云南大学学报（自然科学版），2004, 26 (2): 112~114　　CN 53-1045/N　ISSN 0258-7971
Journal of Yunnan University

Poisson 分布中未知参数的精确置信区间*

杨建红[1,2], 胡　俊[1,2], 王清生[1]
(1.云南大学 统计系,云南 昆明　650091; 2.云南农业大学 基础与信息工程学院,云南 昆明　650201)

摘要: 用假设检验法构造了 Poisson 分布中未知参数的置信下限和上限, 及精确置信区间, 为实际应用提供了理论依据.

关键词: Poison 分布; 假设检验; 置信下限上限; 精确置信区间
中图分类号: O 212.2　　文献标识码: A　　文章编号: 0258-7971(2004)02-0112-03

Poisson 分布是一类重要的离散分布, 具有 Poisson 分布的随机现象在现实世界中是比较普遍的. 例如电话交换台收到的电话呼唤次数; 一页书中的印刷错误; 一段时间内的交通事故数等稀疏现象都可以用 Poisson 分布来描述. 应用中讨论 Poisson 分布未知参数的置信区间时, 一般是用大样本理论构造未知参数的近似置信区间. 但是当样本不是很大时, 用这样的方法构造出来的置信区间就不是很理想, 因此有一定的局限性.

文献[1]介绍了用假设检验法构造参数的置信区间, 文献[2, 3]介绍了用枢轴量法构造置信区间. 本文用文献[1]中介绍的假设检验方法构造了 Poisson 分布中未知参数 λ 的精确置信区间, 并且根据文献[4]中介绍的数学方法表示出了上下限表达式, 克服了样本多少的局限性, 为实际应用提供了理论依据. 类似的研究见文献[5, 6].

定理 1　设 $x_1, x_2, \cdots, x_n$ 是独立且服从参数为 $\lambda(\in(0,\infty))$ 的 Poissson 分布的随机变量, $s=\sum_{i=1}^{n} x_i$, 则

(i) $\underline{\lambda}(s,\alpha)$ 是 λ 的置信水平为 $1-\alpha$ 的置信下限, 其中 $\underline{\lambda}(s,\alpha)=\begin{cases}0, & s=0,\\ \dfrac{1}{2n}\chi^2_{2s}(\alpha), & s\geqslant 1.\end{cases}$

$\chi^2_m(\alpha)$ 表示自由度为 m 的 χ^2-分布的 α 分位点;

(ii) $\overline{\lambda}(s,\alpha)$ 是 λ 的置信水平为 $1-\alpha$ 的置信上限, 其中 $\overline{\lambda}(s,\alpha)=\dfrac{1}{2n}\chi^2_{2(s+1)}(1-\alpha)$, $s\geqslant 1$;

(iii) 当 $\alpha_1+\alpha_2=\alpha$ 时, $(\underline{\lambda}(s,\alpha_1), \overline{\lambda}(s,\alpha_2))$ 是 λ 的置信水平为 $1-\alpha$ 的置信区间.

证明　(i) 为了得到参数 λ 的精确置信下限, 考虑假设 $H: \lambda\leqslant\lambda_0, (\lambda>0)$ 的水平为 α 的检验

$$\delta_{k(\lambda_0,\alpha)}(x)=\begin{cases}1, & s\geqslant k,\\ 0, & s<k,\end{cases}$$

其中 $s=\sum_{i=0}^{n}x_i$, $k=k(\lambda_0,\alpha)$. 这里 $k(\lambda_0,\alpha)$ 是使得 $\sum_{r\geqslant k}\dfrac{(n\lambda)^r}{r!}e^{-n\lambda}\leqslant\alpha$ 的最小非负整数, s 服从参数为 $n\lambda$ 的 Poisson 分布.

记
$$c(x)=\{\lambda: \delta_{k(\lambda,\alpha)=0}\}=\{\lambda: s<k(\lambda,\alpha)\}=\{\lambda: s\leqslant k(\lambda,\alpha)-1\},$$
则
$$p\{\lambda\in c(x)\}=p\{\delta_{k(\lambda,\alpha)}(x)=0\}=1-p\{\delta_{k(\lambda,\alpha)}(x)=1\}\geqslant 1-\alpha,$$
即 $c(x)$ 是 λ 的水平为 $1-\alpha$ 的置信区间.

