文章编号: 1008-1542(2005)03-0191-03

样本区间的概率分布及其性质

陈光曙[1], 韦相和[2]

(1. 淮阴师范学院数学系, 江苏淮安, 223300; 2. 淮阴师范学院计算机科学系, 江苏淮安, 223001)

摘　要: 设$(X_1, X_2, \cdots, X_n)$为服从$I=[0,1]$上的均匀分布的简单随机样本, 它们将$[0,1]$分成$(n+1)$个样本区间, 以$Y_0, Y_1, \cdots, Y_n$分别表示这些样本区间的长度, 利用顺序统计量的性质讨论$Y_i(1\leqslant i\leqslant n+1)$的概率性质。

关键词: 顺序统计量; 均匀分布; 样本区间

中图分类号: O211.1　　文献标识码: A

Properties of sample spacing probability distribution

CHEN Guang-shu[1], WEI Xiang-he[2]

(1. Department of Mathematics, Huaiyin Teachers' College, Huai'an Jiangsu, 223300, China; 2. Department of Computer Science and Technology, Huaiyin Teachers' College, Huai'an Jiangsu 223001, China)

Abstract: Let$(X_1, X_2, \cdots, X_n)$be a simple random sample of size n chosen according to uniform distribution on the unit interval $[0,1]$. Then the interval is divided into$(n+1)$sample spacing of lengths $Y_0, Y_1, \cdots, Y_n$ respectively. In this paper, we attained the properities of sample spacing probability distribution by discussing the behaviour of order statistics.

Key words: order statistics; uniform distribution; sample spacing

均匀分布是概率统计中的重要分布, 具有广泛的应用价值。由于随机选取的点集被广泛应用于许多概率模型中(如: 流行病学、遗传学及交通流理论等), 因此, 对样本区间及其相关统计量的研究也引起了众多学者的关注[1~3]。设随机变量Z在区间$[a,b]$上服从均匀分布, 对Z进行压缩变换$X=\frac{Z-a}{b-a}$, 则X在$I=[0,1]$上服从均匀分布。因此, 本文着重研究$[0,1]$上均匀分布的样本区间。

设$(X_1, X_2, \cdots, X_n)$为服从$I=[0,1]$上均匀分布的变量为n的简单随机样本, 且全不相等, 若有相等的则只取1个作代表值。记$Y_0, Y_1, \cdots, Y_n$分别为它们所产生的$(n+1)$个样本区间的长度, 以$X_{(1)}<X_{(2)}<\cdots<X_{(n)}$表示$(X_1, X_2, \cdots, X_n)$的顺序统计量, 令$X_{(0)}=0$, $X_{(n+1)}=1$, 则$Y_i=X_{(i+1)}-X_{(i)}$, $i=0,1,\cdots,n$, 易知: $\sum_{i=0}^{n}Y_i=1$。在文献[4]中, 作者利用顺序统计量的性质分别得到了样本空间长度$Y_i(0\leqslant i\leqslant n)$的分布概率密度以及$(Y_i, Y_j)(0\leqslant i<j\leqslant n)$的联合分布概率密度。本文在文献[1]的基础上进一步研究了样本区间长度的相关概率性质。

1　主要结果

在文献[1]中, 利用顺序统计量的性质分别得到了样本空间长度$Y_i(0\leqslant i\leqslant n)$的分布概率密度以及

收稿日期: 2005-01-12; 修回日期: 2005-04-02; 责任编辑: 王士忠
基金项目: 江苏省教育厅自然科学基金资助项目(03KJB110012)
作者简介: 陈光曙(1961-), 男, 江苏淮安人, 副教授, 主要从事概率论与数理统计方面的研究。

$(Y_i, Y_j)(0 \leqslant i < j \leqslant n)$ 的联合分布概率密度。为方便起见, 给出如下引理:

引理 **1** 若总体 $X \sim U(0, 1)$, ($U(0, 1)$ 表示[0, 1]区间上的均匀分布), $X_1, X_2, \cdots, X_n$ 为来自总体 X 的容量为 n 的简单随机样本, $Y_0, Y_1, \cdots, Y_n$ 分别为它们所产生的 $(n+1)$ 个样本区间的长度, 则 Y_i 的分布概率密度为

$$f(y) = \begin{cases} n(1-y)^{n-1}, & 0 < y < 1, \\ 0, & \text{其他}, \end{cases} \tag{1}$$

