第五章 数理统计的基本概念
§5.3 次序统计量及其分布

2023年 3月

2 3 4 2 4 1 1 3 2

设$X_1, X_2, \cdots, X_n$是来自总体$X \sim F(x)$的一个样本. 对每次抽样的可能的观测值$x_1, x_2, \cdots, x_n$, 将它们从小到大排列为

$X_{(i)}$

$$\boxed{x_{(1)}, x_{(2)}, \cdots, x_{(n)}.}$$

$X_{(1)}$

1	2	3
1	2	4
2	3	4

定义5.3.1

第i个次序统计量$X_{(i)}$是上述样本$X_1, X_2, \cdots, X_n$的一个函数, 不论样本$X_1, X_2, \cdots, X_n$取怎样的观测值, 它总是取其中的$x_{(i)}$为观测值.

显然，

$$X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}$$

我们称它为最大次序统计量；在第三章我们已经知道它的分布函数为

总体 $X \sim F(x)$

$$\begin{aligned} F_{\max}(x) &= P(X_{(n)} \le x) \\ &= P(X_1 \le x, X_2 \le x, \cdots, X_n \le x) \\ &= (F(x))^n. \end{aligned}$$

$F_n(x)$

三.(12分)设$\xi_1,\xi_2,\cdots,\xi_n,\cdots$为独立同分布的随机变量序列，共同的密度函数为

$$p_\xi(x)=\begin{cases}\frac{1}{\beta}, & 0<x<\beta;\\ 0, & \text{其他};\end{cases}$$

其中$\beta>0$为常数. 证明:

$$\eta_n=\max(\xi_1,\xi_2,\cdots,\xi_n)\xrightarrow{P}\beta,\quad n\to\infty.$$

类似地,

$$X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}.$$

我们称它为最小次序统计量, 它的分布函数为

$$\begin{aligned} F_{\min}(x) &= P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x, X_2 > x, \cdots, X_n > x) \\ &= 1 - (1 - F(x))^n. \end{aligned}$$

$X \sim \boxed{U(0,1)}$

例5.3.1

（P40例3.16）设$X_1, X_2, \cdots, X_n$是n个独立同分布的随机变量，都服从均匀分布$U[0,1]$，试求$\xi = \max\{X_1, X_2, \cdots, X_n\} = X_{(n)}$

和$\eta = \min\{X_1, X_2, \cdots, X_n\}$ 的分布.

$X_{(1)}$

例5.3.1

（*P40例3.16*）设$X_1, X_2, \cdots, X_n$是n个独立同分布的随机变量，都服从均匀分布$U[0,1]$，试求$\xi = \max\{X_1, X_2, \cdots, X_n\}$和$\eta = \min\{X_1, X_2, \cdots, X_n\}$ 的分布.

【解】$X_1, X_2, \cdots, X_n$共同的分布函数为

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

于是ξ的分布函数为

$$F_\xi(x) = F_{\max}(x) = \boxed{(F(x))^n} = \begin{cases} 0, & x < 0 \\ x^n, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}.$$

从而ξ的密度函数为

$$p_\xi(x) = \begin{cases} nx^{n-1}, & 0 < x < 1 \\ 0, & \text{其它} \end{cases}$$

于是ξ的分布函数为

$$\begin{aligned} F_\xi(x) &= F_{\max}(x) = (F(x))^n \\ &= \begin{cases} 0, & x < 0 \\ x^n, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}. \end{aligned}$$

从而ξ的密度函数为

$$p_\xi(x) = \begin{cases} nx^{n-1}, & 0 < x < 1 \\ 0, & \text{其它} \end{cases}$$

即$\xi \sim Be(n,1)$.

$X_{(n)} \sim Be(n,1)$

$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$

$= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

【复习】

若随机变量ξ 的密度函数为

$$p(x) = \begin{cases} \boxed{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}} x^{a-1}(1-x)^{b-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

$a = n$

$b = 1$

则称ξ 服从贝塔分布，记作$\xi \sim Be(a,b)$，其中$a > 0, b > 0$ 都是形状参数.

$\int_0^1 p(x)\,dx = 1$

$n x^{n-1}$

$Be(n,1)$

同理η的分布函数为

$$\begin{aligned} F_\eta(x) &= F_{\min}(x) = 1-(1-F(x))^n \\ &= \begin{cases} 0, & x<0 \\ 1-(1-x)^n, & 0\le x<1 \\ 1, & x\ge 1 \end{cases}. \end{aligned}$$

从而η的密度函数为

$$p_\eta(x) = \begin{cases} n(1-x)^{n-1}, & 0<x<1 \\ 0, & \text{其它} \end{cases}$$

$X_{(1)} \sim Be(1,n)$

同理η的分布函数为

$$
\begin{aligned}
F_\eta(x) &= F_{\min}(x) = 1-(1-F(x))^n \\
&= \begin{cases} 0, & x<0 \\ 1-(1-x)^n, & 0\le x<1 \\ 1, & x\ge 1 \end{cases}.
\end{aligned}
$$

从而η的密度函数为

$$
p_\eta(x) = \begin{cases} n(1-x)^{n-1}, & 0<x<1 \\ 0, & \text{其它} \end{cases}
$$

即$\eta \sim Be(1,n)$.

