次序统计量

次序统计量: 即若 $X_1, X_2, \cdots, X_n$ i.i.d. $\sim F$, 将其按大小排列为 $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$,则称 $\left(X_{(1)}, X_{(2)}, \cdots, X_{(n)}\right)$ 为次序统计量, 它的任一部分, 如 $X_{(i)}$, 和 $\left(X_{(i)}, X_{(j)}\right)(1 \leq i<j \leq n)$ 等也称为次序统计量.

样本 $X_1, \ldots, X_n$ 的秩次统计量由序列 $X_{(1)}, \ldots, X_{(n)}$ 给出，其中这些数值按升序排列，$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$。特别地，第一个和最后一个秩次统计量分别等于

$$X_{(1)}=\min _{1 \leq i \leq n} X_i \quad \text { and } \quad X_{(n)}=\max _{1 \leq i \leq n} X_i$$

对于来自分布 F 的给定样本 $X_1, \ldots, X_n$ 的第 i 个秩次统计量 $X_{(i)}$，我们有 $\mathrm{E}\, F\left(X_{(i)}\right)=i /(n+1)$（见描述性统计习题 8）。因此，我们可以预期点集 $\left\{\left(i /(n+1), F\left(x_{(i)}\right)\right): i=1, \ldots, n\right\}$ 在 xy 平面上大致位于直线 $y=x$ 上。对于点集

$$\left\{\left(F^{-1}\left(\frac{i}{n+1}\right), x_{(i)}\right): i=1, \ldots, n\right\}$$

同样可以做出这样的预期。更一般地，如果样本 $x_1, \ldots, x_n$ 来自位置-尺度族的元素 $F_{a, b}$，那么我们预期上述点集位于直线 $y=a+b x$ 上；毕竟我们有 $x_{(i)} \approx F_{a, b}^{-1}(i /(n+1))=a+b F^{-1}(i /(n+1))$。这里的对照是把一个位置-尺度族中的次序统计量和该族的"标准分布"的分位值做对照.

次序统计量的分布

1. $X_{(n)}$ 的分布

$$P\left(X_{(n)} \leq x\right)=P\left(X_{1} \leq x, \cdots, X_{n} \leq x\right)=F^{n}(x)$$

2. $X_{(1)}$ 的分布

$$P\left(X_{(1)} \leq x\right)=1-P\left(X_{(1)}>x\right)=1-P\left(X_{1}>x, \cdots, X_{n}>x\right)=1-(1-F(x))^{n}$$

3. $X_{(m)}$ 的分布 $(1<m<n)$

$$\begin{aligned}
f_{X_{(m)}}(x) d x & \approx P\left(x<X_{(m)} \leq x+d x\right) \\
& =\frac{n!}{(m-1)!1!(n-m)!} F^{m-1}(x) f(x) d x[1-F(x+d x)]^{n-m}
\end{aligned}$$

因此两边同时除以 dx, 并令 $dx \to 0$, 得到

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

4. $X_{(i)}, X_{(j)}$ 的联合密度

f_{i j}(x, y)= \begin{cases}\frac{n!}{(i-1)!(j-i-1)!(n-j)!}(F(x))^{i-1}(F(y)-F(x))^{j-i-1} & \ \quad \times(1-F(y))^{n-j} f(x) f(y), & \text { 当 } x<y, \ 0, & \text { 其它. }\end{cases}

5. $\left(X_{(1)}, \cdots, X_{(n)}\right)$的联合密度为

g\left(x*{(1)}, \cdots, x*{(n)}\right)= \begin{cases}n!f\left(x*{(1)}\right) \cdots f\left(x*{(n)}\right), & \text { 当 } x*{(1)}<x*{(2)}<\cdots<x_{(n)}, \ 0, & \text { 其它. }\end{cases}

6.极差$X_{(n)}-X_{(1)}$的分布(== 需要时候看 ==)作下列变换

\left{\begin{array} { l }

{ V = X *{ (j) } - X*{ (i) } } \

{ Z = X _ { (i) } }

\end{array} \Longleftrightarrow \left{\begin{array}{l}

X_{(i)}=Z \

X_{(j)}=V+Z

\end{array}\right.\right.

变换的$Jacobian$行列式为 : $|J|=\left|\frac{\partial\left(X_{(i)}, X_{(j)}\right)}{\partial V \partial Z}\right|=1,\left(X_{(i)}, X_{(j)}\right)$的联合分布密度由前给出, 故$(V, Z$

g_{i j}(v, z)= \begin{cases}\frac{n!}{(i-1)!(j-i-1)!(n-j)!}(F(z))^{i-1}(F(v+z)-F(z))^{j-i-1} & \tag{2.1}\ \quad \times(1-F(v+z))^{n-j} f(z) f(v+z), & \text { 当 } v>0, \ 0, & \text { 其它. }\end{cases}

从而易知V的密度为

f*{V}(v)=\int*{-\infty}^{\infty} g_{i, j}(v, z) d z

特别, 当取$i=1, j=n$得到$\left(R, X_{(1)}\right)$的联合密度

g_{1, n}(v, z)= \begin{cases}\frac{n!}{(n-2)!}(F(v+z)-F(z))^{n-2} f(v+z) f(z), & \text { 当 } v>0, \tag{2.2}\ 0, & \text { 当 } v \leq 0 .\end{cases}

R的边缘密度为$\int_{-\infty}^{\infty} g_{1,n}(v,z)dz$. 当样本来自均匀分布$U(\theta_1,\theta_2)$时，极差$R=X_{(n)}-X_{(1)}$的分

f_V (v) = n (n - 1) \cdot \frac{v^{n - 2} (\lambda - v)}{\lambda^{n + 2}}, \quad 0 < v < \lambda.

因此，极差$R=X_{(n)}-X_{(1)}$的具体概率密度函数为：

f_R (r) = n (n - 1) \cdot \frac{(\theta_2 - \theta_1 - r) \cdot r^{n - 2}}{(\theta_2 - \theta_1)^{n + 2}}, \quad 0 < r < \theta_2 - \theta_1.

这个结果表明，极差R的分布取决于样本容量n和区间长度$\theta_2-\theta_1$。密度函数在区间$(0,\theta_2-\theta_1$