见文献[1]中的命题 4。

引理 **2** $Y_i, Y_j (0 \leqslant i < j \leqslant n)$ 的联合分布概率密度为

$$f(x, y) = \begin{cases} n(n-1)(1-x-y)^{n-2}, & x \geqslant 0, y \geqslant 0, x+y \leqslant 1, \\ 0, & \text{其他}, \end{cases} \tag{2}$$

见文献[1]中的命题 5。

利用引理 1、引理 2, 可以讨论 Y_i 以及 (Y_i, Y_j) 的相关概率性质, 给出如下定理。

定理 1 设 $Y_i (0 \leqslant i \leqslant n)$ 是来自总体 $X \sim U(0, 1)$ 的简单随机样本 $X_1, X_2, \cdots, X_n$ 所产生的样本区间长度, 则

$$1) E(Y_i) = \frac{1}{n+1}; \quad 2) E(Y_i^k) = \frac{k!}{(n+1)(n+2)\cdots(n+k)} \quad (k \text{ 为正整数}); \tag{3}$$

$$3) D(Y_i) = \frac{n}{(n+1)^2(n+2)}; \quad 4) \operatorname{cov}(Y_i, Y_j) = -\frac{1}{(n+1)^2(n+2)};$$

5) Y_i 与 Y_j 的相关系数 $\rho_{Y_i, Y_j} = -\frac{1}{n}$。

证明 ①由引理 1 知:

$$E(Y_i) = \int_0^1 y n(1-y)^{n-1} \mathrm{d}y = -\int_0^1 y \mathrm{d}(1-y)^n = \int_0^1 (1-y)^n \mathrm{d}y = \frac{1}{n+1};$$

$$②E(Y_i^k) = \int_0^1 y^k n(1-y)^{n-1} \mathrm{d}y = -\int_0^1 y^k \mathrm{d}(1-y)^n = k\int_0^1 y^{k-1}(1-y)^n \mathrm{d}y =$$

$$\frac{k}{n+1}\int_0^1 y^{k-1} \mathrm{d}(1-y)^{n+1} = \frac{k(k-1)}{n+1}\int_0^1 y^{k-2}(1-y)^{n+1} \mathrm{d}y =$$

$$\frac{k(k-1)}{(n+1)(n+2)}\int_0^1 y^{k-2} \mathrm{d}(1-y)^{n+2} = \cdots = \frac{k!}{(n+1)(n+2)\cdots(n+k)};$$

③由式(3), 令 $k = 2$ 得: $E(Y_i^2) = \frac{2}{(n+1)(n+2)}$; 由方差的性质, 易知:

$$D(Y_i) = E(Y_i^2) - [E(Y_i)]^2 = \frac{2}{(n+1)(n+2)} - \left(\frac{1}{n+1}\right)^2 = \frac{n}{(n+1)^2(n+2)};$$

④由引理 2, 得:

$$E(Y_i, Y_j) = \iint\limits_{\substack{x+y \leqslant 1 \\ x \geqslant 0, y \geqslant 0}} xyn(n-1)(1-x-y)^{n-2} \mathrm{d}x\,\mathrm{d}y =$$

$$\int_0^1 \int_0^{1-x} xyn(n-1)(1-x-y)^{n-2} \mathrm{d}x\,\mathrm{d}y = \int_0^1 x\left[\int_0^{1-x} yn(n-1)(1-x-y)^{n-2} \mathrm{d}y\right] \mathrm{d}x =$$

$$\int_0^1 x\left[\int_0^{1-x} n(1-x-y)^{n-1} \mathrm{d}y\right] \mathrm{d}x = \int_0^1 x(1-x)^n \mathrm{d}x = -\frac{1}{n+1}\int_0^1 x \mathrm{d}(1-x)^{n+1} =$$

$$\frac{1}{n+1}\int_0^1 (1-x)^{n+1} \mathrm{d}x = \frac{1}{(n+1)(n+2)},$$

所以得:

$$\operatorname{cov}(Y_i, Y_j) = E(Y_i, Y_j) - E(Y_i)E(Y_j) = \frac{1}{(n+1)(n+2)} - \left(\frac{1}{n+1}\right)^2 = -\frac{1}{(n+1)^2(n+2)}。$$

$$⑤ \rho_{Y_i, Y_j} = \frac{\operatorname{cov}(Y_i, Y_j)}{\sqrt{D(Y_i)D(Y_j)}} = \frac{-\frac{1}{(n+1)^2(n+2)}}{\frac{n}{(n+1)^2(n+2)}} = -\frac{1}{n}。$$

