_{ij},$$

for $k = 1, \ldots, d$.

The approximation theorem now states:

Theorem 2.2. *Let W be as above such that $\lambda_k = \mathbb{E}[W_k] \in (0, \infty)$ for $k = 1, \ldots, d$. Furthermore, let $Z_k : \Omega \to \mathbb{N}$ be Poisson distributed random variable with mean λ_k for $k = 1, \ldots, d$. Then*

$$\mathrm{d_{TV}}(W, Z) \leq 2 \cdot \sum_{k=1}^{d} B_{1,k} + B_{2,k} + B_{3,k},$$

where $Z : \Omega \to \mathbb{N}^d$ is defined by $Z = (Z_1, \ldots, Z_d)$.

Proof. Our strategy will be to apply the proof of Theorem 2.1 recursively. First note that for $A \subset \mathbb{N}^d$

$$\begin{aligned} \mathbb{P}[W \in A] - \mathbb{P}[Z \in A] &= \mathbb{E}[\chi_A(W)] - \mathbb{E}[\chi_A(Z)] \\ &= \sum_{r=1}^{d} \mathbb{E}[\chi_A(Z_1, \ldots, Z_{r-1}, W_r, W_{r+1} \ldots, W_d)] \\ &\quad -\mathbb{E}[\chi_A(Z_1, \ldots, Z_{r-1}, Z_r, W_{r+1}, \ldots, W_d)]. \end{aligned}$$

Let us write

$$\begin{aligned} t_r &= \mathbb{E}[\chi_A(Z_1, \ldots, Z_{r-1}, W_r, W_{r+1} \ldots, W_d)] \\ &\quad -\mathbb{E}[\chi_A(Z_1, \ldots, Z_{r-1}, Z_r, W_{r+1}, \ldots, W_d)]. \end{aligned}$$

Let $g_{A,r} : \mathbb{N}^d \to \mathbb{R}$ be the function satisfying $g_{A,r}(k) = 0$ when $k_r = 0$ and

$$\lambda_r g_{A,r}(k + e_r) - k_r g_{A,r}(k) = \chi_A(k) - \mathbb{E}[\chi_A(k_1, \ldots, k_{r-1}, Z_r, k_{r+1}, \ldots, k_d)] \tag{2.2}$$

for all $k \in \mathbb{N}^d$, where $e_r \in \mathbb{N}^d$ is defined by

$$(e_r)_j = \begin{cases} 1 & \text{if } j = r \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\begin{aligned} t_r &= \mathbb{E}\left[\lambda_r g_{A,r}(Z_1,\ldots,Z_{r-1},W_r+1,W_{r+1},\ldots,W_d)\right] \\ &\quad -\mathbb{E}\left[W_r g_{A,r}(Z_1,\ldots,Z_{r-1},W_r,W_{r+1},\ldots,W_d)\right] \end{aligned}$$

Like in the proof of Theorem 2.1, we define random variables $S_{i,r}, T_{i,r} : \Omega \to \mathbb{N}$ by

$$S_{i,r} = \sum_{j\in\mathcal{I}_r\setminus(\mathcal{D}_i\cup\{i\}} X_j \text{ and } T_{i,r} = \sum_{j\in\mathcal{D}_i\cap\mathcal{I}_r} X_j$$

for all i, r. We again observe that

$$W_r = S_{i,r} + T_{i,r} + X_i$$

for all i, r. As such, with a similar computation to the one in the proof of Theorem 2.1, we obtain

$$\begin{aligned} t_r &= \sum_{i\in\mathcal{I}_r} \mathbb{E}\left[X_i\right]\mathbb{E}\left[g_{A,r}(Z_1,\ldots,Z_{r-1},S_{i,r}+T_{i,r}+X_i+1,S_{i,r+1}+T_{i,r+1},\right. \\ &\qquad \left.\ldots,S_{i,d}+T_{i,d}) - g_{A,r}(Z_1,\ldots,Z_{r-1},S_{i,r}+1,\ldots,S_{i,d})\right] \\ &\quad -\mathbb{E}\left[X_i(g_{A,r}(Z_1,\ldots,Z_{r-1},S_{i,r}+T_{i,r}+1,S_{i,r+1}+T_{i,r+1},\ldots,\right. \\ &\qquad \left. S_{i,d}+T_{i,d}) - g_{A,r}(Z_1,\ldots,Z_{r-1},S_{i,r}+1,S_{i,r+1},\ldots,S_{i,d}))\right]. \end{aligned}$$

Note that for all $k_1,\ldots,k_{r-1},k_{r+1},\ldots,k_d$:

$$\chi_A(k_1,\ldots,k_{r-1},Z_r,k_{r+1},\ldots,k_d) = \chi_{A'}(Z_r),$$

where $A' \subset \mathbb{N}$ is defined by

$$A' = \{k\in\mathbb{N};\ (k_1,\ldots,k_{r-1},k,k_{r+1},\ldots,k_d)\in A\}.$$

As such (2.2) is an instance of Stein's equation and Proposition 1.10 applies, from which we obtain

$$\sup_{k\in\mathbb{N}^d}\{|g_{A,r}(k)|\} \leq 1.$$

This means that for all $k, j \in \mathbb{N}^d$:

$$|g_{A,r}(k+j) - g_{A,r}(k)| \leq 2\sum_{s=1}^{d} j_s.$$

Hence

$$|t_r| \leq 2 \sum_{i \in \mathcal{I}_r} \mathbb{E}\left[X_i\right] \mathbb{E}\left[X_i + T_{i,r} + \ldots T_{i,d}\right] + \mathbb{E}\left[X_i(T_{i,r} + \ldots + T_{i,d})\right].$$

Filling this in in our original bound, we obtain

$$\mathbb{P}\left[W \in A\right] - \mathbb{P}\left[Z \in A\right] \leq 2 \cdot \sum_{k=1}^{d} B_{1,k} + B_{2,k} + B_{3,k}.$$

□

2.3 Exercises

Exercise 2.1.

