LIMITING THEOREMS ON 'CASE' REPORTING

ARNI S.R. SRINIVASA RAO

Centre for Ecological Sciences, Indian Institute of Science
Bangalore, INDIA.
email: arni.srs.rao@ces.iisc.ernet.in

***Applied Mathematics Letters* (US) (*In Press*)**

Abstract. The relation between reported disease cases and actual cases (hypothetical) is defined. Certain situations where such sequences of relation converge or diverge is studied.

Key words: Monotone convergence, Poisson process,series.

Acknowledgement. I thank the Department of Science and Technology, New Delhi for funding this research (SR/FTP/MS-12/2001).

Proposed running title: Limiting theorems

AMS subject classifications: 40A05, 26A48

1. Adjustment of under-reporting

Usually, reporting of disease cases in epidemiology is not complete. Here in this work these cases are treated as points on the positive real line and focus their mathematical properties.

Definition 1. A function $\phi(y)$ that determines the closeness of reported cases R to total (actual) cases T is defined as a *efficiency function* of cases.

Definition 2. Let $R_1, R_2, R_3 ... R_N$ and $T_1, T_2, T_3, ..., T_N$ be the numbers of reported and total cases in the time intervals $[\Omega_0, \Omega_1)$, $[\Omega_1, \Omega_2)$, $[\Omega_2, \Omega_3), ...[\Omega_{N-1}, \Omega_N)$, where Ω_N can be current calender time or previous to the current calender time. If $y_j (j = 1, 2, 3, ...N)$ be the estimated percentage of cases that were not reported for the year j and is available, then define *total cases* T_i at time i as the ratio of reported cases R_i $(R_i > 0)$ and the efficiency function $\phi(y_i)$. This means

$$T_i = \frac{R_i}{\phi(y_i)} \quad \text{for all } i \leq N \quad (\text{where } \phi(y_i) = 1 - y_i) \text{ and } y_i \neq 1. \tag{1}$$

We modify definition 2 by removing the condition on time i.

Definition 3. Define T_i as the ratio of $R_i (R_i > 0)$ and $\phi(y_i)$ for all $i \in \mathbf{R}^+$i.e.

$$T_i = \frac{R_i}{\phi(y_i)} \quad \text{for all } i \in \mathbf{R}^+ \quad (\text{here } R_i > 0 \text{ and } \phi(y_i) = 1 - y_i) \text{ and } y_i \neq 1. \tag{2}$$

Later in section 2 we see that actually $0 < y < 1$. If $\phi(y_i)$ is a constant in (1), then $R \propto T$ or $R = CT$, where C is proportionality constant $(= \phi(y))$.

Remark 1. $\phi(y) \to 1$ implies $R \to T$. An important generalization is proved by the following theorem.

Theorem 1. *Let $R, T \in \mathbf{R}^+ - \{0\}$. For every $\alpha-$neighbourhood $N_\alpha(T)$ of T, there exist a $\delta-$neighbourhood $N_\delta(1)$ of 1, around 1, with the property that for all $\phi_n(y) \in N_\delta(1)$ $(\phi(y) \neq 1)$, it follows that $R_n \in N_\alpha(T_n)$. To put this simple, we can state the theorem as below:*

$$\phi_n(y) \in N_\delta(1) \ (\phi(y) \neq 1) \ \Rightarrow \ R_n \in N_\alpha(T_n)$$

Proof. Let $\alpha > 0$. $\phi_n(y) \in N_\delta(1) \Rightarrow | \ \phi_n(y) - 1 \ | < \delta$. By the definition (3), $T_n = \frac{R_n}{\phi_n(y)}$, which implies,

$$\begin{aligned} | \ R_n - T_n \ | &= | \ R_n - \frac{R_n}{\phi_n(y)} \ | \qquad \text{for all } R_n \in (R_n), \\ &\Rightarrow | \ \frac{R_n}{\phi_n(y)} (\phi_n(y) - 1) \ | \qquad \text{for all } R_n \in (R_n), \\ &\Rightarrow | \ \frac{R_n}{\phi_n(y)} \ || \ \phi_n(y) - 1 \ | \qquad \text{for all } R_n \in (R_n). \end{aligned} \tag{3}$$

