应用概率统计　第十九卷
第一期　2003 年 2 月

Chinese Journal of Applied Probability
and Statistics Vol. 19 No. 1 Feb. 2003

# 二元极值分布的一个性质 *

史　道　济
(天津大学数学系；南开大学天津大学刘徽应用数学中心，天津，300072)

**摘　要**

本文考虑二元极值的相关结构，通过一个变量变换，使变换后的变量基本上是独立的，并给出了它们的随机表示．由此能非常容易地在计算机上产生二元极值分布伪随机向量，以及计算一类常用统计量的数字特征，这是研究某些统计量渐近分布的基础．

**关键词**：二元极值分布，Copula，相关结构，独立性，随机表示．
**学科分类号**：O212.4.

## §1. 引　　论

关于二元极值分布的讨论主要集中在极值的相关结构上．设 $G(x,y)$ 是一个二元极值分布，它可以表示为

$$G(x,y)=C\{F_1(x),F_2(y)\}, \tag{1.1}$$

其中 $F_1(x)$、$F_2(y)$ 是边缘分布，而

$$C(u,v)=\mathsf{P}\{F_1(X)\le u,F_2(Y)\le v\}=\exp\Big\{\log(uv)A\Big(\frac{\log u}{\log(uv)}\Big)\Big\},\qquad 0\le u,v\le 1 \tag{1.2}$$

是 copula，函数 $A$ 定义在 $[0,1]$ 上，满足

$$\max(t,1-t)\le A(t)\le 1,\qquad 0\le t\le 1,$$

称为相关函数．Copula 的概念最早由 [1] 提出的，后来 [2,3] 对此做了专门介绍．[4](第九章) 及 [5] 给出了二元极值分布中相关函数 $A$ 的参数形式的例子．

关于 $A$ 的非参数估计已有不少讨论，首先是由 [6] 提出的，以后 [7] 给出了它的渐近性质，[8] 讨论了 $A$ 的核密度估计，[9] 及 [10] 分别给出了一个半参数及非参数估计，最近 [11] 又提出了二个相关函数的新的估计方法．

设 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 是二元极值分布 (1.1) 的 $n$ 个观测，假定边缘分布 $F_1(x)$，$F_2(y)$ 已知，记 $(U_i,V_i)=(F_1(X_i),F_2(Y_i))$，$i=1,\cdots,n$，它们是 Copula(1.2) 的观测，考虑

$$Z_i=\log(U_i)/\log(U_iV_i),\qquad i=1,\cdots,n.$$

用 $Z_{(1)},\cdots,Z_{(n)}$ 表示它们的次序统计量，定义

$$Q_i=\Big\{\prod_{k=1}^{i}\frac{Z_{(k)}}{1-Z_{(k)}}\Big\}^{1/n},\qquad 1\le i\le n,$$

如果所有的 $Z_i$ 各不相同，由 [10] 给出的 (1.2) 中 $A$ 的非参数估计为

$$A_n(t)=\begin{cases}(1-t)Q_n^{1-p(t)}, & 0\le t\le Z_{(1)},\\ t^{i/n}(1-t)^{1-i/n}Q_n^{1-p(t)}Q_i^{-1}, & Z_{(i)}\le t\le Z_{(i+1)},\ 1\le i\le n-1,\\ tQ_n^{-p(t)}, & Z_{(n)}\le t\le 1,\end{cases} \tag{1.3}$$

* 本项目受国家自然科学基金 (19871061) 及南开大学天津大学刘徽应用数学中心的支持．
本文 2000 年 11 月 13 日收到，2001 年 9 月 3 日收到修改稿．

其中 $p$ 是 $[0.1]$ 上的有界函数，满足 $p(0)=1-p(1)=1$, [10] 证明了 $A_n$ 是 $A$ 的一个渐近无偏估计，且是一致的、强相合的.

