应用简报

混合整参数贝塔分布及其应用

陈启明　　顾龙全

(上海师范大学数理学院, 上海, 200234)

摘　要

本文给出了一个新的指数型分布, 它是整参数贝塔分布Be$(k+1,n-k)$对于整数n的截0至k的截尾泊松分布的混合分布. 论文给出了这一分布的统计意义和数字特征, 并通过一个例子说明了它的一个应用.

关键词: 贝塔分布, 截0至k的截尾泊松分布, 混合整参数贝塔—泊淞分布, 容许限.

学科分类号: O212.

《应用概率统计》版权所用

§1. 引　　言

标准均匀分布, 或更为一般的贝塔分布是定义在$(0,1)$区间上的最常用的连续型分布, 它在随机数生成及新分布的构造中起着非常重要的作用. 而在离散型分布中, 泊松分布在理论及在金融、保险、生物医药等领域都有重要的应用, 特别是它关于其它连续分布的混合, 例如泊松分布关于伽玛分布的混合分布构成负二项分布, 它是一类重要的风险模型[4]. 它关于其它连续分布的混合可参见文献[5]. 上面讨论的是离散型分布对于其连续型参数(分布)的混合分布, 其结果是离散分布. 本文考虑的是连续型分布对于其离散型参数(分布)的混合分布, 给出了一个新的连续型分布. 此分布定义在$(0,1)$区间上, 可视作连续型整参数贝塔分布Be$(k+1,n-k)$对于离散型参数n服从参数为λ的截0至k的截尾泊松分布的混合分布, 称之为参数为(k,λ)的混合整参数贝塔—泊淞分布, 记为BP(k,λ). 本文第二节将直接给出此分布的密度函数及主要的数字特征, 第三节给出了此分布的由来, 第四节给出了一个应用例子.

§2. 分布及其数字特征

参数为(k,λ)的混合整参数贝塔—泊淞分布BP(k,λ)的密度函数为

$$g_k(x)=\frac{\lambda}{f_k(\lambda)}\frac{(\lambda x)^k}{k!}e^{-\lambda x},\qquad x\in(0,1),\tag{2.1}$$

本文2009年5月5日收到.

其中$k=1,2,\cdots$, 参数函数

$$f_k(\lambda)=1-\sum_{i=0}^{k}\frac{\lambda^i}{i!}e^{-\lambda},\qquad \lambda>0.$$

由[1]知$g_k(x)$是一个指数型分布密度, 其分布函数及一、二阶(原点)矩分别为

$$G_k(x)=\frac{1}{f_k(\lambda)}\Big(1-\sum_{j=0}^{k}\frac{1}{j!}(\lambda x)^j e^{-\lambda x}\Big), \tag{2.2}$$

$$m_1=\frac{k+1}{\lambda f_k(\lambda)}\Big(1-\sum_{j=0}^{k}\frac{1}{j!}\lambda^j e^{-\lambda}\Big), \tag{2.3}$$

$$m_2=\frac{(k+1)(k+2)}{\lambda^2 f_k(\lambda)}\Big(1-\sum_{j=0}^{k+1}\frac{1}{j!}\lambda^j e^{-\lambda}\Big). \tag{2.4}$$

特别当$k=0$或1时的参数函数、密度函数与一、二阶矩如表下表所示

表1　$k=0,1$时BP(k,λ)的分布及数字特征

k	0	1
$f_k(\lambda)$	$1-e^{-\lambda}$	$1-(1+\lambda)e^{-\lambda}$
$g_k(x)$	$\dfrac{\lambda}{1-e^{-\lambda}}e^{-\lambda x}$	$\dfrac{\lambda^2 x}{1-(1+\lambda)e^{-\lambda}}e^{-\lambda x}$
m_1	$\dfrac{1}{\lambda(1-e^{-\lambda})}(1-e^{-\lambda})$	$\dfrac{2[1-(1+\lambda+\lambda^2)e^{-\lambda}]}{1-(1+\lambda)e^{-\lambda}}$
m_2	$\dfrac{2[1-(1+\lambda)e^{-\lambda}]}{\lambda^2(1-e^{-\lambda})}$	$\dfrac{6[1-(1+\lambda+\lambda^2/2)e^{-\lambda}]}{\lambda^2[1-(1+\lambda)e^{-\lambda}]}$