(a) Let $X_1, X_2 : \Omega \to \mathbb{N}^r$ be random variables. Show that

$$\mathrm{d_{TV}}\left(X_1, X_2\right) = \sum_{k \in \mathbb{N}^r} \left|\mathbb{P}\left[X_1 = k\right] - \mathbb{P}\left[X_2 = k\right]\right|.$$

(b) Let $r \in \mathbb{N}$ and $\lambda_1, \lambda_2 \in (0, \infty)^r$ and let $\mathbf{X}_1 = (X_{1,1}, \ldots, X_{1,r}), \mathbf{X}_2 = (X_{2,1}, \ldots, X_{2,r}) : \Omega \to \mathbb{N}^r$ be random variables so that $X_{i,j}$ are Poisson distributed with mean $\lambda_{i,j}$ and pairwise independent. Show that

$$\mathrm{d_{TV}}\left(\mathbf{X}_1, \mathbf{X}_2\right) = \mathcal{O}\left(\sum_{i=1}^{r} |\lambda_{1,i} - \lambda_{2,i}|\right)$$

as $\sum_{i=1}^{r} |\lambda_{1,i} - \lambda_{2,i}| \to 0$.

Exercise 2.2. Let $\lambda \in (0, \infty)$. For all $n \in \mathbb{N}$, let $\{X_{i,n}\}_{i=1}^{n}$ be independent Bernoulli variables so that

$$\mathbb{E}\left[X_{i,n}\right] = \lambda/n.$$

Furthermore, define

$$W_n = \sum_{i=1}^{n} X_{i,n}$$

and let Z_λ be a Poisson ditributed random variable with mean λ. Show that

$$W_n \overset{\text{TV}}{\to} Z_\lambda$$

as $n \to \infty$.

Exercise 2.3. Random geometric graphs: let $\mathbb{T}^2$ denote the 2-dimensional torus. That is

$$\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2,$$

where $\mathbb{Z}^2 \curvearrowright \mathbb{R}^2$ by translations. Figure 2.1 shows a cartoon of $\mathbb{T}^2$.

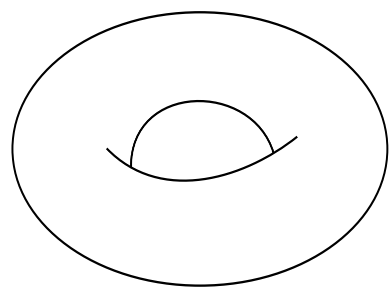

Figure 2.1: A torus.

The Lebesgue measure on $\mathbb{R}^2$ is invariant under the $\mathbb{Z}^2$ action and hence descends to a measure on $\mathbb{T}^2$. $[0,1]^2 \subset \mathbb{R}^2$ forms a fundamental domain for this action. As such the total area of $\mathbb{T}^2$ under this measure is 1. In other words, we obtain a probability space $(\mathbb{T}^2, \mathbb{P})$. $\mathbb{T}^2$ also comes with a distance: the Euclidean distance function $\mathrm{d}_{\mathbb{R}^2} : \mathbb{R}^2 \times \mathbb{R}^2 \to [0,\infty)$ is also invariant under the $\mathbb{Z}^2$ action and hence descends to a distance

$$\mathrm{d}_{\mathbb{T}^2} : \mathbb{T}^2 \times \mathbb{T}^2 \to [0,\infty).$$

Set $r_n = n^{-3/4}$. Given $x_1, \ldots, x_n \in \mathbb{T}^2$ we define a graph as follows. The points $x_1, \ldots, x_n$ will be the vertices of our graph. We connect x_i and x_j by an edge if and only if $\mathrm{d}_{\mathbb{T}^2}(x_i, x_j) \leq r_n$.

(a) Let $X_n : (\mathbb{T}^2)^n \to \mathbb{N}$ count the number of triangles (triples of vertices that are all connected by an edge) in the graph associated to the points $x_1, \ldots, x_n$. Show that

$$\mathbb{E}\left[X_n\right] \to \frac{1}{6} \int_{B_1(0)} \int_{B_1(0)} h(y_1, y_2) dy_1 dy_3$$

as $n \to \infty$, where $B_1(0) \subset \mathbb{R}^2$ denotes the unit ball around the origin in $\mathbb{R}^2$ and

$$h(y_1, y_2) = \begin{cases} 1 & \text{if } \mathrm{d}_{\mathbb{R}^2}(y_1, y_2) \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(b) Set

$$\lambda = \frac{1}{6} \int_{B_1(0)} \int_{B_1(0)} h(y_1, y_2) dy_1 dy_3$$

and let $Z_\lambda : \Omega \to \mathbb{N}$ be a Poisson random variable with mean λ. Show that

$$X_n \overset{\mathrm{TV}}{\to} Z_\lambda$$

as $n \to \infty$.