为了求置信下限, 我们需要证明下列几点:

(a) $k(\lambda,\alpha)$ 是 λ 的单调上升函数;

(b) $\lim_{\lambda\to\lambda_0} k(\lambda,\alpha)=k(\lambda_0,\alpha)$;

(c) $k(\lambda,\alpha)$ 作为 λ 的函数在其间断点处跃迁 1, 且 $k(\lambda,\alpha)$ 是左连续的;

(d) $k(0,\alpha)=1$; 当 $\lambda\to\infty$ 时, $k(\lambda,\alpha)\to\infty$;

因为 $k(\lambda,\alpha)$ 只取非负整数, 当(a), (b) 成立时(c), (d) 也成立.

验证**(a)**　因为 $p(s<j)=\sum_{i=0}^{j-1}\dfrac{(n\lambda)^i}{i!}e^{-n\lambda}$ 则

$$\frac{dp(s<j)}{d\lambda}=\frac{d}{d\lambda}\left\{e^{-n\lambda}+\sum_{i=1}^{j-1}\frac{(n\lambda)^i}{i!}e^{-n\lambda}\right\}=$$

* 收稿日期: 2003-09-25
基金项目: 国家自然科学基金资助项目(10261009); 国家教育部科研重点基金资助项目(00246).
作者简介: 杨建红(1972-　), 女, 云南人, 硕士生, 主要从事多元统计方面的研究.

$$-ne^{-n\lambda}+\sum_{i=1}^{j-1}\left[\frac{n(n\lambda)^{i-1}}{(i-1)!}-\frac{n(n\lambda)^{i}}{i!}\right]e^{-n\lambda}=$$

$$-ne^{-n\lambda}+ne^{-n\lambda}-\frac{n(n\lambda)^{j-1}}{(j-1)!}e^{-n\lambda}=$$

$$-\frac{n(n\lambda)^{j-1}}{(j-1)!}e^{-n\lambda}<0.$$

即 $p(s<j)$ 是 λ 的非增函数, 又因为 $p(s\geqslant j)=1-p(s<j)$, 所以 $p(s\geqslant j)$ 是 λ 的非降函数, 并且对于固定的 λ, $p(s\geqslant j)$ 是 j 的非增函数.

因此, 若 $\lambda_1<\lambda_2$ 时有 $k(\lambda_1,\alpha)>k(\lambda_2,\alpha)$, 即 $k(\lambda_1,\alpha)-1\geqslant k(\lambda_2,\alpha)$, 那么

$$\begin{aligned}\alpha&\geqslant p_{\lambda_2}\{s\geqslant k(\lambda_2,\alpha)\}\geqslant\\&p_{\lambda_2}\{s\geqslant k(\lambda_1,\alpha)-1\}\geqslant\\&p_{\lambda_1}\{s\geqslant k(\lambda_1,\alpha)-1\}>\alpha,\end{aligned}$$

这是矛盾的. 所以当 $\lambda_1<\lambda_2$ 时, 有 $k(\lambda_1,\alpha)<k(\lambda_2,\alpha)$ 即 $k(\lambda,\alpha)$ 是 λ 的单调增加函数.

验证**(b)**　若 λ_0 是 $k(\lambda,\alpha)$ 的连续点, (b) 显然成立. 若 λ_0 是 $k(\lambda,\alpha)$ 的间断点时, 记 $j=\lim\limits_{\lambda\to\lambda_0^-}k(\lambda,\alpha)$. 则由(a) 知 $j\leqslant k(\lambda_0,\alpha)$, 且 $p_{\lambda_0}(s\geqslant j)\leqslant\alpha$, $\forall\lambda<\lambda_0$. 另一方面由 $k(\lambda_0,\alpha)$ 的意义可知: $j\geqslant k(\lambda_0,\alpha)$, 所以 $j=k(\lambda_0,\alpha)$. 即(b) 成立.