§3.　分布的由来

设随机变量X服从分布$F(x,\theta)$, 其形式未知, $x_1,x_2,\cdots,x_n$为来自于分布$F(x,\theta)$的容量为n的样本, $x_{(1)},x_{(2)},\cdots,x_{(n)}$为其次序统计量, 令$y_k=F(x_{(k)},\theta)$, 则$y_1$, y_n和$w=y_n-y_1$的密度函数分别为

$$h_1(y_1|n)=n(1-y_1)^{n-1},$$
$$h_n(y_n|n)=ny_n^{n-1},$$
$$h(w|n)=n(n-1)w^{n-2}(1-w),$$

即它们服从贝塔分布: $y_1\sim \mathrm{Be}(1,n)$, $y_n\sim \mathrm{Be}(n,1)$, $w\sim \mathrm{Be}(n-1,2)$.

应用中样本容量n有时是随机的. 例如, 推算某类(多产)家畜幼崽的体重容许限与容许区间, 一窝家畜(样本)的幼崽数n即是个随机变量.

假定样本量n服从参数为λ的截尾泊松分布, 在此先考虑三个特殊的截尾分布:

《应用概率统计》版权所用

1. 截0的泊松分布, 其分布律为

$$\mathrm{P}_0(n)=\frac{1}{1-e^{-\lambda}}\frac{\lambda^n}{n!}e^{-\lambda},\qquad n=1,2,3,\cdots. \tag{3.1}$$

2. 截0, 1的泊松分布, 其分布律为

$$\mathrm{P}_{0-1}(n)=\frac{1}{1-e^{-\lambda}-\lambda e^{-\lambda}}\frac{\lambda^n}{n!}e^{-\lambda},\qquad n=2,3,\cdots. \tag{3.2}$$

3. 截0至k的泊松分布, 其分布律为

$$\mathrm{P}_{0-k}(n)=\frac{1}{\left[1-\sum\limits_{i=0}^{k}\frac{\lambda^i}{i!}e^{-\lambda}\right]}\sum_{n=k+1}^{\infty}\frac{\lambda^n}{n!}e^{-\lambda},\qquad n=k+1,k+2,\cdots. \tag{3.3}$$

这时y_1对于截0的泊松分布$\mathrm{P}_0(n)$的边际(或称混合)分布的密度函数为

$$h_1(y_1)=\frac{\lambda}{1-e^{-\lambda}}e^{-\lambda y_1},\qquad 0<y_1<1. \tag{3.4}$$

利用y_1与y_n的相对性, 把(3.4)式中y_1换成$(1-y_n)$得y_n对于截0的泊松分布$\mathrm{P}_0(n)$的边际(或称混合)分布的密度函数为

$$h_n(y_n)=\frac{\lambda}{1-e^{-\lambda}}e^{-\lambda(1-y_n)},\qquad 0<y_n<1. \tag{3.5}$$

w对于截0, 1的泊松分布$\mathrm{P}_{0-1}(n)$的边际(或称混合)分布的密度函数为

$$\begin{aligned}
h(w) &= \sum_{n=2}^{\infty}h(w|n)\mathrm{P}(n)\\
&= \sum_{n=2}^{\infty}n(n-1)w^{n-2}(1-w)\frac{1}{1-e^{-\lambda}-\lambda e^{-\lambda}}\frac{\lambda^n}{n!}e^{-\lambda}\\
&= \frac{\lambda^2}{1-e^{-\lambda}-\lambda e^{-\lambda}}(1-w)e^{-\lambda(1-w)},\qquad 0<w<1.
\end{aligned} \tag{3.6}$$

