应用概率统计　第二十五卷
第四期　2009年8月

Chinese Journal of Applied Probability
and Statistics Vol.25 No.4 Aug. 2009

广义多元Beta分布 *

陆志峰[1,2]　　王　娟[2,3]

([1]南通大学理学院, 南通, 226007; [2]东南大学数学系, 南京, 210096)

([3]江苏经贸职业技术学院, 南京, 211168)

摘　要

本文把广义Beta分布(Eugene (2001))推广到了多元的情形, 研究了多元Beta分布的矩母函数, 以及广义多元Beta分布的边际分布、条件分布及回归函数. 给出了他们在次序统计量中的应用.

关键词: Beta分布, 多元Beta分布, 广义多元Beta分布, 次序统计量.

学科分类号: O212.4.

§1. 引　　言

在实际问题中, 有很多问题相当复杂, 不能简单地用一些如正态分布、二项分布、负二项分布、Poisson分布、对数正态分布、Weibull分布、指数分布、Γ分布、极值分布等我们所熟知的概率分布来建模. 因此, 我们就有必要发展一些含有更多参数的广义的概率分布来处理那些复杂的情况.

Eugene (2001)[4]引进了广义Beta分布, Eugene, Lee and Famoye (2002)[5], Famoye, Lee and Eugene (2004)[6], Gupta and Nadarajah (2004)[3]等讨论了Beta-normal分布及其应用.

Eugene (2001)[4]定义了如下的广义Beta分布:

设$F(x)$是某随机变量的分布函数, 其概率函数为$f(x)$, α和β是给定的正实数, 则称随机变量X服从广义Beta分布, 若其分布函数为

$$G_F(x;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^{F(x)}t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t, \tag{1.1}$$

或其密度函数为

$$g_F(x;\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}f(x)F(x)^{\alpha-1}(1-F(x))^{\beta-1}, \tag{1.2}$$

记作$X\sim \mathrm{GB}(\alpha,\beta,F)$或$X\sim \mathrm{GB}(x,\alpha,\beta,F)$.

*应用型本科院校“十一五”国家课题(FIB070335-A2-02)、南通大学自然科学基金项目(06Z014)资助.
本文2006年2月14日收到, 2006年10月31日收到修改稿.

广义Beta分布某些特例早已为我们大家所熟知, 如McDonald (1984)[9], Parker (1999)[13]等的广义I型Beta分布与广义II型Beta分布, Prentice (1976)[12]等讨论的Logist VI型分布以及更为熟知的次序统计量的分布.

考虑分布函数

$$F(x)=\begin{cases}0, & x<0;\\ \left(\dfrac{x}{\Upsilon}\right)^b, & 0\le x<\Upsilon;\\ 1, & x\ge\Upsilon,\end{cases}\tag{1.3}$$

其中$b>0$.

$$G(y)=\begin{cases}0, & y<0;\\ \dfrac{(y/\Upsilon)^b}{1+(y/\Upsilon)^b}, & y\ge 0,\end{cases}\tag{1.4}$$

其中$\Upsilon>0$, $b>0$.

代入(1.2)得

$$h_F(x,\alpha,\beta)=\begin{cases}\dfrac{\Gamma(\alpha+\beta)bx^{b\alpha-1}(1-(x/\Upsilon)^b)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)\Upsilon^{\alpha b}}, & 0\le x\le\Upsilon;\\ 0, & \text{其他}.\end{cases}\tag{1.5}$$

$$h_G(y,\alpha,\beta)=\begin{cases}\dfrac{by^{b\alpha-1}}{\Upsilon^{b\alpha}B(\alpha,\beta)(1+(y/\Upsilon)^b)^{\alpha+\beta}}, & y\ge 0;\\ 0, & y<0.\end{cases}\tag{1.6}$$

(1.5)式与(1.6)式就是McDonald (1984)[9]在研究收入分布时引进的广义I型Beta分布与广义II型Beta分布. 1999年Parker在文[13]中用(1.6)式研究工人收入分布. 在(1.6)式中令$\Upsilon=1$, $b=1$, 即得到所谓的Fisher Z分布族(茆诗松, 王静龙, 濮晓龙 (1998)[1] p.11), 在(1.6)式中令$\Upsilon=\sqrt[2]{n}$, $b=2$, $\alpha=1/2$, $\beta=n/2$, 即得到所谓的自由度为n的t分布(茆诗松, 王静龙, 濮晓龙 (1998)[1] p.13).

考虑分布函数

$$E(z)=\frac{1}{1+\exp[-(z-\mu)/\theta]},\qquad \theta>0.\tag{1.7}$$

代入(1.2)得

$$h_E(z,\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\theta\Gamma(\alpha)\Gamma(\beta)}\Big\{1+\exp\Big(-\frac{z-\mu}{\theta}\Big)\Big\}^{-\alpha-\beta}\exp\Big(-\beta\frac{z-\mu}{\theta}\Big).\tag{1.8}$$

(1.8)式就是所谓的Logist VI型分布, 1976年Prentice在文[12]中建议, 对二元反映数据建模时, 用Logist VI型分布代替一般的Logist分布; 1980年Kalbfleish和Prentice在文[7]中把Logist VI型分布应用于生存分析; 1991年McDonald在文[10]中讨论了Logist VI型分布; 2004年Nadarajah在文[11]中讨论了Logist VI型分布的Fisher信息阵.