Since $\delta-$neighbourhood around 1 has radius at the most $1-\pi$ (for $\pi > 0$), it follows that for a given r,

$$| \ \frac{R_n}{\phi_n(y)} \ | = | \ \frac{R_n}{1 - y_n} \ | < T_n \tag{4}$$

Now choosing $\delta = \dfrac{\alpha\phi_n(y)}{R_n}$, we have

$$\mid R_n - T_n \mid = \mid \frac{R_n}{\phi_n(y)} \mid\mid \phi_n(y) - 1 \mid < \alpha.$$

Hence the proof. □

Remark 2. $P(\mid T_n - R_n \mid < \alpha) = 1$ implies $lim\ \phi_n(y) \to 1$. It is not difficult to prove $\phi_n(y) \in N_\delta(1)\ (\phi(y) \neq 1) \ \Leftrightarrow\ R_n \in N_\alpha(T_n)$

2. SOME PRELIMINARY RESULTS ON EFFICIENCY FUNCTION

Let $y_1 \geq y_2 \geq y_3 ...$, then,

$$\Rightarrow \quad 1 - y_1 \leq 1 - y_2 \leq 1 - y_3 ...$$

$$\Rightarrow \quad \phi(y_1) \leq \phi(y_2) \leq \phi(y_3) \leq ... \tag{5}$$

This implies that if the percentage of cases that were not reported over the period decreases then the efficiency function will increase with the time i. Now, take $y_1 < y_2 < y_3 ...$ then

$$\Rightarrow \quad 1 - y_1 > 1 - y_2 > 1 - y_3 > ...$$

$$\Rightarrow \quad \phi(y_1) > \phi(y_2) > \phi(y_3) > ... \tag{6}$$

This means ϕ is strictly a monotonic decreasing function.

Lemma 1. *If R is always under-reported and the efficiency function non-decreasing, then sequence (R_n) converges.*

Proof. Since R is always under-reported they will never be equal to T. This means sequence (R_n) is bounded above. That the efficiency function is non-decreasing implies, that the percentage of cases not reported is non-increasing (5). This means $R_1 \leq R_2 \leq R_3 \leq ....$ So R is bounded and also monotone, hence by using MCT sequence (R_n) converges. □

Another proof for theorem 1 is given below:

Proof. In general from (3), $T = R\phi(y)^{-1} = R(1-y)^{-1}$, where $y \neq 1$. Let us assume that y will never be extinct. Higher value of y indicates poor reporting and lower value of y indicates good reporting of cases. So

$$0 < y < 1 \tag{7}$$

This means

$$1 > y > y^2 > y^3 > ... > 0$$

This means sequence (y^n) is bounded and monotone. Hence by MCT (y^n) converges to a limit $\mathbf{a}$ (say). But $(y^{2n}) = y^n.y^n = \mathbf{a^2}$ (algebraic limit theorem). Therefore $\mathbf{a} = 0$ and hence sequence (y^n) converges to zero. This implies the sequence $(\phi_n(y)) \to 1$ as $n \to \infty$, so $R \to T$ or $R_n \in N_\alpha(T_n)$. □

Remark 3. $0 < y < 1 \Rightarrow \mid y \mid < 1$ and R cannot be *zero*. So

$$\begin{aligned} T_i &= R_i(1 - y_i)^{-1} \\ &= \sum_{k=0}^{\infty} R_i y_i^k \end{aligned} \tag{8}$$

which is a geometric series and it is known that if $y_i = 1$ and $R_i \neq 0$ then the series diverges. Also total cases reported can be written as $R_i(1-y_i^n)(1-y_i)^{-1}$ as $n \to \infty$.

