§1.4 顺序统计量

一、顺序统计量的定义

定义1.4.1 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1.4.1 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 样本$X_1, X_2, \cdots, X_n$中不超过 x 的个数. 它表示对总体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

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

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

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

附证：

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

$$= \frac{n!}{k!(n-k)!}t^k(1-t)^{n-k}\Big|_0^{F(x)} + \frac{n!}{k!(n-k-1)!}\int_0^{F(x)} t^k(1-t)^{n-k-1}dt$$

$$= C_n^k[F(x)]^k[1-F(x)]^{n-k} + \frac{n!}{k!(n-k-1)!}\int_0^{F(x)} t^k(1-t)^{n-k-1}dt$$

最后一项为 $\dfrac{n!}{(n-1)!0!}\displaystyle\int_0^{F(x)} t^{n-1}dt = [F(x)]^n$

所以有 $F_{X_{(k)}}(x) = \displaystyle\sum_{i=k}^{n} C_n^i[F(x)]^i[1-F(x)]^{n-i}$

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

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

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

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

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

于是

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

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

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

三、多个顺序统计量的联合分布

定理1.4.2 在定理1.4.1的记号下，顺序统计量 $(X_{(i)}, X_{(j)})$ $(i < j)$的联合分布函数为

$$F_{i,j}(x, y) = \begin{cases} P\{X_{(j)} \le y\} & x \ge y; \\ P\{X_{(j)} \le y\} - P\{X_{(i)} > x, X_{(j)} \le y\} & x < y. \end{cases}$$

其中 $P\{X_{(j)} \le y\} = \dfrac{n!}{(j-1)!(n-j)!} \displaystyle\int_0^{F(y)} t^{k-1}(1-t)^{n-k}\, dt,$

$P\{X_{(i)} > x, X_{(j)} \le y\}$

$$= \sum_{0 \le k \le i\text{-}1,}\ \sum_{0 \le l \le n-j} \frac{n!}{k!(n-k-l)!l!}[F(x)]^k$$

$$\times [F(y) - F(x)]^{n-k-l}[1 - F(y)]^l.$$

又若总体X为连续型，其密度函数为$f(x)$. 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. $f_{i,j}(x,y) = 0$.

证明： 根据分布函数的定义

$$F_{i,j}(x,y) = P\{X_{(i)} \le x, X_{(j)} \le y\}$$
$$= \begin{cases} P\{X_{(j)} \le y\} & x \ge y; \\ P\{X_{(j)} \le y\} - P\{X_{(i)} > x, X_{(j)} \le y\} & x < y. \end{cases}$$

事件$\{X_{(i)} > x, X_{(j)} \le y\}$等价于$X_1, X_2, \cdots, X_n$中落入区间$(-\infty, x]$内的不超过$i-1$个，落入$(y, +\infty)$内的不超过$n-j$个. 根据多项分布的概率得

$$P\{X_{(i)} > x, X_{(j)} \le y\}$$

$$= \sum_{0 \le k \le i-1,} \sum_{0 \le l \le n-j} \frac{n!}{k!(n-k-l)!l!}[F(x)]^k \times [F(y)-F(x)]^{n-k-l}[1-F(y)]^l$$

再根据定理1.4.1，即得到$F_{i,j}(x, y)$的表达式.

当总体X为连续型时，$F_{i,j}(x, y)$对x, y求偏导得

$$\frac{\partial F_{i,j}(x, y)}{\partial x} = -\sum_{0 \le l \le n-j} \frac{n!}{(i-1)!(n-i-l)!l!}[F(x)]^{i-1} \times [F(y)-F(x)]^{n-i-l}[1-F(y)]^l f(x).$$

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

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

因而$(X_{(i)}, X_{(j)})$的密度函数为 (这里$x<y$)

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

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

注：$(X_{(1)}, X_{(n)})$的密度函数为

$$f_{1,n}(x,y) = \begin{cases} n(n-1)[F(y)-F(x)]^{n-2} f(x) f(y), x<y \\ 0, \qquad \text{其他,} \end{cases}$$