$X_{(n)}$ $X_{(1)}$

定理5.3.1

设总体X的分布函数为$F(x)$, 且有密度函数$f(x) > 0, a \le x \le b$. 设$X_1, X_2, \cdots, X_n$是来自该总体的一个样本, 则第i个次序统计量$X_{(i)}$的密度函数为

Tail

$$g_i(y) = \frac{n!}{(i-1)!(n-i)!} \times [F(y)]^{i-1}[1-F(y)]^{n-i} f(y), a < y < b,$$

$$= 0, \qquad else.$$

【证明】 我们用所谓的“微元密度法”来证明这一结论. 任意给定$a<y<b$, 将$(-\infty,\infty)$ 分成三个部分:

$(-\infty,y]$, $(y,y+\Delta y]$, $(y+\Delta y,\infty)$.

则$P(X_{(i)}\in(y,y+\Delta y])\doteq P($ 恰有$i-1$个在第一个区间,1个第二个区间, $n-i$ 个在第三个区间$)$

$$\begin{aligned} = \quad & C_n^{i-1}C_{n-i+1}^1[F(y)]^{i-1}\left(F(y+\Delta y)-F(y)\right) \\ & [1-F(y+\Delta y)]^{n-i} \end{aligned}$$

$$\begin{aligned} = \quad & \frac{n!}{(i-1)!(n-i)!}[F(y)]^{i-1}\left(F(y+\Delta y)-F(y)\right) \\ & [1-F(y+\Delta y)]^{n-i}. \end{aligned}$$

从而

$$
\begin{aligned}
& g_i(y) \\
= \quad & \lim_{\Delta y \to 0} \frac{P(X_{(i)} \in (y, y+\Delta y])}{\Delta y} \\
= \quad & C_n^{i-1} C_{n-i+1}^1 [F(y)]^{i-1} \\
& \lim_{\Delta y \to 0} \frac{(F(y+\Delta y) - F(y))}{\Delta y} [1 - F(y+\Delta y)]^{n-i} \\
= \quad & \frac{n!}{(i-1)!(n-i)!} [F(y)]^{i-1} [1 - F(y)]^{n-i} f(y).
\end{aligned}
$$

例5.3.2

设总体X有密度函数 $n=4$ $i=3$

$$f(x)=\begin{cases} 2x, & 0<x<1 \\ 0, & \text{其它}. \end{cases}$$

$F(x)$

X_1, X_2, X_3, X_4是来自该总体的一个容量为4的样本，求第三个次序统计量$X_{(3)}$的密度函数$\boxed{g_3(x)}$和分布函数$G_3(x)$，并计算概率$P(X_{(3)} > \frac{1}{2})$.

例5.3.2

设总体X有密度函数

$$f(x)=\begin{cases}2x, & 0<x<1\\ 0, & 其它.\end{cases}$$

X_1,X_2,X_3,X_4是来自该总体的一个容量为4的样本，求第三个次序统计量$X_{(3)}$的密度函数$g_3(x)$和分布函数$G_3(x)$，并计算概率$P(X_{(3)}>\frac{1}{2})$.

【解】总体X的分布函数为

$$F(x)=\begin{cases}0, & x<0\\ x^2, & 0\le x<1\\ 1, & x\ge 1\end{cases}$$

于是当$0 < x < 1$时，$X_{(3)}$的密度函数为

$$g_3(x) = \frac{4!}{(3-1)!(4-3)!}[F(x)]^2[1-F(x)]^{4-3}f(x)$$

于是当$0 < x < 1$时，$X_{(3)}$的密度函数为

$$
\begin{aligned}
g_3(x) &= \frac{4!}{(3-1)!(4-3)!}[F(x)]^2[1-F(x)]^{4-3}f(x) \\
&= \frac{4!}{2!}[x^2]^2[1-x^2]\cdot 2x \\
&= 24x^5(1-x^2),
\end{aligned}
$$

其他情形时，$g_3(x) = 0$.