由定理 1 易知下列推论。

推论 **1**　Y_i 与 Y_j 总是负相关的。

推论 **2**　当容量 n 不断增大时, Y_i 与 Y_j 的相关性逐渐减弱, 且趋于 0。

定理 **2**　设 $\boldsymbol{B}$ 为 $Y_0, Y_1, \cdots, Y_n$ 的协方差矩阵, 则 $\boldsymbol{B}$ 的行列式为 $|\boldsymbol{B}| = 0$。

证明　由协方差的定义知

$$\boldsymbol{B} = (\mathrm{cov}(Y_i, Y_j))_{(n+1)\times(n+1)} = \frac{1}{(n+1)^2(n+2)} \begin{pmatrix} n & -1 & -1 & \cdots & -1 \\ -1 & n & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & n \end{pmatrix}_{(n+1)\times(n+1)},$$

取 $\boldsymbol{B}$ 的行列式, 得

$$|\boldsymbol{B}| = \left[\frac{1}{(n+1)^2(n+2)}\right]^{n+1} \begin{vmatrix} n & -1 & -1 & \cdots & -1 \\ -1 & n & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & n \end{vmatrix}_{(n+1)\times(n+1)},$$

分别将 2, 3, …, $n+1$ 行对应加到第 1 行, 得

$$|\boldsymbol{B}| = \left[\frac{1}{(n+1)^2(n+2)}\right]^{n+1} \begin{vmatrix} 0 & 0 & 0 & \cdots & 0 \\ -1 & n & -1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & n \end{vmatrix}_{(n+1)\times(n+1)} = 0。$$

由引理 1、引理 2 知, 当 $0 < y < 1$ 时, $Y_j = y$ 的条件下, Y_i 的条件分布概率密度为

$$f_{Y_i|Y_j}(x|y) = \frac{f(x,y)}{f(y)} = \begin{cases} \dfrac{(n-1)(1-x-y)^{n-2}}{(1-y)^{n-1}}, & x \geqslant 0, y > 0, x+y < 1, \\ 0, & \text{其他。} \end{cases}$$

于是, 可得如下定理:

定理 **3**　当 $0 < y < 1$ 时, Y_i 在 $Y_j = y$ 条件下的条件数学期望为

$$E(Y_i|Y_j) = \frac{1}{n}(1-y)。$$

证明　当 $0 < y < 1$ 时,

$$E(Y_i|Y_j) = \int_0^{1-y} x \frac{(n-1)(1-x-y)^{n-2}}{(1-y)^{n-1}}\mathrm{d}x = -\frac{1}{(1-y)^{n-1}}\int_0^{1-y} x\,\mathrm{d}(1-x-y)^{n-1} =$$

$$\frac{1}{(1-y)^{n-1}}\int_0^{1-y}(1-x-y)^{n-1}\mathrm{d}x = \frac{1}{n(1-y)^{n-1}}\cdot(1-y)^n = \frac{1}{n}(1-y)。$$

参考文献:

[1] JOHN H. On the asymptotic behaviour of samplespacing[J]. Math Proc Camb Phil Soc, 1981, 90: 293-295.

[2] 王金涛, 黄丽琨, 叶　鹰. 一种均匀分布的修正分布 UR-n 分布[J]. 应用数学, 2004, 17: 84-87.

[3] 韩金舫, 王清印, 刘金宪. 关于区间矩阵的稳定度[J]. 河北科技大学学报, 2002, 23(1): 1-4.

[4] 邓宇辉. 样本区间的概率分布[J]. 数学杂志, 2004, 24(6): 685-689.