当$\alpha = 1\,(p,1,a)$; $\beta = 1\,(1,p,a)$时就是所谓的标准Logist分布(Logist I型分布, Logist II型分布, Logist III型分布).

设$X_1,\cdots,X_n$为i.i.d.样本, 且$X_1 \sim f(x)$为绝对连续型分布, 因此有$\mathsf{P}(X_{(i)} = X_{(j)}) = 0$, $i \neq j$; $\mathsf{P}(X_{(1)} < X_{(2)} < \cdots < X_{(n)}) = 1$.

记$Y = (Y_1,\cdots,Y_n)^T = (X_{(1)},\cdots,X_{(n)})^T$, $Y_i = X_{(i)} \sim f_{(i)}(y)$, $Y \sim f(y_1,y_2,\cdots,y_n)$.

$Y_i = X_{(i)}$的分布密度为(见[8]第72页)

$$f_{(i)}(y) = \frac{n!}{(i-1)!(n-i)!} f(y)[F(y)]^{i-1}[1-F(y)]^{n-i}. \tag{1.9}$$

即有$Y_i \sim \mathrm{GB}(i, n-i+1, F(x))$, $i = 1,2,\cdots,m$.

在信息社会中, 有大量的多元数据甚至是矩阵数据, 如千车故障数据(朱道元(2005)[2]), 对每一行都是可以用Weibull分布等来近似拟合, 但由于行与行之间并不独立, 应用已知分布很难描述; 又如, 对大多数慢性病, 对于不同阶段, 其生存能力是不一样的, 由于过程是不可逆. 基于此因素, 有必要发展一种多元的而且包含次序的分布. 一种自然的想法就是推广次序统计量的联合分布.

对应于Beta函数有所谓的Beta分布, 同样, 对应于多元Beta函数也有所谓的多元Beta分布, 本文的第二节回顾了多元Beta分布, 并初步研究了其性质. 本文的第三节定义了广义多元Beta分布, 并研究了其简单性质.

§2. 多元Beta分布

定义 2.1 设$\alpha = (\alpha_0,\cdots,\alpha_m)$, α_i为正实数, $i = 0,1,\cdots,m$, 随机向量$X = (X_1, X_2, \cdots, X_m)$服从多元Beta分布, 若其密度函数$f(x)$为

$$f(x;\alpha) = \frac{\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i\Big)}{\prod\limits_{i=0}^{m}\Gamma(\alpha_i)} \prod_{i=1}^{m+1}(x_i - x_{i-1})^{\alpha_{i-1}-1} I\{A\}, \tag{2.1}$$

其中

$$A = \{x | 0 \equiv x_0 \leq x_1 \leq \cdots \leq x_{m+1} \equiv 1\}. \tag{2.2}$$

记作$X \sim \mathrm{MB}(\alpha)$或$X \sim \mathrm{MB}(x,\alpha)$.

例 1 设α_i, $i = 0,1,\cdots,m$取正整数, 记$n_j = \sum\limits_{i=0}^{j}\alpha_i$, $j = 0,1,\cdots,m$, 则$\mathrm{MB}(\alpha)$就表示均匀分布$U[0,1]$的次序统计量$(U_{n_0:n_m}, U_{n_1:n_m}, \cdots, U_{n_m:n_m})$的分布. 其中$X_{k:n}$表示样本$(X_1,\cdots,X_n)$的第$k$个次序统计量.

引理 2.1 $f(x,\alpha)$如(2.1)式)所定义, A如(2.2)式)所定义, $1\le j\le m$, 则

$$\begin{aligned}\int_{\infty}^{-\infty} f(x)\mathrm{d}x_j &= \frac{\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i\Big)}{\prod\limits_{i=0}^{m}\Gamma(\alpha_i)}\int_{x_{j-1}}^{x_{j+1}}\prod_{i=1}^{m+1}(x_i-x_{i-1})^{\alpha_{i-1}-1}I\{A\}\mathrm{d}x_j\\ &= \frac{\Gamma(\alpha_{j-1})\Gamma(\alpha_j)\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i\Big)}{\Gamma(\alpha_{j-1}+\alpha_j)\prod\limits_{i=0}^{m}\Gamma(\alpha_i)}\Big(\prod_{i=1}^{j-1}(x_i-x_{i-1})^{\alpha_{i-1}-1}\Big)\\ &\quad\cdot(x_{j+1}-x_{j-1})^{\alpha_{j-1}+\alpha_j-1}\Big(\prod_{i=j+2}^{m+1}(x_i-x_{i-1})^{\alpha_{i-1}-1}\Big)I\{\overline{A}\},\end{aligned}\qquad(2.3)$$

其中$\overline{A}=\{x|0\equiv x_0\le x_1\le\cdots\le x_{j-1}\le x_{j+1}\le\cdots\le x_{m+1}\equiv 1\}$.

