MATH 6480/STAT 9300/AMCS 6481, Fall 2023, Homework Set 8

The following Theorem gives a criterion for the convergence to Poisson distribution of correlated random variables. We don't have time to discuss it during the class. In this homework, you will learn how to use it.

Let I be a finite or countably infinite index set. Let $\{X_\alpha : \alpha \in I\}$ be random variables, each taking value zero or one. For each $\alpha \in I$ we choose a set B_α, which should be thought of as the set of indices β such that X_α and X_β are highly dependent. Note that this is not a precise heuristic, and that the theorem below will be true no matter what you pick for B_α – it will be informative only if $\{B_\alpha\}$ is chosen with some care. Let p_α denote $\mathbb{E}X_\alpha$ and $p_{\alpha,\beta} := \mathbb{E}X_\alpha X_\beta$. Now define the following quantities.

$$\begin{aligned} b_1 &:= \sum_{\alpha\in I}\sum_{\beta\in B_\alpha} p_\alpha p_\beta\,, \\ b_2 &:= \sum_{\alpha\in I}\sum_{\alpha\neq\beta\in B_\alpha} p_{\alpha,\beta} \\ b_3 &= \sum_{\alpha\in I} \mathbb{E}\left|\mathbb{E}\left(X_\alpha \,|\, X_\beta : \beta \notin B_\alpha\right) - p_\alpha\right|\,. \end{aligned}$$

Loosely speaking, b_1 measures the total size of the dependence neighborhoods, b_2 measures how many dependent pairs are likely to arise, and b_3 measures how honest we were in creating the dependency neighborhood (how far X_α is from being independent of $\{X_\beta : \beta \notin B_\alpha\}$).

Theorem 1 (Arratia-Goldstein-Gordon (1989))**.** *Let $W = \sum_{\alpha\in I} X_\alpha$ and let Z be a Poisson random variable with mean $\lambda := \mathbb{E}W = \sum_{\alpha\in I} p_\alpha$. Then the total variation distance between W and Z is at most $(b_1 + b_2 + b_3)(1 \wedge \lambda)$, which is of course at most $b_1 + b_2 + b_3$.*

1. Let $n \geq 2$ be an integer and $p \in (0,1)$ be real. For $1 \leq i < j \leq n$, let X_{ij} be IID Bernoulli variables with mean p defined on some probably space $(\Omega_n, \mathcal{F}_n \mathbb{P}_n)$. The Erdös-Rényi random graph $G(n,p)$ is the random graph whose vertices are $[n]$ and for which there is an edge between i and j if and only if $X_{ij} = 1$. Say that $\{r,s,t\}$ is a triangle of $G(n,p)$ if $X_{ij} = 1$ for all distinct $i < j$ in $\{r,s,t\}$.

(a) Find an exponent $\alpha > 0$ such that if $n \to \infty$ and $p = \lambda/n^\alpha$, then the expected number of triangles in $G(n,p)$ converges to a finite nonzero value C_λ and determine C_λ in terms of λ.
(b) Prove or disprove that the number of triangles in $G(n, \lambda/n^\alpha)$ converges in distribution to a Poisson with mean C_λ. (Use Arratia-Goldstein-Gordon Theorem)

Solution

(a) Let $A_{r,s,t}$ be the event that $\{r,s,t\}$ form a triangle. i.e. $X_{rs} = X_{rt} = X_{st} = 1$, then let $N = \sum_{\{r,s,t\} \in G(n,p)} \mathbf{1}_{A_{r,s,t}}$ be the random variable indicating the number of triangles in $G(n,p)$. For $1 \leq i \leq j \leq n$, $\mathbb{P}(X_{ij} = 1) = p$ and then $\mathbb{P}_{\{r,s,t\} \in G(n,p)}(A_{r,s,t}) = \mathbb{P}(\{r,s,t\} \text{ is a triangle}) = \mathbb{P}(X_{rs} = X_{rt} = X_{st} = 1) = p^3$.

$$\mathbb{E}(\#\text{ of triangles }) = \mathbb{E}(N) = \sum \mathbb{E}_{\{r,s,t\} \in G(n,p)}(\mathbf{1}_{A_{r,s,t}}) = \sum \mathbb{P}_{\{r,s,t\} \in G(n,p)}(A_{r,s,t}) =$$
$$p^3 \begin{pmatrix} [n-1] \\ 3 \end{pmatrix} = \frac{[n]!}{3!([n]-3)!} p^3 = \frac{[n][n-1][n-2]}{6} p^3 = \frac{\lambda^3 [n][n-1][n-2]}{6 \cdot n^{3\alpha}}$$

where $[n-1]$ choose 3 means that there is such number of triples exists, and p^3 is the probability of each triple being triangle.
For the expected value converges, we need $\alpha = 1$, and then the expected value converges to $C_\lambda = \frac{\lambda^3}{6}$

