

具有 $L^d(\mu, \Sigma)$ 或 $M^d(\mu, \Sigma)$ 分布的成分向量的独立性

杨维权 余锦华
(数学系)

摘 要

讨论了具有 $L^d(\mu, \Sigma)$ 或 $M^d(\mu, \Sigma)$ 分布的随机成分向量的独立性, 得到某些新的结果。

关键词 成分向量, 对数一比变换, 基独立, 剖分独立性, 子成分不变性, 完全子成分独立性

记 $S^d = \{(x_1, \dots, x_{d+1})' : x_i > 0, \sum_{i=1}^{d+1} x_i = 1\}$

$$R_+^{d+1} = \{(w_1, \dots, w_{d+1})' : w_i > 0, i = 1, \dots, d+1\}$$

若 $X^{(d+1)} = (x_1, \dots, x_{d+1})' \in S^d$, 则称 $X^{(d+1)}$ 为 d 维成分向量, 对于任何 $W^{(d+1)} = (w_1, \dots, w_{d+1})' \in R_+^{d+1}$, 若经变换:

$$x_i = w_i / \sum_{j=1}^{d+1} w_j \quad (1)$$

有 $X^{(d+1)} \in S^d$, 则称 $W^{(d+1)}$ 为 $X^{(d+1)}$ 的基向量。

定义 1 设 $X^{(d+1)} \in S^d$, 作对数-比变换:

$$y_i = \lg(x_i/x_{d+1}) \quad (i = 1, \dots, d) \quad (2)$$

若 $Y^{(d)} = (y_1, \dots, y_d)'$ 服从 $N_d(\mu, \Sigma)$ 分布, 则称 $X^{(d+1)}$ 服从 $L^d(\mu, \Sigma)$ 分布。

定义 2 设 $X^{(d+1)} \in S^d$, 作变换:

$$z_i = \lg[x_i / (1 - \sum_{j=1}^i x_j)] \quad (i = 1, \dots, d) \quad (3)$$

本文1987年5月收到

若 $Z^{(d)} = (z_1, \dots, z_d)'$ 服从 $N_d(\mu, \Sigma)$ 分布, 则称 $X^{(d+1)}$ 服从 $M^d(\mu, \Sigma)$ 分布.

由于 $X^{(d+1)}$ 的分量之和为 1, 因而分量之间的独立性不存在, 需引进较弱的几种独立性的概念.

$$\text{记 } X^{(d+1)} = \begin{pmatrix} X^{(c)} \\ \dots \\ X_{(c)} \end{pmatrix}, \quad X^{(c)} = \begin{pmatrix} x_1 \\ \vdots \\ x_c \end{pmatrix}, \quad X_{(c)} = \begin{pmatrix} x_{c+1} \\ \vdots \\ x_{d+1} \end{pmatrix},$$

$$t_1 = \sum_{i=1}^c x_i \triangleq T(X^{(c)}), \quad t_2 = \sum_{i=c+1}^{d+1} x_i \triangleq T(X_{(c)}),$$

$$t^{(2)} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix},$$

$$A_{(1)} = C(X^{(c)}) = X^{(c)} / T(X^{(c)}),$$

$$A_{(2)} = C(X_{(c)}) = X_{(c)} / T(X_{(c)}), \quad 1 < c < d+1.$$

则称 $(A_{(1)}, A_{(2)}, t^{(2)})$ 为 $X^{(d+1)}$ 的二段剖分, $A_{(1)}, A_{(2)}$ 为 $X^{(d+1)}$ 的两个子成分向量. $t^{(2)}$ 也为一成分向量. 如(1)式所示, $X^{(c)}, X_{(c)}$ 分别为 $A_{(1)}, A_{(2)}$ 的基向量.

定义 3 若 $C(X^{(c)}) \text{II} C(X_{(c)}) \text{II} T(X^{(c)})$,

则称 $X^{(d+1)}$ 具有剖分独立性 (II 表示相互独立的记号, 以下同).