证明: 只需注意下面的β积分:

$$\int_{x_{j-1}}^{x_{j+1}}(x_j-x_{j-1})^{\alpha_{j-1}-1}(x_{j+1}-x_j)^{\alpha_j-1}\mathrm{d}x_j=\frac{\Gamma(\alpha_{j-1})\Gamma(\alpha_j)}{\Gamma(\alpha_{j-1}+\alpha_j)}(x_{j+1}-x_{j-1})^{\alpha_{j-1}+\alpha_j-1}.$$

□

对(2.1)式进行多元Mellin变换[14], 即得到所谓的母函数: $M_X(s)=\mathsf{E}X^s=\mathsf{E}\Big(\prod\limits_{i=1}^{m}X_i^{s_i}\Big)$, 其中$s_i$, $i=1,2,\cdots m$是复数, 我们仅计算当$m=2$情形, 一般情况同理可得.

$$\begin{aligned}M_X(s) &= \mathsf{E}(X_1^{s_1}X_2^{s_2})\\ &= \frac{\Gamma(\alpha_0+\alpha_1+\alpha_2)}{\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_0^1\mathrm{d}x_2\int_0^{x_2}x_2^{s_2}x_1^{s_1+\alpha_0-1}(x_2-x_1)^{\alpha_1-1}(1-x_2)^{\alpha_2-1}\mathrm{d}x_1\\ &= \frac{\Gamma(\alpha_0+\alpha_1+\alpha_2)}{\Gamma(\alpha_0)\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_0^1 x_2^{\alpha_0+\alpha_1+s_1+s_2-1}(1-x_2)^{\alpha_2-1}\mathrm{d}x_2\\ &\quad\cdot\int_0^1 t^{s_1+\alpha_0-1}(1-t)^{\alpha_1-1}\mathrm{d}t,\qquad t\triangleq\frac{x_1}{x_2}\\ &= \frac{\Gamma(\alpha_0+\alpha_1+\alpha_2)\Gamma(\alpha_0+s_1)\Gamma(\alpha_0+\alpha_1+s_1+s_2)}{\Gamma(\alpha_0)\Gamma(\alpha_0+\alpha_1+s_1)\Gamma(\alpha_0+\alpha_1+\alpha_2+s_1+s_2)}.\end{aligned}$$

一般地我们有:

定理 2.1 设$X\sim\mathrm{MB}(\alpha)$, 则其母函数为

$$M_X(s)=\frac{\Gamma\Big(\sum\limits_{j=0}^{m}\alpha_i\Big)}{\Gamma(\alpha_0)}\prod_{i=0}^{m-1}\frac{\Gamma\Big(\sum\limits_{j=0}^{i}(\alpha_j+s_{j+1})\Big)}{\Gamma\Big(\alpha_0+\sum\limits_{j=0}^{i}(\alpha_{j+1}+s_{j+1})\Big)}.\qquad(2.4)$$

当$m=1$时, 有

推论 2.1 设$X\sim \mathrm{B}(\alpha,\beta)$, 则其母函数为

$$M_X(s)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\frac{\Gamma(\alpha+s)}{\Gamma(\alpha+\beta+s)}. \tag{2.5}$$

推论 2.2 设$X\sim \mathrm{MB}(\alpha)$, $0<j_1<j_2<\cdots<j_k\le m$, $\widetilde{x}=(x_{j_1},x_{j_2},\cdots,x_{j_k})$, $\widetilde{X}=(X_{j_1},X_{j_2},\cdots,X_{j_k})$, $\widetilde{\alpha}=\Big(\sum\limits_{i=0}^{j_1-1}\alpha_i,\sum\limits_{i=j_1}^{j_2-1}\alpha_i,\cdots,\sum\limits_{i=j_k}^{m}\alpha_i\Big)$, 则$\widetilde{X}\sim \mathrm{MB}(\widetilde{x},\widetilde{\alpha})$.

推论 2.3 设$X\sim \mathrm{MB}(\alpha)$, $0<j\le m$, 则$X_j\sim \mathrm{B}\Big(\sum\limits_{i=0}^{j-1}\alpha_i,\sum\limits_{i=j}^{m}\alpha_i\Big)$.

推论 2.4 设$X\sim \mathrm{MB}(\alpha)$, $0<j\le m$, 则

$$\mathsf{E}(X_j)=\frac{\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i\Big)}{\Gamma\Big(\sum\limits_{i=0}^{j-1}\alpha_i\Big)}\frac{\Gamma\Big(\sum\limits_{i=0}^{j-1}\alpha_i+1\Big)}{\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i+1\Big)}=\frac{\sum\limits_{i=0}^{j-1}\alpha_i}{\sum\limits_{i=0}^{m}\alpha_i}.$$

§3. 广义多元Beta分布

定义 3.1 设$\alpha=(\alpha_0,\cdots,\alpha_m)$, α_i为正实数, $i=0,1,\cdots,m$, $\mathcal{F}=(F_1(x_1),\cdots,F_m(x_m))$, $F_i(x_i)$, $i=1,2,\cdots,m$是m个分布函数, 则称随机向量X服从广义多元Beta分布, 若其密度函数$g(x)$为

$$g_{\mathcal{F}}(x;\alpha)=\frac{\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i\Big)}{\prod\limits_{i=0}^{m}\Gamma(\alpha_i)}\prod_{i=1}^{m}f_i(x_i)\prod_{i=1}^{m+1}(F_i(x_i)-F_{i-1}(x_{i-1}))^{\alpha_{i-1}-1}I\{A\}, \tag{3.1}$$

其中$A=\{x|0\equiv F_0(x_0)\le F_1(x_1)\le\cdots\le F_{m+1}(x_{m+1})\equiv 1\}$. 记作$X\sim \mathrm{GMB}(\alpha,\mathcal{F})$或$X\sim \mathrm{GMB}(x,\alpha,\mathcal{F})$.

