·学习·发现·交流·

疾病潜伏期容许限容许区间的推算

赵飞 蔡全才 陈启明 姜庆五

【导读】 为从不完整调查资料推算疾病潜伏期容许限和容许区间，假定观察次数（样本量）服从Poisson分布，引用复合整参数β-分布法计算疾病的潜伏期容许限和容许区间。获得以样本最小顺序统计量与最大顺序统计量为容许限，或以样本最大、最小顺序统计量差为容许区间的容许概率容许度，以及样本量服从Poisson分布参数的对应关系式。根据不完整调查资料提供的信息，结合数值举例，讨论如何有效合理地选择样本量的取样单位。

【关键词】 潜伏期容许限；潜伏期容许区间；复合整参数β-分布；样本量分布参数

Estimation on tolerance limits and tolerance interval regarding the disease incubation *ZHAO Fei*[1,2], *CAI Quan-cai*[3], *CHEN Qi-ming*[4], *JIANG Qing-wu*[1,2]. 1 *Department of Epidemiology, School of Public Health, Fudan University, Shanghai* 200032, China; 2 *Key Laboratory on Public Health Safety, Ministry of Education*; 3 *Changhai Hospital, Second Military Medical University*; 4 *Mathematical Institute, Shanghai Normal University*

Corresponding author: *JIANG Qing-wu*, Email: jiangqw@fudan.edu.cn

This work was supported by grants from the National High Technology Research and Development Program of China (*863 Program*) (No. 2006AA02Z402) *and Construction of Key Discipline of Shanghai Municipality* (No. B118).

【Introduction】 To estimate the tolerance limit and the tolerance interval of disease incubation, under the theory that the observations (samples) were subject to Poisson distribution, the tolerance limits and tolerance interval of disease incubation were calculated based on beta-distribution with integer parameter. Expressions on the relation were obtained, including the statistics on tolerance limits of both minimum and maximum orders while the tolerance was the difference between minimum and maximum order statistics and the parameters of Poisson distribution on the sample size. Using the incomplete observations as an example, reasonable unit of sample size was considered and chosen.

【Key words】 Incubation tolerance limits; Incubation tolerance interval; Compound β-distribution with integer parameters; Parameters of distribution on sample size

疾病潜伏期是传染病流行病学研究的重要指标。疾病潜伏期均数、方差等数据特征估计及其容许限、容许区间推算是流行病学调查与资料分析的基础内容。传染病发病参数包括患者发病时间、接触（感染者）时间、潜伏期及其容许限和容许区间。发病时间包括患者接触（感染者）时间与接触后至发病的潜伏期。接触时间可观测时，发病参数估计包括潜伏期均数和方差估计及潜伏期容许限和容许区间估计；接触时间模糊（不能确切观测）时，发病参数估计包括接触时间与潜伏期均数和方差估计及接触时间与潜伏期容许限和容许区间估计，均可采用分布法和非分布法进行估计。对于完整调查资料，可用一般统计学方法估计疾病潜伏期并推算其容许限或容许区间[1-3]。对于不完整调查资料，蔡全才等[4,5]提出了“假定感染时间在接触期内服从均匀分布”的潜伏期估计“条件非参数法”，姜庆五等[6]提出了“假定感染时间在接触期内服从均匀分布且接触期（区间大小）服从伽玛分布”的潜伏期估计“条件参数法”。笔者应用复合整参数β-分布也提出了不完整观测值潜伏期容许限、容许区间的推算方法，为疾病潜伏期的计算提供理论依据。

DOI: 10.3760/cma.j.issn.0254-6450.2011.12.024

基金项目：国家高技术研究发展计划（“863”计划）(2006AA02Z402)；上海市重点学科建设项目(B118)

作者单位：200032上海，复旦大学公共卫生学院、公共卫生安全教育部重点实验室（赵飞、姜庆五）；第二军医大学长海医院（蔡全才）；上海师范大学数理学院（陈启明）

通信作者：姜庆五，Email: jiangqw@fudan.edu.cn

一、基本原理

记T(>0)表示感染期，F(t)表示（个体）感染时间分布函数，$t_1<t_2<\cdots<t_n$是T的n容量样本次序统计量观测值。则随机变量X=F(t)服从（标准）均匀

分布，X～U(0,1)。此时$X_{(1)}=F(t_1)$、$X_{(n)}=F(t_n)$，以及$W=X_{(n)}-X_{(1)}$的密度函数[6]

$$h_1(x_{(1)}\mid n)=n(1-x_{(1)})^{n-1}$$

$$h_1(x_{(n)}\mid n)=nx_{(n)}^{n-1}$$

$$h(w\mid n)=n(n-1)w^{n-2}(1-w)$$

分别服从(a=1,b=n),(a=n,b=1),(a=$n-1$,b=2)的β-分布：$X_{(1)}\sim Be(1,n)$，$X_{(n)}\sim Be(n,1)$，$W\sim Be(n-1,2)$。