因此, 定义 $\underline{\lambda}(s)=\inf\{\lambda: k(\lambda,\alpha)=s+1\}$ 则

$$c(x)=\{\lambda: k(\lambda,\alpha)=s+1\}=\begin{cases}[\underline{\lambda}(s),+\infty), & s>0,\\ [0,+\infty), & s=0.\end{cases}$$

由前述知, $c(x)$ 为 λ 的水平为 $1-\alpha$ 的置信区间. 当 $s>0$ 时有 $k(\underline{\lambda}(s))=s$, 当 $s=0$ 时 $\underline{\lambda}(s)=0$, 因此 $\underline{\lambda}(s)$ 为 λ 的水平为 $1-\alpha$ 的置信下限.

当 $s>0$ 时, 由 $\underline{\lambda}(s)$ 的意义可知: 存在点列 $\{\lambda_i\}$, 使得 $\lambda_i\to\underline{\lambda}(s)$, $i\to\infty$, 且 $k(\lambda_i,\alpha)=s+1$, 故由 $k(\lambda_i,\alpha)$ 的意义知: $\sum\limits_{r\geqslant s}\frac{(n\lambda_i)^r}{r!}e^{-n\lambda_i}\geqslant\alpha$ 令 $i\to\infty$ 得

$$\sum_{r\geqslant s}\frac{(n\underline{\lambda}(s))^r}{r!}e^{-n\underline{\lambda}(s)}\geqslant\alpha,\tag{1}$$

而由 $k(\underline{\lambda}(s),\alpha)=s$ 知

$$\sum_{r\geqslant s}\frac{(n\underline{\lambda}(s))^r}{r!}e^{-n\underline{\lambda}(s)}\leqslant\alpha,\tag{2}$$

由(1), (2) 知

$$\sum_{r\geqslant s}\frac{(n\underline{\lambda}(s))^r}{r!}e^{-n\underline{\lambda}(s)}=\alpha.\tag{3}$$

又因为

$$\sum_{r=0}^{s-1}\frac{(n\underline{\lambda}(s))^r}{r!}e^{-n\underline{\lambda}(s)}+\sum_{r\geqslant s}\frac{(n\underline{\lambda}(s))^r}{r!}e^{-n\underline{\lambda}(s)}=1,$$

所以(3) 式可变为

$$\sum_{r=0}^{s-1}\frac{(n\underline{\lambda}(s))^r}{r!}e^{-n\underline{\lambda}(s)}=1-\alpha.\tag{4}$$

即, 参数 λ 的精确下限 $\underline{\lambda}(s)$ 为关于 λ 的方程

$$\sum_{r=0}^{s-1}\frac{(n\lambda)^r}{r!}e^{-n\lambda}=1-\alpha\tag{5}$$

的解, 由验证(a): $P(T<s)=\sum\limits_{r\geqslant s}\frac{(n\lambda)^r}{r!}e^{-n\lambda}$ 是关于 λ 的单调函数, 所以(5) 的解是唯一的.

(ii) 为了得到参数 λ 的精确置信上限考虑假设 $H: \lambda\geqslant\lambda_0$, $(\lambda>0)$ 的水平为 α 的检验

$$\delta_{(\lambda_0,\alpha)}(x)=\begin{cases}1, & s\leqslant j,\\ 0, & s>j,\end{cases}$$

其中 $s=\sum\limits_{i=0}^{n}x_i$, $j=j(\lambda_0,\alpha)$. 这里 $j(\lambda_0,\alpha)$ 是使得 $\sum\limits_{r=0}^{j}\frac{(n\lambda)^r}{r!}e^{-n\lambda}\leqslant\alpha$ 的最大非负整数, s 服从参数为 $n\lambda$ 的 Poisson 分布. 记

$$\begin{aligned}D(x)&=\{\lambda: \delta_{(\lambda,\alpha)}=0\}=\{\lambda: s>j(\lambda,\alpha)\}=\\&\{\lambda: s\geqslant j(\lambda,\alpha)-1\},\end{aligned}\tag{6}$$

则 $p\{\lambda\in D(x)\}=p\{\delta_{j(\lambda,\alpha)}(x)=0\}=1-p\{\delta_{(\lambda,\alpha)}(x)=1\}\geqslant1-\alpha$ 即 $D(x)$ 是 λ 的水平为 $1-\alpha$ 的置信区间.