特别当 $F_1(x)=F_2(y)=\exp(-\exp(-x))$, 即 Gumbel 边缘情况，由 (1.3) 定义的统计量 $A_n(t)$ 实际上完全由 $\sum_{i=1}^{n}|X_i-Y_i|$ 所决定. 联系 [12] 给出的 $W=Y-X$ 的分布

$$D(w)=\frac{e^w}{1+e^w}+\frac{A'(w)}{A(w)},$$

以及 $V=\exp(-x)+\exp(-y)$ 与 $W$ 的联合分布

$$\mathsf{P}(V\le v,W\le w)=(1-e^{-vA(w)})D(w)-v\int_{-\infty}^{w}\mathrm{d}\Big(t\frac{A(t)}{1+e^t}e^{-vA(t)}D(t)\Big).$$

显然，这个分布形式仍然比较复杂，不便进一步讨论.

关于 $A$ 的参数模型也有不少文章讨论， Logistic 模型是在理论上和实际中最常见的，有关的文章也较多，例如 [5] 及 [13,14] 是关于二元极值的讨论，而 [15-20] 是关于多元极值的. 对嵌套 Logistic 模型也有一些文章，见 [21-22]. 而其它参数模型的理论研究及应用则牵涉较少，似乎还没有专门的文章. 对每个指定的参数模型，实际上也就规定了极值的相关结构，对它们的讨论，实际上也是关于相应的相关结构性质的研究.

另一种途径是讨论一般二元分布的尾部相关特征， [23,24] 提出了尾部的几乎独立的概念， [25] 讨论了二元 Cauchy 分布的尾部相关特征.

我们则想从另一个角度讨论相关性，即能否通过一个适当的变换，使变换后的变量是独立的，或虽然不能完全独立，但至少在某种意义下具有类似于独立变量的方便性质，我们称这种性质为基本独立性. 这个想法在 [14,17,21] 已经实现. 他们考虑了二元及多元 Logistic 模型， 3 元嵌套 Logistic 模型，都得到了所希望的变换，那么对一般的二元极值分布是否可以找到具有如此性质的变换呢？我们在下一节给出回答，第 3 节是几个例子，第 4 节是对所讨论问题的展望.

## §2. 主要结果

考虑二元极值 Copula(1.2). 为方便起见，我们只限于可微模型，即相关函数存在二阶导数，不难求得 (1.2) 的密度函数为

$$c(u,v)=C(u,v)\frac{A^2(t)}{uv}\Big\{1+\frac{\log u-\log v}{\log(uv)}\frac{A'(t)}{A(t)}-\frac{\log u\log v}{(\log(uv))^2}\Big(\frac{A'(t)}{A(t)}\Big)^2-\frac{\log u\log v}{(\log(uv))^3}\frac{A''(t)}{A^2(t)}\Big\},\tag{2.1}$$

其中 $t=\log u/\log(uv)$, 作变换

$$s=-\log(uv)A\Big(\frac{\log u}{\log(uv)}\Big),\qquad t=\frac{\log u}{\log(uv)},$$

或

$$\begin{cases}u=\exp\Big(-\dfrac{st}{A(t)}\Big),\\ v=\exp\Big(-\dfrac{s(1-t)}{A(t)}\Big).\end{cases}\tag{2.2}$$

可以得到 $(S,T)$ 的联合分布密度函数

$$g(s,t)=e^{-s}\Big\{\Big(1-t\frac{A'(t)}{A(t)}\Big)\Big(1+(1-t)\frac{A'(t)}{A(t)}\Big)s+t(1-t)\frac{A''(t)}{A(t)}\Big\}.\tag{2.3}$$

如果记

$$p_1(t)=\Big(1-t\frac{A'(t)}{A(t)}\Big)\Big(1+(1-t)\frac{A'(t)}{A(t)}\Big),\tag{2.4}$$

$$p_2(t)=t(1-t)\frac{A''(t)}{A(t)},\tag{2.5}$$

那么，(2.3) 成为

$$g(s,t)=e^{-s}(p_1(t)s+p_2(t)). \tag{2.6}$$

由 (2.6) 可见，虽然 $(S,T)$ 不是独立的，但它们之间的相关结构非常简单．显然 $T$ 的边缘分布密度函数

$$p(t)=p_1(t)+p_2(t)=1+(1-2t)\frac{A'(t)}{A(t)}-t(1-t)\Big\{\Big(\frac{A'(t)}{A(t)}\Big)^2-\frac{A''(t)}{A(t)}\Big\}. \tag{2.7}$$