定理 3.1 设$X\sim \mathrm{GMB}(\alpha,\mathcal{F})$, $0<j\le m$, 则$X_j\sim \mathrm{GB}\Big(\sum\limits_{i=0}^{j-1}\alpha_i,\sum\limits_{i=j}^{m}\alpha_i,F_j(x_j)\Big)$.

证明:

$$\begin{aligned}f(x_j)&=\frac{\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i\Big)}{\prod\limits_{i=0}^{m}\Gamma(\alpha_i)}\int_D\prod_{i=1}^{m}f_i(x_i)\\&\quad\cdot\prod_{i=1}^{m+1}(F_i(x_i)-F_{i-1}(x_{i-1}))^{\alpha_{i-1}-1}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_{j-1}\mathrm{d}x_{j+1}\cdots\mathrm{d}x_m\end{aligned}$$

$$
\begin{aligned}
&= \frac{\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i\Big)}{\prod\limits_{i=0}^{m}\Gamma(\alpha_i)} f_j(x_j) F_j^{\sum\limits_{i=0}^{j-1}\alpha_i-1}(x_j)(1-F_j(x_j))^{\sum\limits_{i=j}^{m}\alpha_i-1}\\
&\quad\cdot\int_{D_1} s_1^{\alpha_0-1}(s_2-s_1)^{\alpha_1-1}\cdots(1-s_{j-1})^{\alpha_{j-1}-1}\mathrm{d}s_1\cdots\mathrm{d}s_{j-1}\\
&\quad\cdot\int_{D_2} t_{j+1}^{\alpha_j-1}(t_{j+2}-t_{j+1})^{\alpha_{j+1}-1}\cdots(1-t_m)^{\alpha_m-1}\mathrm{d}t_{j+1}\cdots\mathrm{d}t_m\\
&= \frac{\Gamma\Big(\sum\limits_{i=0}^{m}\alpha_i\Big)}{\Gamma\Big(\sum\limits_{i=0}^{j-1}\alpha_i\Big)\Gamma\Big(\sum\limits_{i=j}^{m}\alpha_i\Big)} f_j(x_j) F_j^{\sum\limits_{i=0}^{j-1}\alpha_i-1}(x_j)(1-F_j(x_j))^{\sum\limits_{i=j}^{m}\alpha_i-1},
\end{aligned}
$$

其中

$$
\begin{aligned}
&D=\{x|0\equiv F_0(x_0)\le F_1(x_1)\le\cdots\le F_{m+1}(x_{m+1})\equiv 1\};\\
&D_1=\{x|0<s_1\le s_2\le\cdots\le s_{j-1}\le 1\};\\
&D_2=\{x|0<t_{j+1}\le t_{j+2}\le\cdots\le t_m\le 1\};\\
&s_i=\frac{F_i(x_i)}{F_j(x_j)},\qquad i=1,2,\cdots,j-1;\\
&t_i=\frac{F_i(x_i)-F_j(x_j)}{1-F_j(x_j)},\qquad i=j+1,\cdots,m.
\end{aligned}
$$

所以定理证毕. □

更一般地, 有

定理 3.2 设$X\sim \mathrm{GMB}(\alpha,\mathcal{F})$, $0<j_1<j_2<\cdots<j_k\le m$, $\widetilde{\mathcal{F}}=(F_{j_1},F_{j_2},\cdots,F_{j_k})$, $\widetilde{x}=(x_{j_1},x_{j_2},\cdots,x_{j_k})$, $\widetilde{X}=(X_{j_1},X_{j_2},\cdots,X_{j_k})$, $\widetilde{\alpha}=\Big(\sum\limits_{i=0}^{j_1-1}\alpha_i,\sum\limits_{i=j_1}^{j_2-1}\alpha_i,\cdots,\sum\limits_{i=j_k}^{m}\alpha_i\Big)$, 则

$$\widetilde{X}\sim \mathrm{GMB}(\widetilde{x},\widetilde{\alpha},\widetilde{\mathcal{F}}).$$

定理 3.3 设$(X,Y)\sim \mathrm{GMB}(\alpha,\beta,\gamma,F_1,F_2)$, $\widetilde{F}_1(x)=(F_1(x)/F_2(Y))\cdot I\{F_1(x)\le F_2(Y)\}$, $\widetilde{F}_2(y)=[(F_2(y)-F_1(X))/(1-F_1(X))]\cdot I\{F_1(X)\le F_2(y)\}$, 则

$$X|Y\sim \mathrm{GB}(\alpha,\beta,\widetilde{F}_1(x)),\tag{3.2}$$

$$Y|X\sim \mathrm{GB}(\beta,\gamma,\widetilde{F}_2(y)).\tag{3.3}$$

证明: 由于$(X,Y)\sim \mathrm{GMB}(\alpha,\beta,\gamma,F_1,F_2)$, 则由定理3.1知, $X\sim \mathrm{GB}(\alpha,\beta+\gamma,F_1)$, $Y\sim \mathrm{GB}(\alpha+\beta,\gamma,F_2)$, 进而

$$f_{X|Y=y}(x)=\frac{f(x,y)}{f_Y(y)}=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{f_1(x)}{F_2(y)}\Big(\frac{F_1(x)}{F_2(y)}\Big)^{\alpha-1}\Big(1-\frac{F_1(x)}{F_2(y)}\Big)^{\beta-1}\tag{3.4}$$