(b) Let I denote the set of grouped triangles, i.e, I= $\{1 \leqslant r < s < t \leqslant n, X_{rs} = X_{st} = X_{rt} = 1\}$.
Let $A_{r,s,t} = \{\mathbf{1}_\alpha, 1 \leqslant r < s < t \leqslant n\}$ be random variables that indicating whether vertices $\{r,s,t\}$ form a triangle. $\mathbf{1}_\alpha = 1$ when $X_{rs} = X_{st} = X_{rt} = 1$. Let $B_{i,j,k} = \{1_\beta, 1 \leqslant i < j < k \leqslant n\}$ be dependent randoms variables where at least 2 of $\{t,j,k\}$ is equal to a pair in $\{r,s,t\}$ i.e., triangles share a common edge or two triangles are same. The three measures can thus be calculated as following.
i)

$$b_1 = \sum_{\alpha \in I} \sum_{\beta \in B_\alpha} \mathbb{E}\mathbf{1}_\alpha \mathbb{E}\mathbf{1}_\beta = \begin{pmatrix} [n] \\ 3 \end{pmatrix} (3[n]-8) p^5 = \frac{[n][n-1][n-2](3[n]-8)}{6} \left(\frac{\lambda}{n}\right)^5$$

where $\mathbb{E}\mathbf{1}_\alpha = \mathbb{P}(\{r,s,t\} \text{ is triangle}) = p^3$ and $\mathbb{E}\mathbf{1}_\beta = \mathbb{P}(\{i,j,k\} \text{ is triangle given 2 vertices fixed}) = p^2$.
There are $\begin{pmatrix} [n] \\ 3 \end{pmatrix}$ possible combinations of $\{r,s,t\}$ and $3n-8$ possible combinations of $\{i,j,k\}$: consider we already has $A_{r,s,t}$, fix two vertices to construct $A_{i,j,k}$, say $j = s, k = t$ are fixed, then there is $\begin{pmatrix} [n-3] \\ 1 \end{pmatrix} =$ $[n-3]$ possible choice of i for the triangle formed by $\{i,j,k\}$ is different with $A_{r,s,t}$ but share a common edge.

Same replacement can be operated when fix $i = r, j = s$ and $i = r, k = t$. Then there are $3([n] - 3)$ possible choice of edge share triangles. There is also 1 case that $\{i, j, k\} = \{r, s, t\}$. So $3n - 8$ in total.
Then $b_1 \to 0$ as $n \to \infty$
ii)

$$b_2 = \sum_{\alpha \in I} \sum_{\substack{\beta \in B_\alpha \\ \alpha \neq \beta}} \mathbb{E}\mathbf{1}_\alpha \mathbf{1}_\beta = \begin{pmatrix} [n] \\ 3 \end{pmatrix} (3[n] - 9])p^5 = \frac{[n][n-1][n-2](3[n]-9)}{6} \left(\frac{\lambda}{n}\right)^5$$

where $\mathbb{E}\mathbf{1}_\alpha \mathbf{1}_\beta = p^5$ and remove the equality case in b_1. $b_2 \to 0$ as $n \to \infty$
iii)

$$b_3 = \sum_{\alpha \in I} \mathbb{E} \left| \mathbb{E} \left(\mathbf{1}_\alpha \mid \mathbf{1}_\beta, \beta \notin B_\alpha \right) - \mathbb{E}\mathbf{1}_\alpha \right| = 0$$

since $\mathbf{1}_\alpha$ and $\mathbf{1}_\beta$ are independent when $\beta \notin B_\alpha$

By theorem Arratia-Goldstein-Gordon, the total variation distance between $W = \sum_{\{r,s,t\} \in G(n,p)} \mathbf{1}_{A_{r,s,t}}$ and $Poiss(\frac{\lambda^3}{6})$ is $b_1 + b_2 + b_3$, which converges to 0 as $n \to \infty$. Then $W \to Poiss(\frac{\lambda^3}{6})$ in distribution.

2. Let $n = 2^k$ be a power of 2 and let $X_1, \ldots, X_n$ be IID fair 0-1 coin flips. Let

$$V_n = \#\{0 \le j \le n-k : X_{j+1} = \cdots = X_{j+k} = 1\}$$

be the number of times you see the sequence of k ones, and let

$$W_n = \#\{0 \le j \le n-k : X_{j+1} = 0, X_{j+2} = \cdots = X_{j+k} = 1\}$$

be the number of times you see the sequence $0, 1, 1, \ldots, 1$ of a zero and then $k-1$ ones. Prove that one of the sequences $\{V_n\}$ or $\{W_n\}$ satisfies a Poisson limit theorem (use Arratia-Goldstein-Gordon). For the other sequence, you don't need to prove anything, but state a guess as to the limit distribution.