与(2.1)式对照不难发现, (3.4)式或(3.5)式是(2.1)式在$k=0$时的特例, (3.6)式是(2.1)式在$k=1$时的特例. 一般地, (2.1)式可视作均匀分布的第$(k+1)$个次序统计量$y_{k+1}=F(x_{(k+1)},\theta)$的分布对于样本容量$n$服从参数为$\lambda$的截0至$k$的截尾泊松分布$\mathrm{P}_{0-k}(n)$的混合. 事实上, $y_{k+1}=F(x_{(k+1)},\theta)$服从参数为$(k+1,n-k)$的贝塔分布$\mathrm{Be}(k+1,n-k)$, 其密度函数为

$$h(y|n)=\frac{n!}{k!(n-k-1)!}y^k(1-y)^{n-k-1},\qquad 0<y<1, \tag{3.7}$$

《应用概率统计》版权所用

因此y_k的边际分布为

$$\begin{aligned} h_k(y) &= \frac{1}{\left[1-\sum\limits_{i=0}^{k}\frac{\lambda^i}{i!}e^{-\lambda}\right]}\sum_{n=k+1}^{\infty}\frac{n!}{k!(n-k-1)!}y^k(1-y)^{n-k-1}\frac{\lambda^n}{n!}e^{-\lambda} \\ &= \frac{\lambda}{\left[1-\sum\limits_{i=0}^{k}\frac{\lambda^i}{i!}e^{-\lambda}\right]}\sum_{n=k+1}^{\infty}\frac{(\lambda y)^k}{k!}\frac{[\lambda(1-y)]^{n-k-1}}{(n-k-1)!}e^{-\lambda(1-y)}e^{-\lambda y} \\ &= \frac{\lambda}{f_k(\lambda)}\frac{(\lambda y)^k}{k!}e^{-\lambda y}, \end{aligned} \tag{3.8}$$

此即(2.1)式.

§4. 应用举例

先考虑样本容量n给定的情形. 由上一节的讨论知, 对于未知连续型分布$F(x,\theta)$, 以$X_{(1)}$和$X_{(n)}$为分布的(τ,γ)上、下容许限与样本容量有关系[2, 3]

$$\mathsf{P}\{y_n = F(x_{(n)},\theta) \geq \tau\} = \int_{\tau}^{1} n y_n^{n-1}\mathrm{d}y_n = 1-\tau^n \geq \gamma, \tag{4.1}$$

$$\mathsf{P}\{y_1 = F(x_{(1)},\theta) \leq 1-\tau\} = \int_{0}^{1-\tau} n(1-y_1)^{n-1}\mathrm{d}y_1 = 1-\tau^n \geq \gamma, \tag{4.2}$$

即有

$$n \geq \ln(1-\gamma)/\ln\tau. \tag{4.3}$$

以(y_1,y_n)为$F(x,\theta)$的(τ,γ)容许区间与样本容量有关系

$$\begin{aligned} \mathsf{P}\{w=(y_n-y_1)\geq\tau\} &= \int_{\tau}^{1} n(n-1)w^{n-2}(1-w)\mathrm{d}w \\ &= 1-[n\tau^{n-1}-(n-1)\tau^n] \geq \gamma, \end{aligned} \tag{4.4}$$

或

$$n\tau^{n-1}-(n-1)\tau^n \leq 1-\gamma. \tag{4.5}$$

可见这些容许限或容许区间都与一定样本容量n有关.