与

$$f_{Y|X=x}(y)=\frac{f(x,y)}{f_X(x)}=\frac{\Gamma(\beta+\gamma)}{\Gamma(\beta)\Gamma(\gamma)}\frac{f_2(y)}{1-F_1(x)}\Big(\frac{F_2(y)-F_1(x)}{1-F_1(x)}\Big)^{\beta-1}\Big(\frac{1-F_2(y)}{1-F_1(x)}\Big)^{\gamma-1}. \tag{3.5}$$

所以定理证毕. □

定理 3.4 设$(X,Y,Z)\sim\mathrm{GMB}(\alpha,\beta,\gamma,\delta,F_1,F_2,F_3)$, $\widetilde{F}_1(x)=(F_1(x)/F_2(Y))\cdot I\{F_1(x)\le F_2(Y)\}$, $\widetilde{F}_3(z)=[(F_3(z)-F_2(Y))/(1-F_2(Y))]\cdot I\{F_3(z)\ge F_2(y)\}$, 则

$$(X,Z)|Y\overset{\mathrm{d}}{=}(U,V), \tag{3.6}$$

其中U与V独立,

$$U=X|Y\sim\mathrm{GB}(\alpha,\beta,\widetilde{F}_1(x)), \tag{3.7}$$

$$V=Z|Y\sim\mathrm{GB}(\gamma,\delta,\widetilde{F}_3(z)). \tag{3.8}$$

证明: 由于$(X,Y,Z)\sim\mathrm{GMB}(\alpha,\beta,\gamma,\delta,F_1,F_2,F_3)$, 则由定理3.1知, $Y\sim\mathrm{GB}(\alpha+\beta,\gamma+\delta,F_2)$, 进而

$$\begin{aligned}f_{(X,Z)|Y=y}(x,z)&=\frac{f(x,y,z)}{f_Y(y)}\\&=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{f_1(x)}{F_2(y)}\Big(\frac{F_1(x)}{F_2(y)}\Big)^{\alpha-1}\Big(1-\frac{F_1(x)}{F_2(y)}\Big)^{\beta-1}\\&\quad\cdot\frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)}\frac{f_3(z)}{1-F_2(y)}\Big(\frac{F_3(z)-F_2(y)}{1-F_2(y)}\Big)^{\gamma-1}\Big(1-\frac{F_3(z)-F_2(y)}{1-F_2(y)}\Big)^{\delta-1}.\end{aligned}$$

所以定理证毕. □

定理3.4可以推广到一般情形:

定理 3.5 设m维随机向量$X\sim\mathrm{GMB}(\alpha,\mathcal{F})$, 给定$0=j_0<j_1<j_2<\cdots<j_k<j_{k+1}=m+1$, 记$\widetilde{X}_i=(X_{j_i+1},X_{j_i+2},\cdots,X_{j_{i+1}-1})$, $\widetilde{x}_i=(x_{j_i+1},x_{j_i+2},\cdots,x_{j_{i+1}-1})$, $\widetilde{\alpha}_i=(\alpha_{j_i+1},\alpha_{j_i+2},\cdots,\alpha_{j_{i+1}-1})$, $G_{j_i+l}(x_{j_i+l})=[F_{j_i+l}(x_{j_i+l})-F_{j_i}(X_{j_i})]/[F_{j_{i+1}}(X_{j_{i+1}})-F_{j_i}(X_{j_i})]$ (其中$l=1,2,\cdots,j_{i+1}-j_i-1$), $\mathcal{F}_i=(G_{j_i+1},G_{j_i+2},\cdots,G_{j_{i+1}-1})$, $U_i=Y_i|(X_{j_i},X_{j_{i+1}})$, 其中$i=0,1,\cdots,k$.

当$j_{i+1}=j_i+1$时, $\widetilde{X}_i$, U_i与$\mathcal{F}_i$没有定义, 我们约定

$$\mathsf{P}(\widetilde{X}_i=1)=1,\qquad \mathsf{P}(U_i=1)=1, \tag{3.9}$$

则有

$$(\widetilde{X}_0,\widetilde{X}_1,\cdots,\widetilde{X}_k)|(X_{j_1},X_{j_2},\cdots,X_{j_k})\overset{\mathrm{d}}{=}(U_0,U_1,\cdots,U_k), \tag{3.10}$$

其中$U_0,U_1,\cdots,U_k$相互独立, 且$U_i\sim\mathrm{GMB}(\widetilde{\alpha}_i,\mathcal{F}_i(\widetilde{x}_i))$, $i=0,1,\cdots,k$.