而 $S$ 的的边缘分布密度函数为

$$g(s)=((1-\beta)s+\beta)e^{-s},\qquad s>0, \tag{2.8}$$

其中 $\beta=\int_0^1 p_2(t)\mathrm{d}t$. 即 $S$ 的分布是 $1-\beta:\ \beta$ 的 $\Gamma(s,2)$ 与 $\Gamma(s,1)$ 混合 gamma 分布．这里 $\Gamma(s,k)$ 表示参数为 $k$ 的 gamma 分布密度 $\Gamma(s,k)=s^{k-1}e^{-s}/\Gamma(k),\ s>0$, 而 $\Gamma(\cdot)$ 是 gamma 函数．

由此，我们可以方便地求出任何形如 $H_1(S)H_2(T)$ 函数的数学期望：

$$\begin{aligned}\mathsf{E}(H_1(S)H_2(T)) &= \int_0^\infty\int_0^1 H_1(s)H_2(t)e^{-s}(p_1(t)s+p_2(t))\mathrm{d}t\mathrm{d}s\\ &= \int_0^\infty H_1(s)se^{-s}\mathrm{d}s\int_0^1 H_2(t)p_1(t)\mathrm{d}t+\int_0^\infty H_1(s)e^{-s}\mathrm{d}s\int_0^1 H_2(t)p_2(t)\mathrm{d}t.\end{aligned} \tag{2.9}$$

与独立的随机变量比较，(2.9) 给出的结果并没有许多本质的不同，我们不妨称这样的 $S,T$ 为基本独立的随机变量．在下一节讨论的具体例子中，经进一步的变换，由基本独立的随机变量可以得到真正独立的随机变量．

在二元极值研究中，感兴趣的常常是两个随机变量同时取大值的概率，即需要考虑联合生存函数

$$\overline{G}(x,y)=\mathsf{P}(X>x,Y>y). \tag{2.10}$$

相应于 (1.2), 我们有

$$\overline{G}(x,y)=\overline{C}(\overline{F}_1(x),\overline{F}_2(y)), \tag{2.11}$$

其中 $\overline{C}(u,v)=u+v+C(1-u,1-v)-1$, 而 $\overline{F}(x)=\mathsf{P}(X>x)=1-F(x)$ 是边缘生存函数．显然 (1.1) 与 (2.11) 有相同的密度函数 (如果存在的话), 因此可以方便地讨论其中之一即可．

## §3. 例

在极值的理论研究与应用中，最主要的可微参数模型是 logistic 模型及混合模型，较复杂一些的是将它们推广到不对称的情况．因为在某些实际问题中，假定变量是可交换的或对称的，并不合理．本节只给出一般结果在前二种参数模型下的具体形式．为叙述方便起见，且不失一般性，假定边缘分布为指数分布，$F_1(x)=F_2(x)=1-\exp(-x),\ x>0$, 其生存函数 $\overline{F}_1(x)=\overline{F}_2(x)=\exp(-x),\ x>0$ 具有简单形式，相应于变换 (2.2). 我们有

$$\begin{cases}s=(x+y)A\Big(\dfrac{x}{x+y}\Big),\\ t=\dfrac{x}{x+y}.\end{cases} \tag{3.1}$$

**例 1**　Logistic 模型．

Logistic 模型是二元极值中应用最广泛的一个参数模型，其相关函数为

$$A(t)=(t^{1/\alpha}+(1-t)^{1/\alpha}))^\alpha,\qquad 0\le t\le 1,$$

其中 $0<\alpha<1$ 是相关参数，相应的生存函数为

$$\overline{G}(x,y)=\exp\{-(x^{1/\alpha}+y^{1/\alpha})^\alpha\},\qquad x>0,\ y>0.$$