定义 4 若右中性: $X^{(c)} \text{II} C(X_{(c)})$, 左中性: $C(X^{(c)}) \text{II} X_{(c)}$

同时成立, 则称 $X^{(d+1)}$ 具有中性.

定义 5 若左不变性: $C(X^{(c)}) \text{II} T(X^{(c)})$, 右不变性: $C(X_{(c)}) \text{II} T(X_{(c)})$

同时成立, 则称 $X^{(d+1)}$ 具有子成分不变性.

定义 6 若 $C(X^{(c)}) \text{II} C(X_{(c)})$,

则称 $X^{(d+1)}$ 具有子成分独立性.

若对于 $X^{(d+1)}$ 的各分量的任意排列及任意剖分, 上述独立性成立, 则相应地可加上“完全”性, 如完全剖分独立性、完全子成分独立性等. 上述各独立性概念, 可以证明具有如下关系:

右中性、左不变性 (或左中性、右不变性) $\Rightarrow$ 剖分独立性 $\Rightarrow$ 中性 $\Rightarrow$ 子成分独立性 (或子成分不变性)

本文证明了下述结论成立.

命题 1 若 $X^{(d+1)} \in S^d$, 具有独立基向量 $W^{(d+1)} \in R_+^{d+1}$, 则 $X^{(d+1)}$ 具有完全子成分独立性.

证明 设 $c_1, \dots, c_k$ 为 k 个正整数, 满足

$$0 = c_0 < c_1 < \dots < c_k = d+1$$

将 $X^{(d+1)}$ 分成 k 段:

$$X_{(j)} = (x_{c_{j-1}+1}, \dots, x_{c_j})' \quad (j=1, \dots, k).$$

建立 k 个子成分向量 $A_{(1)}, \dots, A_{(k)}$, 其中

$$A_{(j)} = C(X_{(j)}) = X_{(j)} / \sum_{r=c_{j-1}+1}^{c_j} x_r \quad (j=1, \dots, k).$$

另一方面, 由于

$$X^{(d+1)} = C(W^{(d+1)}) = \frac{1}{t} W^{(d+1)}, \quad t \triangleq T(W^{(d+1)})$$

于是

$$A_{(j)} = C(W_{(j)}) = W_{(j)} / \sum_{r=c_{j-1}+1}^{c_j} w_r,$$

$$W_{(j)} = (w_{c_{j-1}+1}, \dots, w_{c_j})', \quad (j=1, \dots, k)$$

由假设知 $w_1, \dots, w_{d+1}$ 相互独立, 所以 $A_{(1)}, \dots, A_{(k)}$ 是 k 个相互独立的子成分向量, 即 $X^{(d+1)}$ 具有完全子成分独立性.

命题2 若 $X^{(d+1)} \in Sd$, 具有完全子成分独立性, 经变换(1)式得到 $Y^{(d)}$, 则 $Y^{(d)}$ 的协方差阵为

$$\Sigma = \text{diag}(\lambda_1, \dots, \lambda_d) + \lambda_{d+1} J_d \quad (4)$$

其中 J_d 是元素全为1的 d 阶方阵.

证明 对于任何 $1 \leq i \neq j \neq k \neq l \leq d+1$, 若 $C(x_i, x_j) \amalg C(x_k, x_l)$, 即

$$\left(\frac{x_i}{x_i + x_j}, \frac{x_j}{x_i + x_j} \right) \amalg \left(\frac{x_k}{x_k + x_l}, \frac{x_l}{x_k + x_l} \right)$$

于是

$$\left(\frac{1}{1 + x_j/x_i}, \frac{1}{1 + x_i/x_j} \right) \amalg \left(\frac{1}{1 + x_l/x_k}, \frac{1}{1 + x_k/x_l} \right)$$

也即 x_i/x_j 与 x_k/x_l 相互独立, 从而

$$\text{cov}(\lg \frac{x_i}{x_j}, \lg \frac{x_k}{x_l}) = 0$$

$$\sigma_{ij} \triangleq \text{cov}(y_i, y_j) = \text{cov}(\lg \frac{x_i}{x_{d+1}}, \lg \frac{x_j}{x_{d+1}})$$

$$= \text{cov}(\lg \frac{x_i}{x_{d+1}}, \lg \frac{x_j}{x_k} + \lg \frac{x_k}{x_{d+1}}) = \text{cov}(\lg \frac{x_i}{x_{d+1}}, \lg \frac{x_k}{x_{d+1}}) \triangleq \sigma_{ik}$$