为了求置信上限, 我们需要证明下列几点:

(e) $j(\lambda,\alpha)$ 是 λ 的单调上升函数;

(f) $\lim\limits_{\lambda\to\lambda_0^+}j(\lambda,\alpha)=j(\lambda_0,\alpha)$;

(g) $j(\lambda,\alpha)$ 作为 λ 的函数在其间断点处跃迁 1, 且 $k(\lambda,\alpha)$ 是右连续的;

(h) $k(0,\alpha)=-1$; 当 $\lambda\to\infty$ 时, $k(\lambda,\alpha)\to\infty$; 因为 $j(\lambda,\alpha)$ 只取非负整数, 当(e), (f) 成立时 (g), (h) 也成立. 验证的方法与(i) 相同. 令

$$\bar{\lambda}(s)=\sup\{\lambda: j(\lambda,\alpha)=s-1\},$$

则由此及(6) 知

$$D(x)=\{\lambda: j(\lambda,\alpha)\leqslant s-1\}=\begin{cases}[0,\bar{\lambda}(s)), & 0<s<\infty,\\ [0,+\infty), & s=0,\end{cases}$$

由前述知, $D(x)$ 为 λ 的水平为 $1-\alpha$ 的置信上限.

当 $s\geqslant0$ 时, 有 $j(\bar{\lambda}(s))=s$, 因此 $\bar{\lambda}(s)$ 为 λ 的置信上限.

当 $s>0$ 时, 由 $\bar{\lambda}(s)$ 的意义知存在点列 $\{\lambda_i\}$ 使得 $\lambda_i\to\bar{\lambda}(s)$, $i\to\infty$, 且由 $j(\lambda_i,\alpha)$ 的意义知

$$\sum_{r=0}^{s}\frac{(n\lambda_i)^r}{r!}e^{-n\lambda_i}\geqslant\alpha,$$

令 $i \to \infty$ 有 $\lambda_i \to \bar{\lambda}(s)$ 即

$$\sum_{r=0}^{s} \frac{(n\bar{\lambda}(s))^r}{r!} e^{-n\bar{\lambda}(s)} \geqslant \alpha, \tag{7}$$

而由 $j(\bar{\lambda}(s), \alpha) = s$ 知

$$\sum_{r=0}^{s} \frac{(n\bar{\lambda}(s))^r}{r!} e^{-n\bar{\lambda}(s)} \leqslant \alpha, \tag{8}$$

由(7),(8)知

$$\sum_{r=0}^{s} \frac{(n\bar{\lambda}(s))^r}{r!} e^{-n\bar{\lambda}(s)} = \alpha$$ 即参数 λ 的精确上限是关于 λ 的方程

$$\sum_{r=0}^{s} \frac{(n\lambda)^r}{r!} e^{-n\lambda} = \alpha \tag{9}$$

的解. 并由(a)的证明知 $\sum_{r=0}^{s} \frac{(n\lambda)^r}{r!} e^{-n\lambda}$ 是关于 λ 的单调函数所以(9)的解是唯一的.