分布函数为

$$G_3(x) = \begin{cases} 0, & x < 0 \\ x^6(4 - 3x^2), & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

分布函数为

$$G_3(x)=\begin{cases}0, & x<0\\ x^6(4-3x^2), & 0\le x<1\\ 1, & x\ge 1\end{cases}$$

$$P(X_{(3)}>\frac{1}{2}) \;=\; 1-G_3(\frac{1}{2})$$

分布函数为

$$G_3(x) = \begin{cases} 0, & x < 0 \\ x^6(4-3x^2), & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

$$\begin{aligned} P(X_{(3)} > \frac{1}{2}) &= 1 - G_3(\frac{1}{2}) \\ &= 1 - \frac{1}{2^6}[4 - 3(\frac{1}{2})^2] = \frac{243}{256}. \end{aligned}$$

思考

(1)设总体X服从$(0,1)$上的均匀分布，求次序统计量$\boxed{X_{(i)}}$, $i=1,2,\cdots,n$的分布.

(2) P260 Ex 5.31.

$$g_i(x) = \begin{cases} \dfrac{n!}{(i-1)!\,(n-i)!} x^{i-1} (1-x)^{n-i}, & 0<x<1 \\ 0 & \text{else} \end{cases}$$

$$\sim Be(i, n-i+1)$$

思考

(1)设总体X服从$(0,1)$上的均匀分布，求次序统计量$X_{(i)}, i=1,2,\cdots,n$的分布.

(2) P260 Ex 5.31.

【解】 $X_{(i)} \sim Be(i, n-i+1), i=1,2,\cdots,n$.

2022复旦大学432统计学试题

8. (10分) $X_1, X_2, \ldots X_6 \overset{i.id.}{\sim} U(0,1)$,

求 $\mathrm{Var}\left(2X_{(2)} + 3X_{(3)}\right)$.

$\frac{61}{98}$

问题：你能计算出$(X_{(i)}, X_{(j)})$的联合分布吗?

定理5.3.2

设总体X的分布函数为$F(x)$, 且有密度函数$f(x)>0, a\le x\le b$. 设$X_1, X_2, \cdots, X_n$是来自该总体的一个样本, 则第i个和第j个次序统计量$X_{(i)}$和$X_{(j)}$ $(i<j)$的联合密度函数为

$$
\begin{aligned}
g_{ij}(y,z) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(y)]^{i-1} \\
&\quad \times [F(z)-F(y)]^{j-i-1}[\overline{F}(z)]^{n-j} f(y)f(z), \\
&\qquad\qquad a<y<z<b, \\
&= 0, \qquad\qquad else.
\end{aligned}
$$

【证明】我们继续用“微元密度法”来证明这一结论. 我们只要讨论$a < y < z < b$的情形:

【证明】我们继续用"微元密度法"来证明这一结论. 我们只要讨论$a < y < z < b$的情形:
将$(-\infty, \infty)$ 分成五个部分:
$(-\infty, y)$, $(y, y + \Delta y]$, $(y + \Delta y, z]$,$(z, z + \Delta z]$,
$(z + \Delta z, \infty)$.

则

$P(X_{(i)} \in (y, y+\Delta y], X_{(j)} \in (z, z+\Delta z])$

$\doteq$ **P(** 恰有$i-1$个在第一个区间,1个第二个区间,
$j-i-1$个在第三个区间,1 个第四个区间,$n-j$个在第五个区间**)**

则

$P(X_{(i)} \in (y, y+\Delta y], X_{(j)} \in (z, z+\Delta z])$

$\doteq$ **P(恰有$i-1$个在第一个区间,1个第二个区间,$j-i-1$个在第三个区间,1 个第四个区间,$n-j$个在第五个区间)**

$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \times [F(y)]^{i-1}(F(y+\Delta y) - F(y))[F(z) - F(y+\Delta y)]^{j-i-i}(F(z+\Delta z) - F(z))[1 - F(z+\Delta z)]^{n-j}.$

定理5.3.3

设$X_1, X_2, \cdots, X_n$是独立同分布的连续型随机变量，有共同的密度函数$f(x) > 0, a \le x \le b$. 则n个次序统计量$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$的联合密度函数为

$$g(y_1, y_2, \cdots, y_n) = \begin{cases} n! \prod_{k=1}^{n} f(y_k), & a < y_1 < y_2 < \cdots < y_n < b \\ 0, & \text{其他}. \end{cases}$$

【*证明】用微元密度法. 对任意$y_1 < y_2 < \cdots < y_n$及任意充分小的$\Delta y_1, \Delta y_2, \cdots, \Delta y_n$, 我们有

$$P(\bigcap_{k=1}^{n}(y_k < X_{(k)} \le y_k + \Delta y_k)$$

【*证明】用微元密度法. 对任意$y_1 < y_2 < \cdots < y_n$及任意充分小的$\Delta y_1, \Delta y_2, \cdots, \Delta y_n$，我们有

$$
\begin{aligned}
& P\left(\bigcap_{k=1}^{n}(y_k < X_{(k)} \le y_k + \Delta y_k)\right) \\
= \ & n! \prod_{k=1}^{n}(F(y_k + \Delta y_k) - F(y_k))
\end{aligned}
$$