定理 3.6 设$X \sim \mathrm{GMB}(\alpha,\mathcal{F})$, $0<i<i+k\le m$, 记$\widetilde{\alpha}=\sum\limits_{j=0}^{i-1}\alpha_j$, $\widetilde{\beta}=\sum\limits_{j=i}^{i+k-1}\alpha_j$, $\widetilde{\gamma}=\sum\limits_{i+k-1}^{m}\alpha_j$, 分布族$\mathcal{F}$有相同的连续分布$F(x)$ (即$F_i(x)=F_{i+k}(x)$且密度函数为$f(x)$), 设$f(x)$有连续的支撑集$\mathcal{D}$, 则当$F(x)$是下面三种情形之一时:

$$F^{11}(x)=\begin{cases}0, & x\le 0;\\ x^{\theta}, & 0<x<1; \qquad (\theta>0),\\ 1, & x\ge 1,\end{cases} \tag{3.11}$$

$$F^{12}(x)=\begin{cases}\mathrm{e}^{x}, & x<0;\\ 1, & x\ge 1,\end{cases} \tag{3.12}$$

$$F^{13}(x)=\begin{cases}(-x)^{\delta}, & x<-1;\\ 1, & x\ge -1,\end{cases} \qquad \Big(\delta<-\frac{1}{\widetilde{\alpha}}\Big). \tag{3.13}$$

有

$$\mathsf{E}\{X_i|X_{i+k}=x\}=ax+b \qquad \text{a.s.} \qquad x\in\mathcal{D}, \tag{3.14}$$

其中, 相应于$F^{11}(x)$有$0<a<1$, $b=0$; 相应于$F^{12}(x)$有$a=1$, $b=\psi(\widetilde{\alpha})-\psi(\widetilde{\alpha}+\widetilde{\beta})$; 相应于$F^{13}(x)$有$a>1$, $b=0$. 式中$\psi(\cdot)$是双Gamma函数.

证明: 由定理3.2与定理3.3可知: $(X_i,X_{i+k})\sim \mathrm{GMB}(\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma},F(x),F(x))$, 所以

$$X_i|(X_{i+k}=x)\sim \mathrm{GB}\Big(\widetilde{\alpha},\widetilde{\beta},\frac{F_i(z)}{F_{i+k}(x)}\Big).$$

当$F(x)=F^{11}(x)$时, 有

$$\begin{aligned}\mathsf{E}(X_i|(X_{i+k}=x)) &= \frac{\Gamma(\widetilde{\alpha}+\widetilde{\beta})}{\Gamma(\widetilde{\alpha})\Gamma(\widetilde{\beta})}\int_0^x \frac{\theta s}{x}\Big(\frac{s}{x}\Big)^{\theta\widetilde{\alpha}-1}\Big(1-\Big(\frac{s}{x}\Big)^{\theta}\Big)^{\widetilde{\beta}-1}\mathrm{d}s\\ &= x\frac{\Gamma(\widetilde{\alpha}+\widetilde{\beta})}{\Gamma(\widetilde{\alpha})\Gamma(\widetilde{\beta})}\int_0^1 t^{\widetilde{\alpha}-1}(1-t)^{\widetilde{\beta}-1}t^{1/\theta}\mathrm{d}t\\ &= x\frac{\Gamma(\widetilde{\alpha}+\widetilde{\beta})\Gamma(\widetilde{\alpha}+1/\theta)}{\Gamma(\widetilde{\alpha})\Gamma(\widetilde{\alpha}+\widetilde{\beta}+1/\theta)}.\end{aligned}$$

记$h(x)=[\Gamma(\widetilde{\alpha}+\widetilde{\beta})\Gamma(\widetilde{\alpha}+x)]/[\Gamma(\widetilde{\alpha})\Gamma(\widetilde{\alpha}+\widetilde{\beta}+x)]>0$, 则$\mathrm{d}\ln h(x)/\mathrm{d}x=\psi(\widetilde{\alpha}+x)-\psi(\widetilde{\alpha}+\widetilde{\beta}+x)$, 又由于$\psi(\cdot)$是单调递增函数, 所以$\mathrm{d}\ln h(x)/\mathrm{d}x<0$, 即$\ln h(x)$单调递减, 进而$h(x)$单调递减. 所以$a=h(1/\theta)<h(0)=1$, 所以$0<a<1$, $b=0$.

同理可证: 当$F(x)=F^{13}(x)$时,

$$\mathsf{E}(X_i|(X_{i+k}=x))=x\frac{\Gamma(\widetilde{\alpha}+\widetilde{\beta})\Gamma(\widetilde{\alpha}+1/\delta)}{\Gamma(\widetilde{\alpha})\Gamma(\widetilde{\alpha}+\widetilde{\beta}+1/\delta)}.$$

由于$\delta < -1/\widetilde{\alpha} < 0$, 所以, $a = h(1/\delta) > h(0) = 1$. 进而有$a > 1$, $b = 0$.

当$F(x) = F^{12}(x)$时,

$$\begin{aligned}\mathsf{E}(X_i|(X_{i+k}=x)) &= \frac{\Gamma(\widetilde{\alpha}+\widetilde{\beta})}{\Gamma(\widetilde{\alpha})\Gamma(\widetilde{\beta})}\int_{-\infty}^{x} s\Big(\frac{\mathrm{e}^s}{\mathrm{e}^x}\Big)^{\widetilde{\alpha}}\Big(1-\frac{\mathrm{e}^s}{\mathrm{e}^x}\Big)^{\widetilde{\beta}-1}\mathrm{d}s \\ &= x+\frac{\Gamma(\widetilde{\alpha}+\widetilde{\beta})}{\Gamma(\widetilde{\alpha})\Gamma(\widetilde{\beta})}\int_0^1 t^{\widetilde{\alpha}-1}(1-t)^{\widetilde{\beta}-1}\log t\mathrm{d}t \\ &= x+\psi(\widetilde{\alpha})-\psi(\widetilde{\alpha}+\widetilde{\beta}).\end{aligned}$$

此时, $a = 1$, $b = \psi(\widetilde{\alpha}) - \psi(\widetilde{\alpha}+\widetilde{\beta})$. 定理证毕. $\square$

