ACTA MATHEMATICA VIETNAMICA
TOM 6 No 2 (1981)

THE ASYMPTOTIC LAW OF THE NUMBER OF PATIENTS IN A SPACE – TIME REGION

NGUYỄN HỮU TRỢ
Institute of Mathematics
Hanoi.

Let us consider n cases of some disease : $(S_i, T_i)_i$ $(i=1,2,..,n)$ where $S_i \in R^2$ and $T_i \in R^1$ is the place and time of onset of i—th case, respectively.

The desire of epidemiologists is to obtain powerful methods of detecting clustering of patients.

If there exists a space-time clustering, i. e. cases in the cluster will be close both in time and space, it would be evidence supporting the hypothesis that the disease is contagious. Conversely, if there is no space - time clustering, the disease is non-contagious.

The hypothesis of no-clustering is equivalent to one that the locations in tim are matched at random with the locations in space, there being a total of $n!$ equiprobable sets of matchings. Thus, one can consider the space coordinates as fixed, while the time coordinates are random variables $T_{\pi(i)}$ $(i=1,2,..., n)$ with the uniform distribution on the probability space of all permutations of the numbers $(1,2,..,n)$, where π is a random permutation of $(1,2,...,n)$.

The statistic which corresponds to the number of cases being close together both in space and time was first discussed by Knox (of [3] and [4]). In terms of the graphical theory, Barton — David (cf. [2]) and, then Abe (cf. [1]) gave sufficient conditions for an asymptotic Poisson distribution of the Knox' statistic and an asymptotic unit normal distribution of it's standardised variable under the hypothesis of no space-time clustering. Mantel (cf. [6]) also gave another statistic which contains the Knox' statistic as special case. In application, the above-mentioned statistics can be considered as tests for space-time clustering.

For any $A_n \subset R^2$, $B_n \subset R^1$ let us denote

$$N_{1n} = N_{1n}(A_n) = \# \{ S_i : S_i \in A_n \quad i = 1,2,\ldots, n \}$$

$$N_{2n} = N_{2n}(B_n) = \# \{ T_i : T_i \in B_n \quad i = 1,2,\ldots, n \}$$

$$N_n = N_n(A_n \times B_n) = \# \{ (S_i, T_{\pi(i)}) : S_i \in A_n, T_{\pi(i)} \in B_n, \; i = 1,2,\ldots, n \}$$

The purpose of the paper is to study the asymptotical behaviour of the laws of N_n when n tends to infinity. Namely, we obtain a necessary and sufficient condition for the convergence of law of N_n to the degenerate, Poissonian or binomial one.

In the following theorem, we denote the law of N_n by

$\mathcal{L}(N_n)$, the Poisson law with expectation λ $(o < \lambda < \infty)$ by $\mathcal{P}(\lambda)$, the binomial law of order N with the parameter $\frac{\lambda}{N}$ $(0 < \lambda < \infty)$ by $\mathcal{B}(N, \frac{\lambda}{N})$ and the distribution assigning mass 1 to o or ∞) by $\mathcal{L}(0)$ (or $\mathcal{L}(\infty)$).

THEOREM

Under the hypothesis of no space-time clustering, we have:

(i) N_n converges in distribution to 0 or ∞ if and only if $\frac{N_{n1} N_{2n}}{n} \to 0$ or ∞ when $n \to \infty$, respectively.

(ii) N_n has an asymptotic Poisson law $\mathcal{P}(\lambda)$ if and only if $\frac{N_{1n} N_{2n}}{n} \to \lambda$ $(0 < \lambda < \infty)$ and

$N_{1n} \to \infty$, $N_{2n} \to \infty$ when $n \to \infty$

(iii) N_n *has an asymptotic binomial law $\mathcal{B}\left(N, \frac{\lambda}{N}\right)$ if and only if $\frac{N_{1n} N_{2n}}{n} \to$ $\to \lambda$ $(0 < \lambda < \infty)$ and either $N_{1n} \to N < \infty$ or $N_{2n} \to N < \infty$ when $n \to \infty$.*

Proof.