此时，变换 (3.1) 成为

$$\begin{cases} s = (x^{1/\alpha} + y^{1/\alpha})^{\alpha}, \\ t = x/(x+y). \end{cases} \tag{3.2}$$

可以求出 (2.4), (2.5) 中的 $p_1(t)$, $p_2(t)$ 分别为

$$\begin{aligned} p_1(t) &= \frac{t^{1/\alpha-1}(1-t)^{1/\alpha-1}}{(t^{1/\alpha}+(1-t)^{1/\alpha})^2}, \\ p_2(t) &= \Big(\frac{1}{\alpha}-1\Big)\frac{t^{1/\alpha-1}(1-t)^{1/\alpha-1}}{(t^{1/\alpha}+(1-t)^{1/\alpha})^2}. \end{aligned}$$

因此 $T$ 的边缘分布密度为

$$p(t) = \frac{1}{\alpha}\frac{t^{1/\alpha-1}(1-t)^{1/\alpha-1}}{(t^{1/\alpha}+(1-t)^{1/\alpha})^2}, \qquad 0<t<1.$$

这个密度函数形式似乎比较复杂，如若作进一步变换

$$t_1 = \frac{t^{1/\alpha}}{t^{1/\alpha}+(1-t)^{1/\alpha}},$$

那么 $T_1$ 即为区间 $(0,1)$ 上的均匀分布变量. 或者说，如果作变换

$$\begin{cases} s = (x^{1/\alpha} + y^{1/\alpha})^{\alpha}, \\ t = x^{1/\alpha}/(x^{1/\alpha} + y^{1/\alpha}), \end{cases} \tag{3.3}$$

那么由 (2.8), $S$ 的边缘分布为 $\beta$: $1-\beta$ 的混合分布，其中 $\beta = \int_0^1 p_2(t)\mathrm{d}t = 1-\alpha$, 即 $S$ 为 $1-\alpha$: $\alpha$ 的混合分布，而 $T_1$ 是均匀分布变量，且 $S$, $T$ 相互独立. 这个结果与 [14, 16] 完全一致，而且这里的结果更深刻，因为这里告诉我们，服从区间 (0,1) 上的均匀分布随机变量 $T_1$ 是两个部分 $p_1(t)$, $p_2(t)$ 叠加的结果. 实际上，我们已经得到了 $X$, $Y$ 的随机表示

$$\begin{cases} X = ST^{\alpha}, \\ Y = S(1-T)^{\alpha}. \end{cases}$$

**例 2** 混合模型.

混合模型的相关函数为

$$A(t) = \theta t^2 - \theta t + 1, \qquad 0<t<1,$$

其中 $0<\theta<1$ 是相关参数，相应的生存函数为

$$\overline{G}(x,y) = \exp\Big\{-(x+y) + \frac{\theta xy}{x+y}\Big\}, \qquad x>0,\ y>0.$$

作变换

$$\begin{cases} s = (x+y) - \dfrac{\theta xy}{x+y}, \\ t = \dfrac{x}{x+y}. \end{cases} \tag{3.4}$$

可以得到

$$\begin{aligned} p_1(t) &= \frac{4-\theta}{(\theta t^2-\theta t+1)^2} - \frac{4}{\theta t^2-\theta t+1} + 1, \\ p_2(t) &= \frac{2t(1-t)\theta}{\theta t^2-\theta t+1}. \end{aligned}$$

$T$ 的边缘分布密度

$$p(t) = \frac{4-\theta}{(\theta t^2-\theta t+1)^2} - \frac{2}{\theta t^2-\theta t+1} - 1, \qquad 0<t<1.$$

$S$ 所服从的混合分布中的比例常数

$$\beta = \int_0^1 p_2(t)\mathrm{d}t = \frac{8}{\sqrt{\theta(4-\theta)}} \arctan\sqrt{\frac{\theta}{4-\theta}} - 2.$$

考虑进一步的变换

$$t_1 = \frac{2t-1}{\theta t^2 - \theta t + 1}, \tag{3.5}$$

不难证明，它将 $[p(t)+1]/2$ 变换为 $[-1,1]$ 上的均匀分布变量，$h(t_1) = 1/2$，$-1 < t < 1$. 因此 $S$ 与 $T_1$ 相互独立，这个结果是本文第一次得到的.