其中 $k \neq i \neq d+1$. 可见当 $i \neq j$ 时, σ_{ij} 都为同一常数, 不妨记作 λ_{d+1} . 当 $i = j$ 时, 记

$$\sigma_{ii} \triangleq \text{cov}(y_i, y_i) \triangleq \lambda_i + \lambda_{d+1} \quad (i=1, \dots, d).$$

从而(4)式成立.

命题3 若 $X^{(d+1)} \in Sd$, 具有 $Ld(\mu, \Sigma)$ 分布, 则 $X^{(d+1)}$ 具有完全子成分独立性的充要条件为(4)式成立.

证明 必要性由命题2知成立. 现证充分性. 若 $Y^{(d)}$ 的协方差阵为(4)式, 即有

$$\sigma_{ij} = \text{cov}(y_i, y_j) = \begin{cases} \lambda_i + \lambda_{d+1}, & i = j \\ \lambda_{d+1}, & i \neq j \end{cases}$$

又因 $y_i = \lg(x_i/x_{d+1})$, $(i=1, \dots, d)$, 于是

$$x_i = \begin{cases} e^{y_i} / (1 + \sum_{j=1}^d e^{y_j}), & (i=1, \dots, d) \\ 1 / (1 + \sum_{j=1}^d e^{y_j}), & (i=d+1) \end{cases}$$

$$X^{(d+1)} = (x_1, \dots, x_d, x_{d+1})' = C(e^{y_1}, \dots, e^{y_d}, 1) \triangleq C(W^{(d+1)})$$

其中 $W^{(d+1)} = (e^{y_1}, \dots, e^{y_d}, 1)' \in R_+^{d+1}$.

如命题1的证明所示, 对 $X^{(d+1)}$ 作 k 段剖分, 则有

$$A_{(1)} = (e^{y_1}, \dots, e^{y_{c_1}})' / \sum_{j=1}^{c_1} e^{y_j} = (1, e^{y_2-y_1}, \dots, e^{y_{c_1}-y_1})' / (1 + \sum_{j=2}^{c_1} e^{y_j-y_1})$$

$$= C(1, e^{y_2-y_1}, \dots, e^{y_{c_1}-y_1})$$

$$A_{(2)} = C(1, e^{y_{c_1+2}-y_{c_1+1}}, \dots, e^{y_{c_2}-y_{c_1+1}})$$

⋮

$$A_{(k)} = C(1, e^{y_{c_{k-1}+1}}, \dots, e^{y_d})$$

对于 $A_{(1)}$ 与 $A_{(2)}$, 由于对任意 $1 < i_1 \leq c_1$, $c_1 + 1 < i_2 \leq c_2$, 有

$$\text{cov}(y_{i_1} - y_1, y_{i_2} - y_{c_1+1}) = \lambda_{d+1} - \lambda_{d+1} - \lambda_{d+1} + \lambda_{d+1} = 0$$

由假设知 $(y_1, \dots, y_{c_1})'$ 及 $(y_{c_1+1}, \dots, y_{c_2})'$ 都服从正态分布, 所以 $(1, y_2 - y_1, \dots, y_{c_1} - y_1)'$ 与 $(1, y_{c_1+2} - y_{c_1+1}, \dots, y_{c_2} - y_{c_1+1})$ 相互独立, 因而 $A_{(1)}$ 与 $A_{(2)}$ 相互独立.

同理可证 $A_{(1)}, A_{(2)}, \dots, A_{(k)}$ 相互独立, 也即 $X^{(d+1)}$ 具有完全子成分独立性.

命题4 若 $X^{(d+1)} \in S^d$, 具有 $M^d(\mu, \Sigma)$ 分布, 且有

$$\Sigma = \text{diag}(\lambda_1, \dots, \lambda_d) \quad (5)$$

则 $X^{(d+1)}$ 具有完全子成分独立性.