Solution

Let I denote the index range of j, $I = \{0 \leqslant j \leqslant n-kq$, then $|I| = n-k+1$. Let $\{\mathbf{1}_j : 0 \leqslant j \leqslant n-k\}$ be random variables s.t, $\mathbf{1}_j = 1$ when $x_{j+1} = 0, x_{j+2} = \cdots = x_{j+k} = 1$. It can be used to indicate whether a sequence of x satisfying the condition. Let $\{\mathbf{1}_i : 0 \leqslant i \leqslant n-k, |i-j| < k\}$ be dependent random variables s.t. $\mathbf{1}_i = 1$ when $x_{i+1} = 0, x_{i+2} = \cdots = x_{i+k} = 1$. Then the three measures can be calculated as below:
i) Fix i, $\mathbb{E}\mathbf{1}_i = \mathbb{P}(x_{i+1} = 0, x_{i+2} = \cdots = x_{i+k} = 1) = \left(\frac{1}{2}\right)^k$. Fix j, $\mathbb{E}\mathbf{1}_j = \mathbb{P}(x_{j+1} = 0, x_{j+2} = \cdots = x_{j+k=1}) = \left(\frac{1}{2}\right)^k$. There are $n-k+1$ possible j can be taken in $0 \leqslant j \leqslant n-k$ and there are maxima of $2k-1$ possible i can be taken in $j-k < i < j+k$. Notice that not all integers in $(j-k, j+k)$ are also in $[0, n-k]$ for some k, i has no enough choice of $2k+1$ integers. Then

$$b_1 = \sum_{j \in I} \sum_{i-j|<k} \mathbb{E}1_i \mathbb{E}1_j < (n-k+1)(2k-1)\frac{1}{4^k} < 2^k \cdot 2k \cdot \frac{1}{4^k} = \frac{k}{2^{k-1}} < \frac{k}{(k-1)^2} \text{ when } k > 5 \to 0$$

Therefore $b_1 \to 0$ as $n \to \infty$.
ii)

$$b_2 = \sum_{j \in 1} \sum_{li-j|<k} \mathbb{E}1_i 1_j = 0$$

$\mathbf{1}_j = 0$ when $\mathbf{1}_i = 1$ and $\mathbf{1}_i = 0$ when $\mathbf{1}_j = 1$. These two events can not happen at the same time.
iii)

$$b_3 = \sum_{j \in \mathbb{I}} \mathbb{E}\left|\mathbb{E}\left(1_j \left|1_m, \right| i-m \left|\geqslant k\right.\right) - \mathbb{E}1_j\right| = \sum_{j \in J} \mathbb{E}\left|\mathbb{E}\left(1_j\right) - \mathbb{E}1_j\right| = 0$$

since 1_j and 1_m are independent given $|i-m| \geqslant k$.
We also need to compute the intensity of the poisson distribution. By construction,

$$\lambda = \sum_{j=0}^{2^k-k} \mathbb{E}(\mathbf{1}_j) = (2^k - k + 1)(\frac{1}{2})^k = 1 - \frac{k+1}{2^k} \to 1$$

as $n \to \infty$. By theorem Arratia-Goldstein-Gordon, the total variation distance between $W = \sum \mathbf{1}_j$ and $Poiss(1)$ is $b_1 + b_2 + b_3$, which converges to 0 as $n \to \infty$. Then $W \to Poiss(1)$ in distribution.

As for V_n, we can see that it delays exponentially given $n = 2^k$. It seems that the distribution of V_n should look like poisson in shape. However, it cannot be a poisson since its intensity is not a constant, compared to W_n without a starting point. A possible distribution may fit is a sum of poisson random variables.

3. A probability model states that people will arrive at random positive times, with the density of arrivals at time t equal to $1/\sqrt{t}$ (that is, arrivals occur at distinct times and the probability of an arrival in the time interval $[t, t+dt]$ is $t^{-1/2}\,dt$) and the numbers of arrivals in disjoint intervals independent. Let $N(t)$ denote the number of arrivals up to time t and let X_k denote the time of the k^{th} arrival.
(a) Give a formula for $\mathbb{P}(N(a)=k, N(b)=l)$.
(b) Compute the density of the pair (X_1, X_2).