由变换 (3.4) 容易得到

$$xy = \frac{s^2 t(1-t)}{A^2(t)}.$$

由 (2.9) 可知，这种形式的数学期望是不难求出的. 实际上，我们有

$$\mathsf{E}(XY) = \frac{2}{4-\theta}\Big(\frac{4}{\sqrt{\theta(4-\theta)}} \arctan\sqrt{\frac{\theta}{4-\theta}} + 1\Big),$$

故由指数分布的性质可知，$X$、$Y$ 之间的相关系数为

$$\rho = \frac{\theta-2}{4-\theta} + \frac{8}{(4-\theta)\sqrt{\theta(4-\theta)}} \arctan\sqrt{\frac{\theta}{4-\theta}}.$$

## §4. 结　　论

由于二元极值分布的相关函数 $A$ 不可能用有限的参数形式来表示. 因此对二元极值分布的讨论，常常在一定的参数模型下进行的. 目前，在各种文献中见到的参数模型已有许多，它们大都是从数学上考虑而提出的，真正在实际中得到应用的主要还是上节例 1 讨论的 Logistic 模型.

关于 $A$ 的半参数，非参数估计，最近有不少文章讨论，但是如果某个参数模型比较适合于所考虑的实际问题，且具有某种稳健性质，那么在此参数模型下的结果，可能比一般的非参数模型好.

变换 (2.2) 的具体执行仍依赖于 $A$ 的形式，例 1、例 2 指出，对给定的 $A$ 的参数形式，必定能得到所希望的结果，而且常常可由此作进一步的变换，使最后的变量相互独立. 此时，许多问题变得相对简单了. 文献 [14-18] 关于 Logistic 模型，[21] 关于嵌套 Logistic 模型的一系列结果就是在变换 (3.3) 下得到的. 特别值得一提的是变换 (3.3) 使在计算机上模拟产生 Logistic 模型下的二元极值伪随机向量变得非常简单，[26] 给出了以上模型随机向量产生的有效算法. 这为研究其它比较复杂的问题提供了一个非常好的基础，如果不能得到理论上的结果，至少也有模拟结果.

我们相信变换 (3.4) 及 (3.5) 对混合模型的研究提供了最好的准备.

由于变换 (2.2) 的一般性，当然适用于二元极值分布的其它参数模型. 当参数模型本身表示比较复杂时，(2.4). (2.5) 中的形式也就更加复杂，如果不能找到进一步的变换使 T 的密度函数 $p(t)$ 有较简单的形式，或者连 $S$ 的混合分布中比例常数 $\beta$ 都不能积出，这些多少都影响到我们的最终目标. 例如关于不对称的 Logistic 模型及不对称的混合模型就是这样，对此我们需要进一步的研究.

## 参 考 文 献


[1] Sklar. A., Functions de répartition à $n$ dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, **8**(1959), 229-231.

[2] Joe, H., *Multivariate Models and Dependence Concepts*, Chapman & Hall, London, 1997.

[3] Nelsen, R.B., *An Introduction to Copulas, Lecture Notes in Statistics Vol. 139*, Springer-Verlag, New York, 1999.

[4] Hutchinson, T.P. and Lai, C.D., *Continuous Bivariate Distributions, Emphasising Applications*, Rumsby Scientific, Adelaide, 1991.

[5] Tawn, J.A., Bivariate extremes value theory: models and estimation, *Biometrika*, **75**(1988), 397–415.