定理 3.7 设$X \sim \mathrm{GMB}(\alpha, \mathcal{F})$, $0 < i < i+k \le m$, 记$\widetilde{\alpha} = \sum\limits_{j=0}^{i-1}\alpha_j$, $\widetilde{\beta} = \sum\limits_{j=i}^{i+k-1}\alpha_j$, $\widetilde{\gamma} = \sum\limits_{i+k-1}^{m}\alpha_j$, 分布族$\mathcal{F}$有相同的连续分布$F(x)$ (即$F_i(x) = F_{i+k}(x)$且密度函数为$f(x)$), 设$f(x)$有连续的支撑集$\mathcal{D}$, 则当$F(x)$是下面三种情形之一时:

$$F^{21}(x) = \begin{cases}0, & x \le -1; \\ 1-(-x)^{\theta}, & -1 < x < 0; \\ 1, & x \ge 0,\end{cases} \qquad (0 < \theta), \tag{3.15}$$

$$F^{22}(x) = \begin{cases}0, & x < 0; \\ 1-\mathrm{e}^{-x}, & x \ge 0,\end{cases} \tag{3.16}$$

$$F^{23}(x) = \begin{cases}0, & x < 1; \\ 1-x^{\delta}, & x \ge 1,\end{cases} \qquad \Big(\delta < -\frac{1}{\widetilde{\gamma}}\Big). \tag{3.17}$$

有

$$\mathsf{E}\{X_{i+k}|X_i = y\} = ay+b \qquad \text{a.s.} \qquad x \in \mathcal{D}, \tag{3.18}$$

其中, 相应于$F^{21}(x)$有$0 < a < 1$, $b = 0$; 相应于$F^{22}(x)$有$a = 1$, $b = \psi(\widetilde{\gamma}) - \psi(\widetilde{\gamma}+\widetilde{\beta})$; 相应于$F^{23}(x)$有$a > 1$, $b = 0$.

证明:

$$X_{i+k}|(X_i = y) \sim \mathrm{GB}\Big(\widetilde{\beta}, \widetilde{\gamma}, \frac{F(z)-F(y)}{1-F(y)}\Big).$$

当$F(x) = F^{21}(x)$时, 有

$$\begin{aligned}\mathsf{E}(X_{i+k}|(X_i=y)) &= \frac{\Gamma(\widetilde{\beta}+\widetilde{\gamma})}{\Gamma(\widetilde{\beta})\Gamma(\widetilde{\gamma})}\int_y^0 \frac{-\theta s}{y}\Big(\frac{s}{y}\Big)^{\theta\widetilde{\gamma}-1}\Big(1-\Big(\frac{s}{y}\Big)^{\theta}\Big)^{\widetilde{\beta}-1}\mathrm{d}s \\ &= y\frac{\Gamma(\widetilde{\beta}+\widetilde{\gamma})}{\Gamma(\widetilde{\beta})\Gamma(\widetilde{\gamma})}\int_0^1 t^{\widetilde{\gamma}-1}(1-t)^{\widetilde{\beta}-1}t^{1/\theta}\mathrm{d}t \\ &= y\frac{\Gamma(\widetilde{\beta}+\widetilde{\gamma})\Gamma(\widetilde{\gamma}+1/\theta)}{\Gamma(\widetilde{\gamma})\Gamma(\widetilde{\beta}+\widetilde{\gamma}+1/\theta)}.\end{aligned}$$

记$l(x) = [\Gamma(\widetilde{\beta}+\widetilde{\gamma})\Gamma(\widetilde{\gamma}+x)]/[\Gamma(\widetilde{\gamma})\Gamma(\widetilde{\beta}+\widetilde{\gamma}+x)] > 0$, 同前$h(x)$一样, $l(x)$单调递减. 所以$a = l(1/\theta) < h(0) = 1$, 所以$0 < a < 1$, $b = 0$.

同理可证: 当$F(x) = F^{23}(x)$时,

$$\mathsf{E}(X_{i+k}|(X_i = y)) = y\frac{\Gamma(\widetilde{\beta}+\widetilde{\gamma})\Gamma(\widetilde{\gamma}+1/\delta)}{\Gamma(\widetilde{\gamma})\Gamma(\widetilde{\beta}+\widetilde{\gamma}+1/\delta)}.$$

由于$\delta < -1/\widetilde{\gamma} < 0$, 所以, $a = l(1/\delta) > l(0) = 1$. 所以$a > 1$, $b = 0$.

设$F(x) = F^{22}(x)$, 则

$$\begin{aligned}\mathsf{E}(X_{i+k}|(X_i = y)) &= \frac{\Gamma(\widetilde{\beta}+\widetilde{\gamma})}{\Gamma(\widetilde{\beta})\Gamma(\widetilde{\gamma})}\int_y^{+\infty} s\Big(\frac{\mathrm{e}^y}{\mathrm{e}^s}\Big)^{\widetilde{\gamma}}\Big(1-\frac{\mathrm{e}^y}{\mathrm{e}^s}\Big)^{\widetilde{\beta}-1}\mathrm{d}s\\ &= y + \frac{\Gamma(\widetilde{\beta}+\widetilde{\gamma})}{\Gamma(\widetilde{\beta})\Gamma(\widetilde{\gamma})}\int_0^1 t^{\widetilde{\gamma}-1}(1-t)^{\widetilde{\beta}-1}\log t\mathrm{d}t\\ &= y + \psi(\widetilde{\gamma}) - \psi(\widetilde{\gamma}+\widetilde{\beta}).\end{aligned}$$