由数学关系易知

$$\sum_{r=0}^{s} \frac{(n\lambda)^r}{r!} e^{-n\lambda} = \int_{\lambda}^{+\infty} \frac{n^{s+1}}{\Gamma(s+1)} u^s e^{-nu} du = \int_{2n\lambda}^{+\infty} \frac{v^s e^{-\frac{v}{2}}}{\Gamma(s+1) \cdot 2^{s+1}} dv, \tag{10}$$

$s = 1, 2, \cdots\cdots$ 则方程(5)可表示为

$$\int_0^{2n\lambda} \frac{1}{\Gamma(s) 2^s} v^{s-1} e^{-\frac{v}{2}} dv = \alpha. \tag{11}$$

方程(9)可表示为

$$\int_0^{2n\lambda} \frac{1}{\Gamma(s+1) 2^{s+1}} v^s e^{-\frac{v}{2}} dv = 1 - \alpha. \tag{12}$$

记 $\chi^2(n)$ 的分布函数为 $F(x) = \int_0^x t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt$, 得到未知参数 λ 的置信水平为 $1-\alpha$ 的置信下限为 $\underline{\lambda}(s) = \frac{1}{2n}\chi^2_{\alpha}(2s)$, 上限为 $\bar{\lambda}(s) = \frac{1}{2}\chi^2_{1-\alpha}(2s+2)$.

(iii) 由(i)知, 方程 $\sum_{r=0}^{s-1} \frac{(n\lambda)^r}{r!} e^{-n\lambda} = 1 - \alpha_1$ 关于 λ 的解是置信水平为 $1-\alpha_1$ 的置信下限, 由(ii)知, 方程 $\sum_{r=0}^{s} \frac{(n\lambda)^r}{r!} e^{-n\lambda} = \alpha_2$ 关于 λ 的解是置信水平为 $1-\alpha_2$ 的置信上限, 则

$$\begin{aligned} & P_\lambda\{\lambda \in [\underline{\lambda}(s), \bar{\lambda}(s)]\} = \\ & P_\lambda\{\underline{\lambda}(s) \leqslant \lambda \leqslant \infty, \text{且 } 0 \leqslant \lambda \leqslant \bar{\lambda}(s)\} = \\ & P_\lambda\{\underline{\lambda}(s) \leqslant \lambda \leqslant \infty\} + P_\lambda\{0 \leqslant \lambda \leqslant \bar{\lambda}(s)\} - \\ & P_\lambda\{\underline{\lambda}(s) \leqslant \lambda \leqslant \infty, \text{或 } 0 \leqslant \lambda \leqslant \bar{\lambda}(s)\} \geqslant \\ & (1-\alpha_1) + (1-\alpha_2) - 1 = 1 - \alpha. \end{aligned}$$

由此得到了未知参数的精确的置信区间

$$[\underline{\lambda}(s), \bar{\lambda}(s)] = \left[\frac{1}{2n}\chi^2_{\alpha_1}(2s), \frac{1}{2n}\chi^2_{1-\alpha_2}(2s+2)\right],$$

$\alpha_1 + \alpha_2 = \alpha$.

参考文献:

[1] 茆诗松, 王静龙, 濮晓龙. 高等数理统计[M]. 北京: 高等教育出版社, 1998.

[2] PETER J B, KJELL A D. Mathematical statistics[M]. San Francisco: Holderr Day, Inc, 1977.

[3] 范金城, 吴可法. 统计推断导引[M]. 北京: 科学出版社, 2001.

[4] 高惠璇. 统计计算[M]. 北京: 北京大学出版社, 1999.

[5] 石 磊, 李兴绪, 陈 飞, 等. 生产函数模型中参数的 Minimax 估计及应用[J]. 云南大学学报(自然科学版), 2003, 25(2): 94—100.

[6] 白 鹏. 线性模型中回归系数的一种新估计[J]. 云南大学学报(自然科学版), 1996, 18(4): 388—392.

Exact confidence interval of unknown parameter in Poisson distribution

YANG Jian-hong[1,2], HU Jun[1,2], WANG Qing-sheng[1]

(1. Department of Statistics, Yunnan University, Kunming 650091, China;

2. Foundation and Information Engineering Academe, Yunnan Agriculture University, Kunming 650201, China)

Abstract: The confidence lower bound, upper bound, and accurate confidence interval of unknown parameter in Poisson distribution are obtained by employing the method of hypothesis testing. It also provided theoretical basis for practical application.

Key words: Poisson distribution; hypothesis testing; confidence lower bound and upper bound; confidence interval

MSC(2000): 62F25