附密度函数的证明

$$F_{i,j}(x,y)=\frac{n!}{(j-1)!(n-j)!}\int_0^{F(y)} t^{k-1}(1-t)^{n-k}dt-\sum_{0\le k\le i\text{-}1,}\sum_{0\le l\le n-j}$$

$$\frac{n!}{k!(n-k-l)!l!}[F(x)]^k[F(y)-F(x)]^{n-k-l}[1-F(y)]^l$$

$$\frac{\partial F_{i,j}(x,y)}{\partial x}=-\sum_{1\le k\le i-1,}\sum_{0\le l\le n-j}\frac{n!}{(k-1)!(n-k-l)!l!}[F(x)]^{k-1}$$

$$\times[F(y)-F(x)]^{n-k-l}[1-F(y)]^l f(x)$$

$$+\sum_{0\le k\le i-1,}\sum_{0\le l\le n-j}\frac{n!}{k!(n-k-l-1)!l!}[F(x)]^k$$

$$\times[F(y)-F(x)]^{n-k-l-1}[1-F(y)]^l f(x)$$

$$
\begin{aligned}
&= -\sum_{0\le m\le i-2,\ 0\le l\le n-j} \frac{n!}{m!(n-m-l-1)!l!}[F(x)]^m \\
&\quad \times [F(y)-F(x)]^{n-m-l-1}[1-F(y)]^l f(x) \\
&\quad + \sum_{0\le k\le i-1,\ 0\le l\le n-j} \frac{n!}{k!(n-k-l-1)!l!}[F(x)]^k \\
&\quad \times [F(y)-F(x)]^{n-k-l-1}[1-F(y)]^l f(x) \\
&\xlongequal{\text{剩余}k=i-1} \sum_{0\le l\le n-j} \frac{n!}{(i-1)!(n-i-l)!l!}[F(x)]^{i-1} \\
&\quad \times [F(y)-F(x)]^{n-i-l}[1-F(y)]^l f(x)
\end{aligned}
$$

$$\frac{\partial F_{i,j}(x,y)}{\partial x \partial y} = \frac{\partial}{\partial y}\left\{\sum_{0\le l\le n-j}\frac{n!}{(i-1)!(n-i-l)!l!}[F(x)]^{i-1}\right.$$

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

$$=\sum_{0\le l\le n-j}\frac{n!}{(i-1)!(n-i-l-1)!l!}[F(x)]^{i-1}$$

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

$$-\sum_{1\le l\le n-j}\frac{n!}{(i-1)!(n-i-l)!(l-1)!}[F(x)]^{i-1}$$

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

$$
\begin{aligned}
&= \sum_{0\le l\le n-j} \frac{n!}{(i-1)!(n-i-l-1)!l!}[F(x)]^{i-1} \\
&\qquad \times [F(y)-F(x)]^{n-i-l-1}[1-F(y)]^{l} f(x)f(y) \\
&\quad - \sum_{0\le m\le n-j-1} \frac{n!}{(i-1)!(n-i-m-1)!m!}[F(x)]^{i-1} \\
&\qquad \times [F(y)-F(x)]^{n-i-m-1}[1-F(y)]^{m} f(x)f(y) \\
&\xlongequal{\text{剩余}l=n-j} \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1} \\
&\qquad \times [F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j} f(x)f(y)
\end{aligned}
$$

定理1.4.3　在定理1.4.1的记号下，顺序统计量 $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$的联合分布密度为

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

证明： 略

实际问题中用到的关于顺序统计量的函数：

(1)样本极差：$R_n = X_{(n)} - X_{(1)}$

(2)样本p分位数：

$$m_p = \begin{cases} X_{([np+1])}, & \text{若 } np \text{ 不是整数;} \\ \dfrac{1}{2}(X_{(np)} + X_{(np+1)}), & \text{若 } np \text{ 是整数.} \end{cases}$$

特别地，当$p=0.5$时，称$m_{0.5}$为样本中位数.

$$m_{0.5}=\begin{cases} X_{(\frac{n+1}{2})}, & n\text{为奇数；} \\ \frac{1}{2}(X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}), & n\text{为偶数.} \end{cases}$$

利用$(X_{(1)}, X_{(n)})$的联合分布不难得到

定理1.4.4 在定理1.4.1的记号下，顺序统计量极差$R_{(n)}=X_{(n)}-X_{(1)}$的分布密度为

$$f_{R_{(n)}}(y)=\begin{cases} n(n-1)\int_{-\infty}^{+\infty}[F(t+y)-F(t)]^{n-2}f(t+y)f(y)\mathrm{d}t, y>0, \\ 0, \qquad \text{其他.} \end{cases}$$

证明： 略

对多数总体而言，要给出样本p分位数的精确分布通常不是一件容易的事，但当$n \to +\infty$时，样本p分位数的渐近分布有比较简单的表达式, 我们这里不加证明地给出如下定理。

