第四节　顺序统计量

一、顺序统计量的定义

定义5.4.1　设$(X_1,X_2,\cdots,X_n)$是从总体X中抽取的一个样本, $(x_1,x_2,\cdots,x_n)$是其一个观测值,将观测值按由小到大的次序重新排列为

$$x_{(1)}\le x_{(2)}\le\cdots\le x_{(n)}$$

当$(X_1,X_2,\cdots,X_n)$取值为$(x_1,x_2,\cdots x_n)$时,定义$X_{(k)}$取值为$x_{(k)}(k=1,2,\cdots,n)$,由此得到的

$$(X_{(1)},X_{(2)},\cdots,X_{(n)})$$

称为样本$(X_1,X_2,\cdots,X_n)$的顺序统计量.对应的 $(x_{(1)},x_{(2)},\cdots,x_{(n)})$成为其观察值.

$X_{(k)}$：称为样本$(X_1, X_2, \cdots, X_n)$的第k个顺序统计量.

特别地，$X_{(1)} = \min\limits_{1 \le i \le n} X_i$ 称为最小顺序统计量.

$X_{(n)} = \max\limits_{1 \le i \le n} X_i$ 称为最大顺序统计量.

$R_{(n)} = X_{(n)} - X_{(1)}$称为样本极差.

注：由于每个$X_{(k)}$都是样本$(X_1, X_2, \cdots, X_n)$的函数，所以$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$都是随机变量. 一般它们不相互独立.

例1：设总体X的分布为仅取0, 1, 2的离散均匀分布，其分布列为

X	0	1	2
P	1/3	1/3	1/3

现从中抽取容量为3的样本，其一切可能取值有$3^3=27$种，列表如下：

X_1	X_2	X_3	X_1	X_2	X_3	X_1	X_2	X_3
0	0	0	1	0	0	2	0	0
0	0	1	1	0	1	2	0	1
0	0	2	1	0	2	2	0	2
0	1	0	1	1	0	2	1	0
0	1	1	1	1	1	2	1	1
0	1	2	1	1	2	2	1	2
0	2	0	1	2	0	2	2	0
0	2	1	1	2	1	2	2	1
0	2	2	1	2	2	2	2	2

从而可给出的$X_{(1)},X_{(2)},X_{(3)}$分布列如下：

$X_{(1)}$	0	1	2
P	$19/27$	$7/27$	$1/27$

$X_{(2)}$	0	1	2
P	$7/27$	$13/27$	$7/27$

$X_{(3)}$	0	1	2
P	$1/27$	$7/27$	$19/27$

我们可以清楚地看到这三个顺序统计量的分布是不相同的.

进一步，我们可以给出两个次序统计量的联合分布，如：$X_{(1)}$与$X_{(2)}$的联合分布列为

$X_{(1)}$ \ $X_{(2)}$	0	1	2
0	7/27	9/27	3/27
1	0	4/27	3/27
2	0	0	1/27

不难看出$\boldsymbol{X}_{(1)}$和$\boldsymbol{X}_{(2)}$是不独立的。

二、单个顺序统计量$X_{(k)}$的分布

定理5.4.1 设总体X的分布函数为$F(x)$，$X_1, X_2, \cdots, X_n$为样本，则第k个顺序统计量$X_{(k)}$的分布函数为

$$F_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}\int_0^{F(x)} t^{k-1}(1-t)^{n-k}\,dt$$

又若总体X为连续型，其密度函数为$f(x)$. 则第k个顺序统计量$X_{(k)}$的密度函数为

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!}\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k} f(x)$$

证明： 根据分布函数的定义，可以得

$$F_{X_{(k)}}(x) = P(X_{(k)} \le x) = P(X_1, \cdots, X_n\text{中至少有}k\text{个} \le x)$$