【*证明】用微元密度法. 对任意$y_1 < y_2 < \cdots < y_n$及任意充分小的$\Delta y_1, \Delta y_2, \cdots, \Delta y_n$，我们有

$$\begin{aligned} & P(\bigcap_{k=1}^{n}(y_k < X_{(k)} \le y_k + \Delta y_k) \\ = \ & n! \prod_{k=1}^{n}(F(y_k + \Delta y_k) - F(y_k)) \end{aligned}$$

于是我们得到所要证明的结果.

例5.3.3

设总体X服从$(0,t)$上的均匀分布，求其n个次序统计量$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$的联合密度函数.

例5.3.3

设总体X服从$(0,t)$上的均匀分布，求其n个次序统计量$(X_{(1)},X_{(2)},\cdots,X_{(n)})$的联合密度函数.

【解】由定理5.3.3知$(X_{(1)},X_{(2)},\cdots,X_{(n)})$**的联合密度函数为**

$$g(y_1,y_2,\cdots,y_n)=\begin{cases}\frac{n!}{t^n}, & 0<y_1<y_2<\cdots<y_n<t\\ 0, & \text{其他}.\end{cases}$$

【补充】下面介绍一下几个常用的次序统计量的函数:

1. 质量管理中常用的样本极差

$$R_n = X_{(n)} - X_{(1)};$$

优点：极差含有总体标准差的信息。因为极差表示样本取值范围的大小，也反映总体取值分散与集中的程度。一般说来，若总体的标准差σ较大（小），从中取出的样本的极差也会大（小）一些。反过来也如此，若样本极差较大（小），表明总体取值较分散（集中），那么相应总体的标准差也较大（小）。

缺点：极差受样本量影响较大。在实际中极差常在小样本$(n \leqslant 10)$的场合使用，而在大样本场合很少使用。这是因为极差仅使用了样本中两个极端点的信息，而把中间的信息都丢弃了，当样本容量越大时，丢充的信息也就越多，从而留下的信息过少，其使用价值就不大了。

2. 样本中位数:

设$X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 是容量为n 的样本的次序统计量，则称如下统计量

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

为该样本中位数。

例　一批砖在交付客户之前要抽检其抗压强度（单位：Mpa)，现从中随机抽取10 块砖，测得其抗压强度为:

6.0	6.5	7.3	4.7	5.8
9.0	10.1	17.2	7.7	4.2

其样本中位数

$$m_{0.5} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{7.3 + 7.7}{2} = 7.5$$

$$\bar{x} = 8.21$$

后经复查发现，样本中的异常值17.2 属抄录之误，原始记录为11.2，把17.2 改正为11.2 后，样本中位数不变，仍为7.5。可样本均值$\bar{x}$ 在修正前后分别为8. 21 与7. 61 ，两者相差0.6 。可见，当样本中出现异常值(是指样本中的个别值，它明显偏离其余观察值）时，样本中位数比样本均值更具有抗击异常值干扰的能力。样本中位数的这种抗干扰性在统计学中称为稳健性。

样本中位数是总体中位数的影子，常用来估计总体中位数$x_{0.5}$(概率方程$F\left(x_{0.5}\right)=0.5$ 的解)，且样本量越大，效果越好。

3. **样本p 分位数**

设$X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$ 是容量为n 的样本的次序统计量，对给定的$p(0<p<1)$, 称

$$m_p = \begin{cases} \frac{1}{2}\left[X_{(np)} + X_{(np+1)}\right], & np \text{ 是整数} \\ X_{([np]+1)}, & np \text{ 不是整数} \end{cases}$$

为该样本的样本p 分位数，其中$[np]$ 为np 的整数部分。样本p分位数m_p是总体p分位数x_p (概率方程$F\left(x_p\right)=p$ 的解）的估计量。

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。幸运的是，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 设总体密度函数为$f(x), x_p$ 为其p 分位数，若$f(x)$ 在x_p 处连续，且$f\left(x_p\right)>0$, 则当$n \rightarrow \infty$ 时，样本p 分位数m_p 的渐近分布为:

$$m_p \sim AN\left(x_p, \frac{p(1-p)}{n \cdot f^2\left(x_p\right)}\right).$$

特别地，对样本中位数，当$n \rightarrow \infty$ 时近似地有

$$m_{0,5} \sim AN\left(x_{0,5}, \frac{1}{4n \cdot f^2\left(x_{0,5}\right)}\right).$$

作业:
P259
5.24, 5.28