由此得到容许概率τ的容许度γ，即(τ,γ)的容许上限

$$\begin{aligned}P[F(t_U)\geqslant\tau]&=P(X_{(n)}\geqslant\tau)=\int_\tau^1 nx_{(n)}^{n-1}dx_{(n)}\\&=1-\tau^n\geqslant\gamma\end{aligned}\tag{1-1}$$

与(τ,γ)的容许下限

$$\begin{aligned}P[1-F(t_L)\geqslant\tau]&=\int_0^{1-\tau}n(1-x_{(1)})^{n-1}dx_{(1)}\\&=1-\tau^n\geqslant\gamma\end{aligned}\tag{1-2}$$

及容许区间

$$\begin{aligned}&P[F(t_n)-F(t_1)\geqslant\tau]=P(X_{(n)}\geqslant X_{(1)}+\tau)\\&=\int_0^{1-\tau}\left[\int_{X(1)+\tau}^1 n(n-1)(x_{(n)}-x_{(1)})^{n-2}dx_{(n)}\right]dx_{(1)}\geqslant\gamma\end{aligned}\tag{1-3}$$

说明感染时间百分位数或其相应容许限的容许概率容许度都与样本量n有关[2,3]。按式(1-1)、(1-2)，同时有

$$n\geqslant \mathrm{Ln}(1-\gamma)/\mathrm{Ln}\tau\tag{1-4}$$

按式(1-3)有

$$n\tau^{n-1}-(n-1)\tau^n\leqslant 1-\gamma\tag{1-5}$$

n满足关系式(1-4)时，样本的最大观测值t_n与最小观测值t_1分别为(τ,γ)的容许上限与(τ,γ)的容许下限。n满足关系式(1-5)时，由样本的最小观测值t_1与最大观测值t_n组成的区间(t_1,t_n)即为(τ,γ)容许区间。

二、不完全观测潜伏期容许限容许区间推算的复合整参数β-分布法

不完整的调查资料只记录了受感染者最早(首次)接触感染者时间($T=t_1$)或最迟一次接触感染者时间($T=t_n$)或同时记录到$T=t_1$和$T=t_n$。t_1与t_n间隔时间即$t_2<t_3<\cdots<t_n$或$t_1<t_2<\cdots<t_{n-1}$或$t_2<t_3<\cdots<t_{n-1}$。相应的接触次数$n-1$或$n-2$未被考虑，故n是≥1的整值离散型随机变量。而复合(混合)整参数β-分布法是指定义在(0,1)区间，可视作连续型整参数β-分布Be(k+1,n−k)对于离散型参数n服从参数为λ的截0至k的截尾Poisson分布的混合分布，称之为参数(k,λ)的复合(混合)整参数β-Poisson分布，即BP(k,λ)[6]。

假定n服从参数为λ的截尾Poisson分布，对于密度函数$h_1(x_{(1)}\mid n)=n(1-x_{(1)})^{n-1}$或$h_n(x_{(n)}\mid n)=nx_{(n)}^{n-1}$，假定$n$服从截"0"Poisson分布

$$P(n)=\frac{1}{1-e^{-\lambda}}\frac{\lambda^n}{n!}e^{-\lambda},n=1,2,3,\cdots\cdots\tag{2-1}$$

对于密度函数$h(w\mid n)=n(n-1)w^{n-2}(1-w)$，假定$n$服从截"0,1"Poisson分布

$$P(n)=\frac{1}{1-e^{-\lambda}-\lambda e^{-\lambda}}\frac{\lambda^n}{n!}e^{-\lambda},n=2,3,4,\cdots\cdots\tag{2-2}$$

此时$X_{(1)}$对于n的边缘(或称复合)分布密度函数为

$$h_1(x_{(1)})=\begin{cases}\dfrac{\lambda}{1-e^{-\lambda}}e^{-\lambda x_{(1)}} & 0<x_{(1)}<1\\ 0 & x_{(1)}\leqslant 0,x_{(1)}\geqslant 1\end{cases}\tag{2-3}$$

利用$X_{(n)}$与$X_{(1)}$的相对性，将式(2-3)中$X_{(1)}$换成($1-X_{(n)}$)得$X_{(n)}$对于n的边缘分布密度函数

$$h_n(x_{(n)})=\begin{cases}\dfrac{\lambda}{1-e^{-\lambda}}e^{-\lambda(1-x_{(n)})} & 0<x_{(n)}<1\\ 0 & x_{(n)}\leqslant 0,x_{(n)}\geqslant 1\end{cases}\tag{2-4}$$