设$v_n(x)\ (-\infty < x < +\infty)$表示 $X_1, X_2, \cdots, X_n$中不超过 x 的个数. 它表示的是对总体X作n次重复独立观测时, 事件$\{X \le x\}$出现的次数, 而$P\{X \le x\} = F(x)$

故有

$$v_n(x) \sim B(n, F(x)),$$

因此

$$F_{X_{(k)}}(x) = \sum_{i=k}^{n} P\{v_n(x) = i\}$$

$$= \sum_{i=k}^{n} C_n^i [F(x)]^i [1 - F(x)]^{n-i}$$

$$\xlongequal{\text{分部积分法}} \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} t^{k-1} (1-t)^{n-k} dt$$

当总体X为连续型时，上式两端对x求导可得$X_{(k)}$的密度函数为

$$f_{X_{(k)}}(x)=\frac{n!}{(k-1)!(n-k)!}\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k}f(x)$$

注 (1) 对最大顺序统计量$X_{(n)}=\max\{X_1,\cdots,X_n\}$,

$$F_{X_{(n)}}(x)=[F(x)]^n,$$

$$f_{X_{(n)}}(x)=n[F(x)]^{n-1}f(x).$$

(2) 对最小顺序统计量$X_{(1)}=\min\{X_1,\cdots,X_n\}$,

$$F_{X_{(1)}}(x)=1-[1-F(x)]^n,$$

$$f_{X_{(1)}}(x)=n[1-F(x)]^{n-1}f(x).$$

例2：设总体X的密度函数为

$$f(x)=3x^2,\quad 0<x<1.$$

现从该总体中抽得一个容量为5的样本，试计算$P(X_{(2)}<1/2)$.

解：先求出$X_{(2)}$的分布. 总体分布函数为

$$F(x)=\begin{cases}0, & x\le 0;\\ x^3, & 0<x<1;\\ 1, & x\ge 1.\end{cases}$$

$$f_{X_{(2)}}(x)=\frac{5!}{(2-1)!(5-2)!}[F(x)]^{2-1}p(x)[1-F(x)]^{5-2}$$

$$=20\cdot x^3\cdot 3x^2\cdot(1-x^3)^3=60x^5(1-x^3)^3,\ 0<x<1$$

于是

$$P(X_{(2)}<\frac{1}{2})=\int_0^{\frac{1}{2}}60x^5(1-x^3)^3dx$$

$$\overset{y=x^3}{=}\int_0^{\frac{1}{8}}20y(1-y)^3dy=\int_{\frac{7}{8}}^{1}20(z^3-z^4)dz$$

$$=5(1-(\frac{7}{8})^4)-4(1-(\frac{7}{8})^5)=0.1207.$$

三、多个顺序统计量的联合分布（略）

定理5.4.2 在定理5.4.1的记号下，若总体X为连续型，分布函数及密度函数分别为$F(x)$，$f(x)$. 则$(X_{(i)},X_{(j)})$的密度函数为

$$f_{i,j}(x,y)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}$$

$$\times[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}f(x)f(y).$$

这里 $x<y$. 当$x>y$时. $f_{i,j}(x,y)=0$.

定理5.4.3　在定理5.4.1的记号下，顺序统计量 $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$的联合分布密度为

$$f(y_1, y_2, \cdots, y_n) = \begin{cases} n! \prod_{i=1}^{n} f(y_i), & y_1 < y_2 < \cdots < y_n. \\ 0, & \text{其他,} \end{cases}$$

定理5.4.4 设总体密度函数为$f(x)$, x_p为其p分位数，$f(x)$在x_p处连续且$f(x_p) > 0$,则当$n \to +\infty$时，样本p分位数m_p的渐近分布为

$$m_p \dot{\sim} N\left(x_p, \frac{p(1-p)}{n[f(x_p)]^2}\right).$$

特别地, 对样本中位数有$m_{0.5} \dot{\sim} N\left(x_{0.5}, \dfrac{1}{n[f(x_{0.5})]^2}\right)$.