再考虑n服从参数为λ的泊松分布的情形. 结合(2.2), 与混合贝塔分布(3.4)、(3.5)、(3.6)式对应, 以$x_{(1)}$为容许下限、τ为容许概率的容许度γ为

$$\mathsf{P}(y_1 \leq 1-\tau) = \int_{0}^{1-\tau} h(y_1)\mathrm{d}y_1 = G_0(1-\tau) = \frac{1-e^{-\lambda(1-\tau)}}{1-e^{-\lambda}} = \gamma. \tag{4.6}$$

把(3.4)中y_1换成$(1-y_n)$, 得以$x_{(n)}$为容许上限、τ为容许概率的容许度γ为

$$\mathsf{P}(y_n \geq \tau) = \int_{\tau}^{1} h(y_n)\mathrm{d}y_n = \frac{1-e^{-\lambda(1-\tau)}}{1-e^{-\lambda}} = \gamma. \tag{4.7}$$

《应用概率统计》版权所用

以$(x_{(1)},x_{(n)})$为容许区间、τ为容许概率的容许度γ为

$$\mathsf{P}(w\geq\tau)=\int_{\tau}^{1}h(w)\mathrm{d}w=G_1(1-\tau)=\frac{1-[1+\lambda(1-\tau)e^{-\lambda(1-\tau)}]}{1-(1+\lambda)e^{-\lambda}}=\gamma. \tag{4.8}$$

(4.6)、(4.7)或(4.8)式表明，以观测样本的最小值$x_{(1)}$为容许下限及以最大值$x_{(n)}$为容许上限，以及以$(x_{(1)},x_{(n)})$为容许区间的容许概率τ与容许度γ，均和参数λ有关．但参数λ与(4.3)与(4.5)中样本量参数n不同，不能同时根据容许概率τ与容许度γ的要求确定，而只能根据已知λ由容许概率τ确定容许度γ或由容许度γ确定容许概率τ．根据已知λ，按(4.6)、(4.7)或(4.8)式，容许概率τ越大，容许度γ越小；反之，容许概率τ越小，容许度γ越大．要同时达到容许概率τ与容许度γ即达到(τ,γ)要求，可根据实际对λ的推算按泊松分布可加性通过变化样本单位增大λ值实现，如并窝或不分窝观测同一家畜幼崽数中的$x_{(1)}$与$x_{(n)}$值．

如以容许限为例，若每窝家畜平均幼崽数估计为4.82，按(4.6)与(4.7)式，一窝幼崽有80%(容许概率τ)体重超过最小重量(容许下限$x_{(1)}$)或不超过最大重量(容许上限$x_{(n)}$)的概率(容许度)为62%；若平均幼崽数估计为$\lambda=9.64$，则80%体重超过$x_{(1)}$或不超过$x_{(n)}$的概率$\gamma=85\%$；若$\lambda=15$，则80%体重超过$x_{(1)}$或不超过$x_{(n)}$的概率$\gamma=95\%$；若$\lambda=20$，则80%体重超过$x_{(1)}$或不超过$x_{(n)}$的概率$\gamma=99\%$．与上述λ值对应的容许度$\gamma=80\%$的容许概率分别为$\tau=0.67, 0.83, 0.89, 0.92$．归纳前后结果即如表2所示．

表2　$k=0,1$时$\mathrm{BP}(k,\lambda)$的分布及数字特征

$\tau=80\%$					$\gamma=80\%$				
γ	0.62	0.85	0.95	0.99	τ	0.67	0.83	0.89	0.92
λ	4.82	9.64	15.00	20.00	λ	4.82	9.64	15.00	20.00

参 考 文 献

[1] 茆诗松，王静龙，濮晓龙，高等数理化统计，高等教育出版社，斯普林格出版社，1998.

[2] 茆诗松主编，统计手册，科学出版社，2003.

[3] Wilks, S., *Mathematical Statistics*, John Wiley & Sons, Inc. New York, 1961.

[4] 胡平，索赔次数基于两种不同分布的应用研究，经济师，**8**(2007), 35.

[5] Ghitany, M.E. and Al-Awadhi, S.A., A unified approach to some mixed Poisson distributions, *Tamsui Oxford Journal of Mathematical Sciences*, **17(2)**(2001), 147–161.

《应用概率统计》版权所用