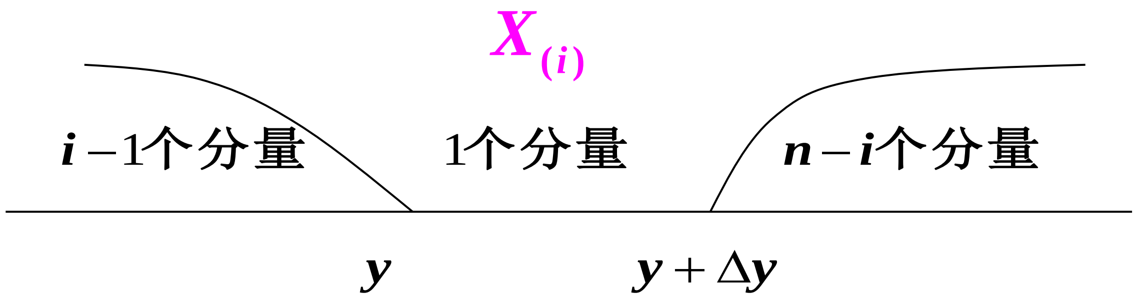
定理1.4.5 设总体密度函数为$f(x)$, x_p为其p分位数，$f(x)$在x_p处连续且$f(x)>0$,则当$n \to +\infty$时, 样本p分位数m_p的渐近分布为

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

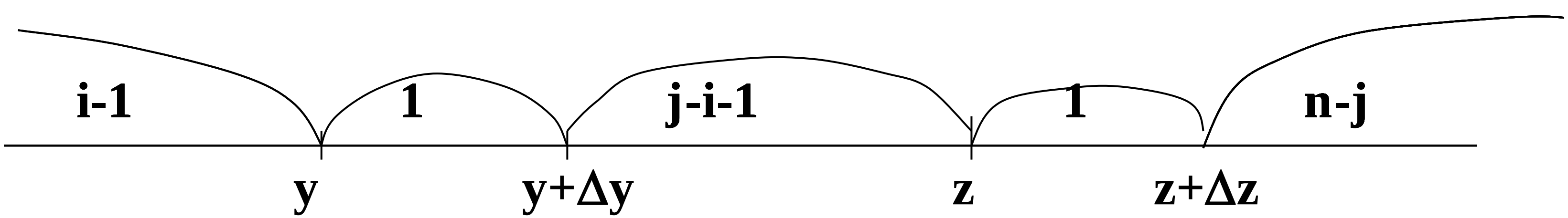
特别地，对样本中位数有$m_{0.5} \dot{\sim} N\left(x_{0.5}, \frac{1}{n[f(x_{0.5})]^2}\right)$

可以用微元密度法求$X_{(i)}$及$(X_{(i)}, X_{(j)})$的分布密度

求$X_{(i)}$的密度函数即求$X_{(i)}$落入无穷小区间$[y, y+\Delta y)$内这一事件A的概率. 利用多项分布求.



$$x_{(i)} \in [y, y+\Delta y),\ x_{(j)} \in [z, z+\Delta z), \quad y < z.$$



例3 设总体X分布为$U(0,1)$, $X_1, X_2, \cdots, X_n$为样本，则$(X_{(1)}, X_{(n)})$的联合密度函数为

$$f_{1,n}(y,z) = n(n-1)(z-y)^{n-2}，0 < y < z < 1.$$

令$R = x_{(n)} - x_{(1)}$，由$R > 0$可以推出

$$0 < X_{(1)} = X_{(n)} - R \le 1 - R$$

则

$$f_R(r) = \int_0^{1-r} n(n-1)[(y+r)-y]^{n-2} dy$$

$$= n(n-1)r^{n-2}(1-r).$$

该分布参数为$(n-1,2)$的贝塔分布.

例4 设总体X为柯西分布，其密度函数为

$$f(x;\theta)=\frac{1}{\pi(1+(x-\theta))^2},\ -\infty<x<+\infty$$

其分布函数为

$$F(x;\theta)=\frac{1}{2}+\frac{1}{\pi}\arctan(x-\theta)$$

易知，θ是该总体的中位数，即$x_{0.5}=\theta.$

$$m_{0.5}\dot{\sim}N(\theta,\frac{\pi^2}{4n}).$$

附录 多项分布

二项分布可以容易地推广到n次重复独立试验且每次试验可能有若干结果的情形. 把每次试验的可能结果记为

$$A_1, A_2, \cdots, A_r$$

记 $p_i = P(A_i) \geq 0,\ \ i = 1, 2, \cdots, r.$ 且$p_1 + p_2 + \cdots + p_r = 1.$

在这种推广的伯努利试验中，不难推得：在n 次试验中事件A_1出现k_1次，A_2出现k_2 次，$\cdots$，A_r出现k_r次的概率为

$$\frac{n!}{k_1!k_2!\cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

这里 $k_1 + k_2 + \cdots + k_r = n$, $k_i \geq 0$. 上述公式称为多项分布. 它是$(p_1 + p_2 + \cdots + p_r)^n$的展开式的一般项,

而且显然有

$$\sum_{\substack{k_1+\cdots+k_r=n \\ k_i \geq 0}} \frac{n!}{k_1!k_2!\cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} = 1.$$