证明 由(3)式知

$$x_i = \begin{cases} e^{y_i} / \prod_{j=1}^i (1 + e^{y_j}), & (i = 1, \dots, d) \\ 1 / \prod_{j=1}^d (1 + e^{y_j}), & (i = d+1) \end{cases}$$

如命题1的证明所示, 对 $X^{(d+1)}$ 作 k 段剖分, 则有

$$A_{(1)} = (x_1, \dots, x_{c_1}) / \sum_{j=1}^{c_1} x_j \triangleq g_1(e^{y_1}, \dots, e^{y_{c_1}})$$

$$A_{(2)} = (x_{c_1+1}, \dots, x_{c_2}) / \sum_{j=c_1+1}^{c_2} x_j \triangleq g_2(e^{y_{c_1+1}}, \dots, e^{y_{c_2}})$$

⋮

$$A_{(k)} \triangleq g_k(e^{y_{c_{k-1}+1}}, \dots, e^{y_d}, 1).$$

由于对任意 $1 \leq i_1 < i_2 \leq c_1$, 有

$$x_{i_2} / x_{i_1} = e^{y_{i_1} - y_{i_2}} \prod_{j=i_1+1}^{i_2} (1 + e^{y_j}),$$

对任何 $1 \leq i_1 < i_2 \leq c_2 - c_1$, 有

$$x_{(c_1+i_2)} / x_{(c_1+i_1)} = e^{y_{(c_1+i_2)} - y_{(c_1+i_1)}} \prod_{j=c_1+i_1+1}^{c_1+i_2} (1 + e^{y_j}),$$

同理, 对其它子成分向量也可知其相应的结构, 由假设知 $y_1, \dots, y_d$ 为独立正态分布变量, 从而 $A_{(1)}, \dots, A_{(k)}$ 相互独立, 即 $X^{(d+1)}$ 具有完全子成分独立性.

由于服从Dirichlet分布的 $X^{(d+1)}$, 具有完全部分独立性, 即有较强的独立性, 因此 $L^d(\mu, \Sigma)$ 及 $M^d(\mu, \Sigma)$ 分布比Dirichlet分布的适用面较宽.

参 考 文 献

- [1] Aitchison, J., The Statistical Analysis of Compositional Data, *J.R. statist. Soc.*, B. 44(1982).
- [2] Aitchison, J., A General Class of Distributions on the Simplex, *J.R. statist. Soc.*, B. 47(1985).

Independence of the Compositional Vectors with $L^d(\mu, \Sigma)$ or $M^d(\mu, \Sigma)$ Distributions

Yang Weiquan Yu Jinghua

Abstract

For compositional vector $X^{(d+1)} = (x, \dots, x_{d+1})'$ of $(d+1)$ components the sample space is the positive simplex

$$S^d = \{(x, \dots, x_{d+1})' : x_i > 0, \sum_{i=1}^{d+1} x_i = 1\}$$

We have proved:

1° If $X^{(d+1)} \in S^d$ has basis independence, then it has complete subcompositional independence.

2° If $X^{(d+1)} \in S^d$ has complete subcompositional independence, then $\text{cov}\{\lg(X^{(d)} / x_{d+1})\} = \text{diag}(\lambda_1, \dots, \lambda_d) + \lambda_{d+1} J_d$.

3° If $X^{(d+1)} \sim L^d(\mu, \Sigma)$, then the following statements are equivalent:

(1) $X^{(d+1)}$ has complete subcompositional independence.

(2) $\text{cov}\{\lg(X^{(d)} / x_{d+1})\} = \text{diag}(\lambda_1, \dots, \lambda_d) + \lambda_{d+1} J_d$.

(3) $X^{(d+1)}$ is logratio uncorrelation.

4° If $X^{(d+1)} \sim M^d(\mu, \Sigma)$, $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_d)$, then $X^{(d+1)}$ has complete subcompositional independence.

Keywords compositional vector, logratio, basis independence, partition independence, subcompositional invariance, complete subcompositional independence