Consider the series $\sum_{n=1}^{\infty} T_n$ and take $u_m = T_1 + T_2 + T_3 + ...T_m$. If u_m converges to U(say), then we can say that $\sum T_n$ converges. Using (8) we can write

$$\begin{aligned} u_m &= \sum_{j=1}^{m}\sum_{k=0}^{\infty} R_j y_j^k \\ &= \sum_{j=1}^{m} R_j(1-y_j)^{-1} \\ &= R_1(1-y_1)^{-1} + R_2(1-y_2)^{-1} + R_3(1-y_3)^{-1} + ... + R_m(1-y_m)^{-1} \end{aligned} \tag{9}$$

Now consider $v_m = R_1 + R_2 + R_3 + ...R_m$. Since $y \neq 1$ and $0 < y < 1$, every term in (9) is greater than the corresponding terms in v_m. Then $u_m > v_m$. If $\phi_j(y)$ and R_j are same with increase in time j in above, then

$$u_m = mR(1-y)^{-1} > mR = v_m \text{ for } m \in \mathbf{N} \tag{10}$$

Therefore

$$v_m > m \text{ and hence } u_m > m \text{ for m} \in \mathbf{N} \tag{11}$$

Thus the series $\sum T_n$ diverges. By the comparison test of convergence we can say that $\sum R_n$ also diverges. If there is no condition on $\phi_j(y)$ and R_j then also every term in (9) is greater than the corresponding term because R is a positive quantity and $0 < y < 1$. In addition, $u_m < u_{m+1} < u_{m+2}$....i.e. (u_m) is monotonic increasing sequence. We have proved a lemma (1) when R_j increases with the time j. Now, we will prove an important theorem if the sequence (R_n) decreases.

Theorem 2. *If sequence (R_n) is decreasing and condition (5) is valid then the sequence (T_n) is convergent.*

Proof. Since (5) is given, the inequality

$$(1-y_1)^{-1} > (1-y_2)^{-1} > (1-y_2)^{-1}... \tag{12}$$

is true. Given that the sequence (R_n) is decreasing, so $R_1 > R_2 > R_3 > ....$ and $R > 0$ from definition (3), so 0 is the *infimum* of (R_n). Hence, multiplying by $R_1, R_2, R_3...$ to the corresponding terms in (12) will not change the inequality, so

$$R_1(1-y_1)^{-1} > R_2(1-y_2)^{-1} > R_3(1-y_3)^{-1} > ... \tag{13}$$

is also true and has same *infimum* as above. hence by MCT sequence (T_n) is convergent.

Hence the proof. □

Corollary 1. *From definition 3 and (7), we have $R_i < T_i$ for all i, hence under the conditions of theorem (2) sequence (R_n) is also convergent.*

If we assume reported cases *i.e.* R will follow Poisson distribution (with parameter β, say) over the time i and there is a decay in percentage of cases not reported *i.e* decay in y is an exponential rate then the nth term of T_n as defined in (3) is

$$\begin{aligned}
(14)\quad R_n(1-y_n)^{-1} &= \frac{e^{-\beta}\beta^{R_n}}{R_n!}\left(1-y_0e^{-d.n}\right)^{-1} \\
&= \frac{e^{-\beta}\beta^{R_n}}{R_n!}\sum_{k=0}^{\infty}\left(y_0e^{-d.n}\right)^k \\
&= \frac{e^{-\beta}\beta^{R_n}}{R_n!}\left(1+y_0e^{-d.n}+y_0^2e^{-2.d.n}+...\right) \\
(15)\qquad &= \frac{e^{-\beta}\beta^{R_n}}{R_n!}+\frac{y_0e^{-(\beta+d.n)}\beta^{R_n}}{R_n!}+\frac{y_0^2e^{-(\beta+2.d.n)}\beta^{R_n}}{R_n!}+...
\end{aligned}$$

Just a reminder here R_n, β, d, y_0 and $n > 0$. y_0 is initial value of y before decay, which is a constant. Every term in (15) is positive and also these terms monotonically decrease (since (7)). All the terms from the second to rest of the terms in (15) converge to the first term as $y_0 \to 1$ and $d \to 0$. Thus,

$$\begin{aligned}
(16)\quad \left\{\frac{e^{-\beta}\beta^{R_n}}{R_n!},\frac{y_0e^{-(\beta+d.n)}\beta^{R_n}}{R_n!},\frac{y_0^2e^{-(\beta+2.d.n)}\beta^{R_n}}{R_n!},\cdots\right\} &\leq \frac{e^{-\beta}\beta^{R_n}}{R_n!} \\
(17)\qquad &< \sum_{R_n=1}^{\infty}\frac{e^{-\beta}\beta^{R_n}}{R_n!}
\end{aligned}$$