W对于n的边缘分布密度函数为

$$\begin{aligned}h(w)&=\sum_{n=2}^{\infty}h(w\mid n)P(n)\\&=\sum_{n=2}^{\infty}n(n-1)w^{n-2}(1-w)\frac{1}{1-e^{-\lambda}-\lambda e^{-\lambda}}\frac{\lambda^n}{n!}e^{-\lambda}\\&=\begin{cases}\dfrac{\lambda^2}{1-e^{-\lambda}-\lambda e^{-\lambda}}(1-w)e^{-\lambda(1-w)} & 0<w<1\\ 0 & w\leqslant 0,w\geqslant 1\end{cases}\end{aligned}\tag{2-5}$$

与$n=1$对应的$X_{(n)}$、$X_{(1)}$的数学期望与方差分别为

$$\begin{aligned}E(x_n)&=E(x_1)=\frac{1}{\lambda}\\D(x_n)&=D(x_1)=\frac{2[1-(1+\lambda)e^{-\lambda}]}{\lambda^2(1-e^{-\lambda})}-\frac{1}{\lambda^2}\end{aligned}\tag{2-6}$$

与$n=2$对应的W其一、二阶矩分别为

$$\begin{aligned}E(w)&=\frac{2[1-(1+\lambda+\lambda^2)e^{-\lambda}]}{1-(1+\lambda)e^{-\lambda}}\\D(w)&=\frac{6[1-(1+\lambda+\lambda^2/2)e^{-\lambda}]}{\lambda^2[1-(1+\lambda)e^{-\lambda}]}\\&\quad-\left[\frac{2[1-(1+\lambda+\lambda^2)e^{-\lambda}]}{1-(1+\lambda)e^{-\lambda}}\right]^2\end{aligned}\tag{2-7}$$

由式(2-3)、(2-4)、(2-5)表示的$X_{(n)}$、$X_{(1)}$与W(对于n)的边缘分布即是复合整参数β-分布BP(k,λ)在k=0,1时的特例[7,8]。这时以t_1为容许下限，以t_n为容许上限，以(t_1,t_n)为容许区间。容许概率为τ的容许度γ分别为

$$\begin{aligned}P(X_{(1)}\leqslant 1-\tau)&=\int_0^{1-\tau}h_1(x)dx=\frac{\lambda}{1-e^{-\lambda}}\int_0^{1-\tau}e^{-\lambda x}dx\\&=\frac{1}{1-e^{-\lambda}}(1-e^{-\lambda(1-\tau)})=\gamma\end{aligned}\tag{2-8}$$

$$P(X_{(n)} \geqslant \tau) = \int_{\tau}^{1} h_n(x)dx = \frac{\lambda}{1-e^{-\lambda}}\int_{\tau}^{1} e^{-\lambda(1-x)}dx = \frac{1}{1-e^{-\lambda}}(1-e^{-\lambda(1-\tau)}) = \gamma \tag{2-9}$$

$$P(W \geqslant \tau) = \int_{\tau}^{1} h(w)dw = \frac{\lambda^2}{1-e^{-\lambda}-\lambda e^{-\lambda}}\int_{\tau}^{1}(1-w)e^{-\lambda(1-x)}dx = \frac{1}{1-(1+\lambda)e^{-\lambda}}\{1-[1+\lambda(1-\tau)e^{-\lambda(1-\tau)}]\} = \gamma \tag{2-10}$$

三、举例

若观测得到个体第一次接触感染时间$t_1=1.72$，或最末一次接触感染时间$t_n=9.08$，则按式(2-8)或(2-9)，容许概率$\tau=80\%$，容许度$\gamma=80\%$，即$(\tau,\gamma)=(0.80,0.80)$。感染时间$t_1<1.72$（容许下限）或$t_n>9.08$（容许上限）对应参数为$\lambda\geqslant9.12$，或按式(2-10)$(\tau,\gamma)=(0.80,0.80)$感染时间在(1.72, 9.08)（容许区间）对应参数为$\lambda\geqslant7.63$。根据潜伏期容许限、容许区间推算公式便可方便计算出疾病潜伏期容许限、容许区间。记t_d为个体发病时间，则t_d-t_1，t_d-t_n以及(t_d-t_n, t_d-t_1)分别即为与(τ,γ)对应的疾病潜伏期容许上、下限及容许区间。若观测得到该个体的发病时间$t_d=12.46$，则在$(\tau,\gamma)=(0.80,0.80)$及上述（不同）$\lambda$值条件下，疾病潜伏期容许上限为$12.46-1.72=10.74$，容许下限为$12.46-9.08=3.38$，以及容许区间为(3.38, 10.74)。

