- **定义3.1** 设$X_1,\ldots,X_n$ $i.i.d.$，将其按数值大小排列为$X_{(1)}\leq X_{(2)}\leq\ldots\leq X_{(n)}$，则$X_{(1)},X_{(2)},\ldots,X_{(n)}$ 称为样本$X_1,\ldots,X_n$的**次序统计量**(Order Statistics)。
- 利用次序统计量可以定义下列统计量

(1) **样本中位数**(Sample Median)

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

注1 $m_{\frac{1}{2}}$反映总体中位数的信息。

注2 当总体分布的p.d.f.对称时，对称中心既是总体中位数，又是总体均值，此时$m_{\frac{1}{2}}$也反映总体均值的信息。

(2) **样本极值**(Extremum of Sample)

$X_{(1)}$和$X_{(n)}$称为样本的极小值和极大值，他们统称为样本极值。

注 极值统计量常用于灾害问题和材料试验的统计分析中。

(3) **样本p分位数**$(0<p<1)$ (Sample p-fractile)

$$X_{(m)},\quad m=[(n+1)p].\quad \text{这里}[\cdot]\text{表示实数}\cdot\text{的整数部分}.$$

注1 其反映总体p分位数信息。

注2 当$p=\frac{1}{2}$，n为奇数时，样本p分位数即为样本中位数。

注3 常用：四分位数，包括低四分位数（第25个百分位数）/下四分位数和高四分位数（第75个百分位数）/上四分位数。

(4) **样本极差**(Sample Range)

$$R=X_{(n)}-X_{(1)}.$$

注 它反映总体分布的散布程度的信息。

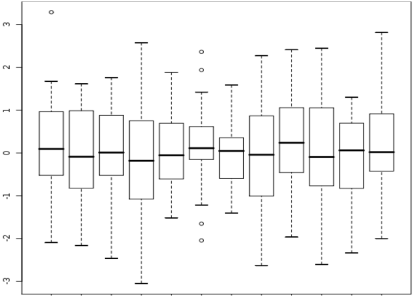
(4) **样本中程数**(Sample Midrange)

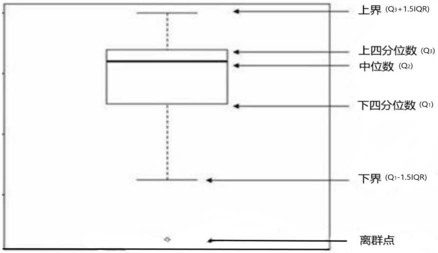
$$mR=\frac{1}{2}\left(X_{(1)}+X_{(n)}\right).$$

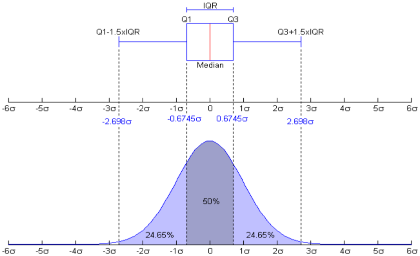
注 它是反映数据集中趋势的一项指标。

应用1*:箱型图

箱型图(Box plot)
--- 利用图形方式显示位置度量（中位数），散度度量（四分位之差）和可能出现的离群点，同时还表明分布的对称性或偏度状态。参考【4】p278。







- 定义3.2[【0】定义1.3.2] 设$X_1,\cdots,X_n$是从累积分布函数（c.d.f.）为F 的总体中抽取的简单样本，对任意实数x，称下列函数

$$\hat{F}_n(x)=\begin{cases}0, & x<X_{(1)}\\ \frac{k}{n}, & X_{(k)}\le x<X_{(k+1)}, k=1,\ldots,n-1\\ 1, & x\ge X_{(n)}\end{cases}$$

为经验分布函数（Empirical distribution function）.

注 $\hat{F}_n(x)=\frac{1}{n}$ "小于等于x的样本个数" $=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{(-\infty,x]}(X_i)$.

- 基本性质
  - 在$x=X_{(k)}, k=1,\ldots,n$处有间断，跳跃幅度$\frac{1}{n}$的阶梯函数；
  - 具有累积分布函数的基本要素：有界、单调非降、右连续；
  - $n\hat{F}_n(x)\sim Binomial(n,F(x))$.
  - 对任意给定x, $\hat{F}_n(x)\xrightarrow{P}F(x)$, as $n\to\infty$.
  - Glivenko-Cantelli定理：$\sup_x|\hat{F}_n(x)-F(x)|\xrightarrow{P}0$, as $n\to\infty$.