In (17) R_n cannot begin from 0 like in general discrete mass functions, because it is a positive quantity. The right hand side of (17) is 'thinly' less than total probability of mass function, so (17)<1. Hence the sequence,

$$(18)\qquad \left(\frac{y_0e^{-(\beta+m.d.n)}\beta^{R_n}}{R_n!}\right)_{m=0,1,2,3...}$$

is bounded. Thus it is convergent.

Now if we consider the series (15), then

$$\begin{aligned}
\frac{e^{-\beta}\beta^{R_n}}{R_n!} &< \sum_{R_n=1}^{\infty}\frac{e^{-\beta}\beta^{R_n}}{R_n!} <\simeq 1 \\
\frac{y_0e^{-(\beta+d.n)}\beta^{R_n}}{R_n!} &< \sum_{R_{n=1}}^{\infty}\frac{y_0e^{-(\beta+d.n)}\beta^{R_n}}{R_n!} <\simeq 1 \\
(19)\qquad \frac{y_0^2e^{-(\beta+2.d.n)}\beta^{R_n}}{R_n!} &< \sum_{R_n=1}^{\infty}\frac{y_0^2e^{-(\beta+2.d.n)}\beta^{R_n}}{R_n!} <\simeq 1 \\
&\cdots \\
\frac{y_0^je^{-(\beta+j.d.n)}\beta^{R_n}}{R_n!} &< \sum\frac{y_0^je^{-(\beta+j.d.n)}\beta^{R_n}}{R_n!} <\simeq 1 \\
&\cdots
\end{aligned}$$

Thus we see series (15) resembles the series $\sum 1$ which is divergent. Hence the sequence

$$(20)\qquad \left(R_n(1-y_n)^{-1}\right)$$

converges. Now, we can state the following theorem based on the results obtained.

Theorem 3. *If $R_n \sim$ Poisson mass function or Poisson process and $y \sim$ exponential decay then the sequence (T_n) is divergent.*

Proof. Given $R_n = \frac{e^{-\beta}\beta^{R_n}}{R_n!}$ and $y = \left(1 - y_0 e^{-d.n}\right)^{-1}$. Here β is Poisson parameter, y_0 is initial value of y, d is decay rate with time i. Now using the definition (3) and from (14) to (20) then the result is proved. □

Though the way in which reporting is introduced and discussed in this work is new, one can read in general about the principles of monotone convergence and divergence from the standard books ([1, 2]).

3. ORDER IN REPORTING

Here one or multiple case reporting is considered. A case could be reported either once or more than once. If it is more than once then it could be possible that reported cases outnumber actual cases in that particular time.

Definition 4. Total (actual) cases at the time i is defined as the sum of reported and error of reporting at the time i over positive real line. i.e. symbolically $T_i = R_i \pm e_i$ for all $i \in \mathbf{R}^+$and e_i is error of reporting.

As per this definition a set $A = \{(R_i, T_i)\,;$ for all $i \in \mathbf{R}\}$ is an ordered set, because either $R_i < T_i$ or $R_i = T_i$ or $T_i < R_i$ will hold true. We can also imagine R_i is around T_i with a radius e_i. As the reporting improves or error of reporting reduces with increase in time i, then the radius e_i approaches *zero*. This results in decreasing the distance between R_i and T_i with increasing in time i. This encourages to state the following theorem.

Theorem 4. *As a monotonic decreasing sequence (e_n) approaches to zero then sequence (R_n) converges to sequence (T_n).*

References

[1] W. Rudin, Principles of Mathematical Analysis, Tata-McGraw-Hill, Singapore, 1976.
[2] EG. Phillips, A Course of Analysis, Cambridge University Press, Cambridge,1948.