四、讨论

公式(1-1)至(1-3)中的(τ,γ)要求与样本量n对应；式(2-8)至(2-10)中(τ,γ)要求与样本量n的分布参数λ对应。参数λ大小与样本量n的取样单位大小有关。取样单位小，n值大，n的分布参数即期望值λ大。由此推出，为实现所要求的容许概率容许度(τ,γ)，与前者增加样本量n相应，后者可缩小感染时间间隔单位以增加样本时间间隔数n。在举例中，若容许概率、容许度分别提高10%，即$(\tau,\gamma)=(0.90,0.90)$，则分别要求$\lambda\geqslant18.23$，$\lambda\geqslant15.27$相当于原来的2倍。故若原样本量$n$及其分布参数$\lambda$单位为天(d)，可改取单位为0.5 d(12 h)。

参 考 文 献

[1] Jiang QW. Epidemiology. Beijing: Science Press, 2003. (in Chinese)
姜庆五. 流行病学. 北京:科学出版社,2003.

[2] Jiang QW, Chen QM. Methods and models of epidemiology. Shanghai: Fudan University Press, 2007. (in Chinese)
姜庆五,陈启明. 流行病学方法与模型. 上海:复旦大学出版社,2007.

[3] Mao SS, Wang JL, Pu XL. Advanced mathematical statistics. Beijing: Higher Education Press, 1998. (in Chinese)
茆诗松,王静龙,濮晓龙. 高等数理统计. 北京:高等教育出版社,1998.

[4] Cai QC, Jiang QW, Cheng X, et al. A non-parametric maximum likelihood estimation method for estimating infection curve of infectious diseases. Acad J Second Military Medical University, 2004(12):1349-1352. (in Chinese)
蔡全才,姜庆五,程翔,等. 一种用于推算感染时间曲线的非参数极大似然估计方法. 第二军医大学学报,2004(12):1349-1352.

[5] Cai QC, Jiang QW, Xu QF, et al. To develop a model on severe acute respiratory syndrome epidemics to quantitatively evaluate the effectiveness of intervention measures. Chin J Epidemiol, 2005, 26(3):153-158. (in Chinese)
蔡全才,姜庆五,徐勤丰,等. 定量评价SARS干预措施效果的传播动力学模型. 中华流行病学杂志,2005,26(3):153-158.

[6] Jiang QW, Zhou YB, Cai QC, et al. Estimate of disease incubation for complete observations. Chin J Prev Med, 2007, 41(2):157-158. (in Chinese)
姜庆五,周艺彪,蔡全才,等. 完全观察潜伏期估计. 中华预防医学杂志,2007,41(2):157-158.

[7] Chen QM, Gu LQ. Analysis and implementation of Beta-distribution with mixed integer parameter. Chin J Appl Prob Stat, 2010(2):220-224. (in Chinese)
陈启明,顾龙全. 混合整参数贝塔分布及其应用. 应用概率统计,2010(2):220-224.

[8] Wang JN, Zhang ZJ, Chen QM, et al. Parametric and non-parametric approaches for human reference range. Fudan University J Med Sci, 2010(6):715-718. (in Chinese)
王静亮,张志杰,陈启明,等. 人体参考值范围的参数与非参数推算方法. 复旦学报(医学版),2010(6):715-718.

（收稿日期:2011-06-17）
（本文编辑:张林东）

疾病潜伏期容许限容许区间的推算

作者： 赵飞， 蔡全才， 陈启明， 姜庆五， ZHAO Fei， CAI Quan-cai， CHEN Qi-ming， JIANG Qing-wu
作者单位： 赵飞,姜庆五,ZHAO Fei,JIANG Qing-wu(复旦大学公共卫生学院、公共卫生安全教育部重点实验室,上海,200032)， 蔡全才,CAI Quan-cai(第二军医大学长海医院)， 陈启明,CHEN Qi-ming(上海师范大学数理学院)

刊名： 中华流行病学杂志ISTIC PKU
英文刊名： Chinese Journal of Epidemiology
年，卷(期)： 2011,32(12)

参考文献(8条)

1. 姜庆五 流行病学 2003

2. 姜庆五;陈启明 流行病学方法与模型 2007

3. 茆诗松;王静龙;濮晓龙 高等数理统计 1998

4. 蔡全才;姜庆五;程翔 一种用于推算感染时间曲线的非参数极大似然估计方法[期刊论文]-第二军医大学学报 2004(12)

5. 蔡全才;姜庆五;徐勤丰 定量评价SARS干预措施效果的传播动力学模型[期刊论文]-中华流行病学杂志 2005(03)

6. 姜庆五;周艺彪;蔡全才 完全观察潜伏期估计[期刊论文]-中华预防医学杂志 2007(02)

7. 陈启明;顾龙全 混合整参数贝塔分布及其应用[期刊论文]-应用概率统计 2010(02)

8. 王静霓;张志杰;陈启明 人体参考值范围的参数与非参数推算方法[期刊论文]-复旦学报（医学版） 2010(06)