[6] Pickands, J., Multivariate extreme value distributions, In Proc. 43rd Sess., Amsterdam: International Statistical Institute, 1981, 859–878.

[7] Deheuvels. P., On the limiting behavior of the Pickands estimatior for bivariate extreme-value distributions, *Statist. Prob. Lett.*, **12**(1991), 429–439.

[8] Smith, R.L., Tawn, J.A. and Yuen, H.K., Statistics of multivariate extremes, *Int. Statist. Rev.*, **58**(1990), 47–58.

[9] Genest. C., Ghoadi, K. & Rivest, L.-P., A semiparametric estimation procedure of dependence parameters in multivariate familas of distribution, *Biometrika*, **82**(1995), 543–552.

[10] Caperaa. P., Fougeres, A.-L. & Gnenest, C., A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*. **84**(1997), 567–577.

[11] Hall, P. and Tajvidi, N., Distribution and dependence function estimation for bivariate extreme-value distribution, *Bernoulli*, **6**(2000), 835–844.

[12] Tiago de Oliveira, J., Bivariate modles for extremes: stataistical dicison, in *Statistical Extremes and Applicatıons*, Ed. J. Tiago de Oliveira, 131–153, Dordrecht, Reidel, 1984.

[13] Joe, H., Smith, R.L. and Weissman, I., Bivariate threshold models for extremes, *J. R. Statist.*, **B 54**(1992), 171–183.

[14] Oakes, D. and Manatunga, A.K., Fisher information for a bivariate extreme value distribution, *Biometrika*, **79**(1992), 827–832.

[15] Shi Daoji, Moment estimation for multivariate extreme value distribution, *Applied Mathematics - A Journal of Chinese Universities*, **B 10**(1995a), 61–68.

[16] Shi Daoji, Multivariate extreme value distribution and its Fisher information matrix, *Acta Mathematicae Applicatae Sinica* (English Series), **11**(1995b), 421–428.

[17] Shi Daoji, Fisher information for a multivariate extreme value distribution, *Biometrika*, **82**(1995c), 664–669.

[18] 史道济，冯燕奇，多元极值分布参数的最大似然估计与分步估计，系统科学与数学，**17**(1997), 244–251.

[19] Coles. S.G. and Tawn, J.A., Modelling multivariate extreme events, *J. R. Statist. Soc.*, **B 53**(1991), 377–392.

[20] Coles, S.G. and Tawn, J.A., Statistical method for multivariate extremes: an apllication to structural design, *Appl. Statist.*, **43**(1994), 1–49.

[21] Shi Daoji and Zhou Shengsheng, Moment estimation for multivariate extreme value distributio in nested logistic model, *Annals of the Institute of Statistical Mathematics*, **51**(1999), 253–264.

[22] Tawn, J.A., Modelling multivariate extreme value distributions, *Biometrika*, **77**(1990), 245–253.

[23] Ledford, A.W. and Tawn, J.A., Statistics for near independence in multivariate extreme values, *Biometrika*, **83**(1996), 169–187.

[24] Ledford, A.W. and Tawn, J.A., Modelling dependence within joint tail regions, *J. R. Statist. Soc.*, **B 59**(1997), 475–499.

[25] 孙炳堃，史道济，二元 Cauchy 分布尾部的相关性，天津大学学报，**33**(2000), 432–434.

[26] 史道济，阮明书，王毓娥，多元极值分布随机向量的抽样方法，应用概率统计，**13**(1997), 75–80.

## A Property for Bivariate Extreme Value Distribution

SHI DAOJI

*(Department of Mathematics, Tianjin University; LiuHui Center for Applied Mathematics, Nankai University & Tainjin University, Tianjin, 300072)*

The dependence structure of bivariate extreme value distribution is considered. A transform of variables is proposed, such that transformed variates are independent basically. Also we obtain the stochastic representation for bivariate extremes. From these it is easy to generate pseudo random vector of bivariate extreme distributed in computer and to calculate numerical characteristic of a class of usual statistics. These are basics to study asymptotic distribution of some statistics.