Since

$$N_{1n} = \sum_{i=1}^{n} 1_{A_n}(S_i)$$

$$N_{2n} = \sum_{i=1}^{n} 1_{B_n}(T_i)$$

$$N_n = \sum_{i=1}^{n} 1_{A_n}(S_i) \, 1_{B_n}(T_{\pi(i)})$$

where $$\mathbf{1}_A(x)=\begin{cases}1 \text{ if } x\in A\\ 0 \text{ if } x\notin A.\end{cases}$$

and under the hypothesia of no space-time clustering, the k-th order factorial moment of N_n can be written as follows:

$$\mu^n_{[k]}=\mathbf{E}N_n[N_n-1]\ldots[N_n-(k-1)]$$

$$=\mathbf{E}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\ldots\sum_{i_k=1}^{n}\mathbf{1}_{A_n}(S_{i_1})\,\mathbf{1}_{B_n}(T_{\pi(i_1)})\ldots\mathbf{1}_{A_n}(S_{i_k})\,\mathbf{1}_{B_n}(T_{\pi i_k})$$

$$(i_p\neq i_q: p\neq q;\ p,\ q=1,2,\ldots k)$$

$$=\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\ldots\sum_{i_k=1}^{n}\mathbf{1}_{A_n}(S_{i_1})\ldots\mathbf{1}_{A_n}(S_{i_k})\,\mathbf{E}\,\mathbf{1}_{B_n}(T_{\pi(i_1)})\ldots\mathbf{1}_{B_n}(T_{\pi(i_k)})$$

$$(i_p\neq i_q: p\neq q;\ p,\ q=1,2,\ldots k)$$

$$=\frac{\left[\underset{(i_p\neq i_q:\ p\neq q,\ p,\ q=1,2,\ldots,k)}{\sum_{i_2=1}^{n}\sum_{i_2=1}^{n}\ldots\sum_{i_k}^{n}}\mathbf{1}_{A_n}(S_{i_1})\ldots\mathbf{1}_{A_n}(S_{i_k})\right]\left[\underset{(i_p\neq iq:\ p\neq q;\ p,q=1,2,\ldots,k)}{\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\ldots\sum_{i_k=1}^{n}}\mathbf{1}_{B_n}(T_{i_1})\ldots\mathbf{1}_{B_n}(T_{i_k})\right]}{n(n-1)\ldots(n-(k-1))}$$

$$=\prod_{i=0}^{k-1}\frac{(N_{1n}-i)(N_{2n}-i)}{n-i}$$

$$=\prod_{i=0}^{k-1}\left[\frac{N_{1n}N_{2n}}{n-i}-\frac{i}{n-i}(N_{1n}+N_{2n})+\frac{i^2}{n-i}\right]$$

for all positive integers k.

Thus, we get the formula:

$$(1)\quad \mu^n_{[k]}=\prod_{i=0}^{k-1}\left[\frac{N_{1n}N_{2n}}{n-i}-\frac{i}{n-i}(N_{1n}+N_{2n})+\frac{i^2}{n-i}\right]$$

The sufficiency:

Suppose that the condition of (i) are satisfied, i.e. $\frac{N_{1n}N_{2n}}{n}\to 0$ (or ∞) when $n\to\infty$. It is obious that the expression in the square brackets of (1) tends to

0 (or ∞) when $n \to \infty$. Consequently $\mu^{n}_{[k]} \to 0$ (or ∞) which by the Frechet – Shohat limit theorem (cf. [5], p. 185), implies that $\mathcal{L}(N_n) \to \mathcal{L}(0)$ or $\mathcal{L}(\infty)$.