Solution

(a) $\mathbb{P}(N(a)=k, N(b)=l) = \mathbb{P}(A\cap B) = \mathbb{P}(A)\cdot\mathbb{P}(B)$, where A is the event that k people arrives in time $[0,a]$ and B is the event that $l-k$ people arrives in time $(a,b]$. A, B are independent.
Notice that this is an nonhomogeneous poisson process. $\mathbb{P}\left(\nu[t,t+dt]\right) = \frac{1}{v}dt$, where ν is a counting measure with same settings as 4.16 in notes07. then the intensity satisfies $\int_a^b \frac{1}{\sqrt{t}}dt$. Then the poisson process can be considered as

$$N(a+b) - N(b) \sim \text{ Poiss } \left(\int_b^{a+b} \frac{1}{\sqrt{t}}dt\right)$$

And the intensity $\int_b^{a+b} \frac{1}{\sqrt{t}}dt = 2(\sqrt{a+b} - \sqrt{b})$. For nonhomogeneous poisson process, $N(0)=0$. Then $N(a)$ and $N(b)-N(a)$ follows poisson distribution as following.

$$N(a) \sim \text{Poiss}(2\sqrt{a}), \quad N(b)-N(a) \sim \text{Poiss}(2(\sqrt{b}-\sqrt{a}))$$

Then the probability $\mathbb{P}(A)$ and $\mathbb{P}(B)$ can be calculated.

$$\mathbb{P}(A) = \mathbb{P}(N(a)=k) = \frac{(2\sqrt{a})^k}{k!}e^{-2\sqrt{a}}, \quad \mathbb{P}(B) = \mathbb{P}(N(b)-N(a)=l-k) = \frac{[2(\sqrt{b}-\sqrt{a})]^{l-k}}{(l-k)!}\cdot e^{-2(\sqrt{b}-\sqrt{a})}$$

Therefore, the formula can be calculated as

$$\mathbb{P}(N(a)=k, N(b)=l) = \mathbb{P}(A)\cdot\mathbb{P}(B) = \frac{(2\sqrt{a})^k\cdot(2\sqrt{b}-2\sqrt{a})^{l-k}}{k!(l-k)!}\cdot e^{-2\sqrt{b}}$$

(b) To compute the joint density of the pair (x_1, x_2), i.e.,$\mathbb{P}\left(X_1=t_1, X_2=t_2\right)$, we need integral the marginal distribution function and the get derivatives, i.e.

$$\frac{d}{dt_2}\frac{d}{dt_1}\int_0^t \mathbb{P}(X_2 \leqslant t_2 | X_1 = t_1)\mathbb{P}(X_1=t_1)dt$$

I will start with distribution of the first arrival.

$$\mathbb{P}\left(X_1 \leqslant t_1\right) = \mathbb{P}\left(N\left(t_1\right) \geqslant 1\right) = 1 - \mathbb{P}\left(N\left(t_1\right)=0\right) = 1 - \frac{\left(2\sqrt{t_1}\right)^0}{0!}e^{-2\sqrt{t_1}} = 1 - e^{-2\sqrt{t_1}}$$

then the density is

$$\mathbb{P}\left(X_1 = t_1\right) = \frac{d}{dt_1}\left(1 - e^{-2\sqrt{t_1}}\right) = \frac{1}{\sqrt{t_1}} e^{-2\sqrt{t_1}}$$

The conditional distribution is presented as below:

$$\begin{aligned}
\mathbb{P}\left(X_2 \leqslant t_2 \mid X_1 = t_1\right) &= \mathbb{P}\left(X_2 - X_1 \leqslant t_2 \ -t_1 \mid X_1 = t_1\right) \\
&= \mathbb{P}\left(N\left(t_2\right) - N\left(t_1\right) \geqslant 1\right) \\
&= 1 - \mathbb{P}\left(N\left(t_2\right) - N\left(t_1\right) = 0\right) \\
&= 1 - e^{-2\left(\sqrt{t_2} - \sqrt{t_1}\right)}
\end{aligned}$$

And the corresponding density is:

$$\begin{aligned}
\mathbb{P}\left(X_2 = t_2 \mid X_1 = t_1\right) &= \mathbb{P}\left(X_2 - X_1 = t_2 - t_1 \mid X_1 = t_1\right) \\
&= \frac{d}{dt_2}\left(1 - e^{-2\left(\sqrt{t_2} - \sqrt{t_1}\right)}\right) \\
&= \frac{1}{\sqrt{t_2}} e^{-2\sqrt{t_2} + 2\sqrt{t_1}}
\end{aligned}$$

Combining two parts we thus get

$$\begin{aligned}
\mathbb{P}\left(X_1 = t_1, X_2 = t_2\right) &= \frac{d}{dt_2}\frac{d}{dt_1}\int_0^{t_1} \mathbb{P}(X_2 \leqslant t_2 | X_1 = t_1)\mathbb{P}(X_1 = t_1)dt \\
&= \frac{1}{\sqrt{t_1 t_2}} e^{-2\sqrt{t_2}}
\end{aligned}$$