此时, $a = 1$, $b = \psi(\widetilde{\gamma}) - \psi(\widetilde{\gamma}+\widetilde{\beta})$. 定理证毕. □

例 2　次序统计量的分布

1. 在(3.1)式中, 令$\alpha_0 = i$, $\alpha_1 = j-i$, $\alpha_2 = n-j+1$ $(1 \le i < j \le n)$, $\mathcal{F} = (F(y), F(z))$, $F(\cdot)$是分布函数, 则有

$$f(y,z) = \frac{n!f(y)f(z)[F(y)]^{i-1}[F(z)-F(y)]^{j-i-1}[1-F(z)]^{n-j}I\{y<z\}}{(i-1)!(j-i-1)!(n-j)!}. \tag{3.19}$$

此式即为总体$F(\cdot)$的次序统计量$(X_{(i)}, X_{(j)}) = (Y, Z)$ $(i < j)$的联合分布(见Shao (1999)[8]第72页).

2. 在(3.1)式中, 令$\alpha_0 = \alpha_1 = \cdots = \alpha_{r-1} = 1$ $(1 \le r \le n)$, $\alpha_r = n-r+1$, $\mathcal{F} = (F(y_1), \cdots, F(y_r))$, $F(\cdot)$是分布函数, 则有

$$f(y_1,\cdots,y_r) = \frac{n!}{(n-r)!}f(y_1)\cdots f(y_r)[1-F(y_r)]^{n-r}I\{y_1<\cdots<y_r\}. \tag{3.20}$$

此式即为总体$F(\cdot)$中的前r个次序统计量$X_{(1)},\cdots,X_{(r)}$的联合分布.

特别地, $r = n$, 即前n个次序统计量$X_{(1)},\cdots,X_{(n)}$的联合分布为(见Shao (1999)[8]第72页)

$$f(y_1,\cdots,y_n) = n!f(y_1)\cdots f(y_n)I\{y_1<y_2<\cdots<y_n\}. \tag{3.21}$$

§4.　广义Beta分布族

由广义Beta分布的定义与广义多元Beta分布可知, 随着分布函数$F(x)$与参数的不同, 就得到不同的分布函数. 把相同类的分布函数所得到的广义Beta分布、广义多元Beta分布统称为一类广义Beta分布族. 如Beta-normal分布族([3–6])、Beta-Weibull分布族、Beta-指数分布族等等, 以上这些分布族的性质有待进一步研究.

参 考 文 献

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

[2] 朱道元, 首届全国部分高校研究生数学建模竞赛, 数学的实践与认识, **35(7)**(2005), 1-15.

[3] Gupta, A.K. and Nadarajah, S., On the moments of the Beta normal distribution, *Communications in Statistics: Theory and Methods*, **33(1)**(2004), 1–13.

[4] Eugene, N., *A Class of Generalized Normal Distribution: Properties, Estimation and Applications*, Central Michigan University, Mount Pleasant, Michigan, 2001.

[5] Eugene, N., Lee, C. and Famoye, F., Beta-normal distribution and its application, *Communications in Statistics: Theory and Methods*, **31(4)**(2002), 497–512.

[6] Famoye, F., Lee, C. and Eugene, N., Beta-normal distributions: bimodality properties and application, *Journal of Modern Applied Statistical Methods*, **3(1)**(2004), 85–103.

[7] Kalbfleish, J.D. and Prentice, R.L., *The Statistical Analysis of Failure Time Data*, John Wiley and Sons, New York, 1980.

[8] Shao, J., *Mathematical Statistics*, Springer-Verlag New York, Inc., 1999.

[9] McDonald, J.B., Some generalized functions for the size distribution of income, *Econometrics*, **52**(1984), 647–663.

[10] McDonald, J.B., Parametric models for partially adaptive estimation with skewed and leptokurtic residuals, *Economic Letters*, **37**(1991), 273–278.

[11] Nadarajah, S., Information Matrix For Logistic Distributions, *Mathematical and Computer Modelling*, **40**(2004), 953–958.

[12] Prentice, R.L., A generalization of the probit and logit methods for dose response curves, *Biometrics*, **32**(1976), 761–768.

[13] Parker, S.C., The generalised Beta as a model for the distribution of earnings, *Economics Letters*, **62**(1999), 197–200.

[14] Titchmarsh, E., *The Theory of Functions*, 2nd ed., Oxford University Press, Oxford, 1939.

Generalized Multivariate Beta Distribution

Lu Zhifeng[1,2]   Wang Juan[2,3]

([1]*Nantong University, Nantong, 226007;* [2]*Southeast University, Nanjin, 210096*)

([3]*Jiangsu Institute of Economic and Trade Technology, Nanjin, 211168*)

This paper extends the (generalized) Beta distribution to the (generalized) multivariate Beta distribution. We also study the moment generating function of multivariate Beta distribution, obtain the marginal distribution, conditional distribution and regression function of the generalized multivariate Beta distribution.

Keywords: Beta distribution, multivariate Beta distribution, generalized multivariate Beta distribution, order statistic.

AMS Subject Classification: 62H10, 60E99, 62G30.