Next, if the conditions of (ii) are satisfied, i.e.

$$\frac{N_{1n}N_{2n}}{n} \to \lambda \ (0 < \lambda < \infty) \text{ and } N_{1n} \to \infty, N_{2n} \to \infty$$

then $\frac{N_{1n}}{n} \to 0$ and $\frac{N_{2n}}{n} \to 0$ as $n \to \infty$.

Consequently, the expression in the square brackets of (1) tends to λ as $n \to \infty$. Thus $\mu^{n}_{[k]} \to \lambda^k$ which, by the Frechet-Shohat limit theorem, follows $\mathcal{L}(N_n) \to \mathcal{P}(\lambda)$

Now suppose that the conditions of (iii) are satisfied.

Then either $\frac{N_{1n}}{n} \to 0$ and $\frac{N_{2n}}{n} \to \frac{\lambda}{N}$ or $\frac{N_{1n}}{n} \to \frac{\lambda}{N}$ and $\frac{N_{2n}}{n} \to 0$ as $n \to \infty$.

In both cases the expression in the square brackets of (1) tends to

$$\lambda - i\,\frac{\lambda}{N} = \frac{\lambda}{N}(N-i),\ i = 0, 1, \dots, k-1 \text{ as } n \to \infty.$$

Consequently,

$$\mu^{n}_{[k]} \to \begin{cases} \left(\frac{\lambda}{N}\right)^k \prod_{i=0}^{k-1}(N-i) & \text{if } k = 1, 2, \dots, N, \\ 0 & \text{if } k = N+1, N+2, \dots \end{cases}$$

which, again by the Erechet-Shohat limit theorem, implies that

$$\mathcal{L}(N_n) \to \mathcal{B}\left(N, \frac{\lambda}{N}\right) \text{as } n \to \infty.$$

The necessity:

We shall first prove the conditions of (ii). The conditions of (i) and (iii) can be obtained by the same method.

Assume that $\mathcal{L}(N_n) \to \mathcal{P}(\lambda)$, but the conditions of (ii) are not satisfied. Then at least one of the following conditions is satisfied:

a) There exists a subsequence $\frac{N_{1nk}N_{2nk}}{n_k}$ of the sequence $\frac{N_{1n}N_{2n}}{n}$ such that $\frac{N_{1nk}N_{2nk}}{n_k} \to \bar{\lambda} \neq \lambda$ as $k \to \infty$.

b) There exists a subseqence N_{1n_k} (or $N_{2n_k)}$ having a a finite limit as $k \to \infty$.

If a) is true, but b) is false by the proof of the above part, the asymptotic law of N_{n_k} is one of the followings: $\mathcal{L}(0)$, $\mathcal{L}(\infty)$, $\mathcal{P}(\overline{\lambda})$or $\mathcal{B}(.,.)$, which contradicts the assumption that $\mathcal{L}(N_{n_k}) \to \mathcal{P}(\lambda)$.

If b) is true, then $N_{1n_k} \leqslant N < \infty$ for some N

It is obvious that $N_{n_k} \leqslant N_{1n_k} \leqslant N$ for all n. Thus, the asymptotic law of N_{n_k} can not be Poissonian. This implies that the conditions of (ii) are necessary.

Received July 15, 1981
in revised form November 20, 1981

REFERENCES

[1] Abe, O. (1969).
A Central limit theorem for the number of edges in the random intersection of two graphs. The Annals of Mat. Stat. Volume 40, Number 1, February 1969.

[2] Barton, D. E and, David, F. N. (1966).
The random intersection of two graphs. Research paper in statistics (F. N. David, Ed.) John Wiley and Sons. In, New York.

[3] Knox, G. (1964 a).
Epidemiology of childhood leukaemia in Northumberland and Durham. Brit J. prev. Soc. Med. 18, 17 – 24.

[4] Knox, G (1964 b).
The detection of space-time interactions. Appl. statist 13. 25 – 29.

[5] Loévé, M,
Probability theory
Second Edition. 1960.

[6] Mantel N. (1967)
The detection of desease clustering and a generalized regression approach. Cancer Res. 27, 209 – 220.