Theorem (3.1)

[【1】定理5.4.3]（离散型） 设$X_1, \dots, X_n$ $\quad i.i.d.$ 取自概率质量函数(p.m.f.) $\mathbb{P}(X = x_i) = p_i$ 的离散型总体X，其中$x_1 < x_2 < \cdots$ 是X的所有可能的取值。定义

$$P_0 = 0; \quad P_1 = p_1; \quad P_2 = p_1 + p_2; \quad \cdots\cdots$$
$$P_i = p_1 + \cdots + p_i; \quad \cdots\cdots$$

以$X_{(1)}, \dots, X_{(n)}$ 为样本$X_1, \dots, X_n$的次序统计量，则

$$\mathbb{P}(X_{(j)} \le x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k},$$

且$\mathbb{P}(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k} \right].$

- **定理3.1的证明** 见【1】.

注 概率质量函数(p.m.f.) 的证明可利用【0】习题2.6证得

$$\mathbb{P}(X_{(j)} = x_i) = j\binom{n}{j}\int_{P_{i-1}}^{P_i} t^{j-1}(1-t)^{n-j}dt.$$

Theorem (3.2)

[【1】定理5.4.4] （连续型） 设简单随机样本$X_1, \ldots, X_n$ 取自累积分布函数*(c.d.f.)*为F，概率密度函数*(p.d.f.)* 为f 的连续型总体, $X_{(1)}, \ldots, X_{(n)}$ 为其次序统计量，则$X_{(j)}$的概率密度函数*(p.d.f.)* 为

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

- **证明** 见【1】或【0】p28. 与定理3.1的证明相似。

- n个次序统计量$X_{(1)},\cdots,X_{(n)}$的联合密度函数

$$g(y_1,y_2,\cdots,y_n)=\begin{cases} n!f(y_1)f(y_2)\cdots f(y_n), & y_1<y_2<\cdots<y_n \\ 0, & \text{其它} \end{cases}$$

- 两个次序统计量的联合分布

Theorem (3.3)

[【0】定理2.3.1，【1】定理5.4.6] 设$X_1,\ldots,X_n$ 取自总体c.d.f.为F，p.d.f.为f的简单样本，则$(X_1,\ldots,X_n)$ 的次序统计量中任意两个$(X_{(i)},X_{(j)}),i<j$的联合密度函数为，

$$f_{ij}(x,y)=\begin{cases} n!\dfrac{[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}}{(i-1)!(j-i-1)!(n-j)!}f(x)f(y), & x<y \\ 0, & \text{其它} \end{cases}$$

- 证明见【0】.

问题 样本极差$R = X_{(n)} - X_{(1)}$，考虑其在连续情形下的概率密度函数（p.d.f.）g_R？（自行练习求解离散情形时的概率质量函数p.m.f.）

- 先考虑一般情形，求$V = X_{(j)} - X_{(i)}, 1 \le i < j \le n$的p.d.f.？
- 作变换

$$\begin{cases} V = & X_{(j)} - X_{(i)} \\ Z = & X_{(i)} \end{cases} \Longleftrightarrow \begin{cases} X_{(i)} = & Z \\ X_{(j)} = & V + Z \end{cases}$$

Jacobi行列式$J = |\frac{\partial(X_{(i)}, X_{(j)})}{\partial V \partial Z}| = 1.$

- 回忆本节定理3.3，由$(X_{(i)}, X_{(j)})$的联合p.d.f. 可得(V, Z)的联合p.d.f.为

$$g_{i,j}(v, z) = f_{i,j}(z, v+z)|J| = f_{i,j}(z, v+z), \quad v > 0.$$

从而可得V的p.d.f.为 $g_V(v) = \int g_{i,j}(v, z) dz.$

- 特别地，当$i=1, j=n$时，(R, Z)的联合p.d.f.为

$$g_{1,n}(r,z)=\begin{cases} n(n-1)[F(r+z)-F(z)]^{n-2}f(r+z)f(z), & r>0 \\ 0, & r\leq 0 \end{cases}$$

- 而R的(边缘)p.d.f.为

$$g_R(r)=\int g_{1,n}(r,z)dz.$$

Example (3.1 – 参考【0】2.3.4节)

设$X_1,\ldots,X_n \quad i.i.d. \sim U(0,1)$, 求

1. $X_{(i)}, i=1,\ldots,n$的p.d.f. ?
2. $X_{(1)},\cdots,X(n)$的联合p.d.f. ?
3. $(X_{(i)},X_{(j)}), 1\leq i<j\leq n$的联合p.d.f. ?
4. 极差R的p.d.f. ?

作业

习题2: Ex. 10, 27.