非参数检验包括:

(一)定性检验法:

正态概率纸检验

(二)数值检验法:

皮尔逊(Pearson) χ^2 拟合检验

柯尔莫哥洛夫检验

斯米尔诺夫检验

Shapiro-Wilk W 检验与Agostino's D 检验（不讲）

Wilcoxon秩和检验

(补) 偏度峰度检验.

三、柯尔莫哥洛夫检验

检验原理: 通过样本的经验分布函数$F_n(x)$与理论分布函数$F_0(x)$的比较, 推断该样本是否来自$F_0(x)$对应的总体.

(一) 参数已知的单总体检验

设总体X的分布函数F(x)未知, $F_0(x)$是一个完全已知的连续型分布函数. 利用样本$X_1,\cdots,X_n$检验假设:

$$\boxed{H_0: F(x)=F_0(x)}$$

1. 构造检验统计量:

(1) 柯尔莫哥洛夫统计量的一般形式:

$$\boxed{D_n=\sup_{-\infty<x<\infty}|F_n(x)-F_0(x)|}.$$

D_n的精确分布见教材P112;

$\sqrt{n}\,D_n$的极限分布见教材P112.

(2) 柯尔莫哥洛夫统计量的化简形式:

先将样本$X_1,\cdots,X_n$从小到大排列成(重复数据合并为一个)

$$X_{(1)}<\cdots<X_{(m)},\ (1\le m\le n)$$

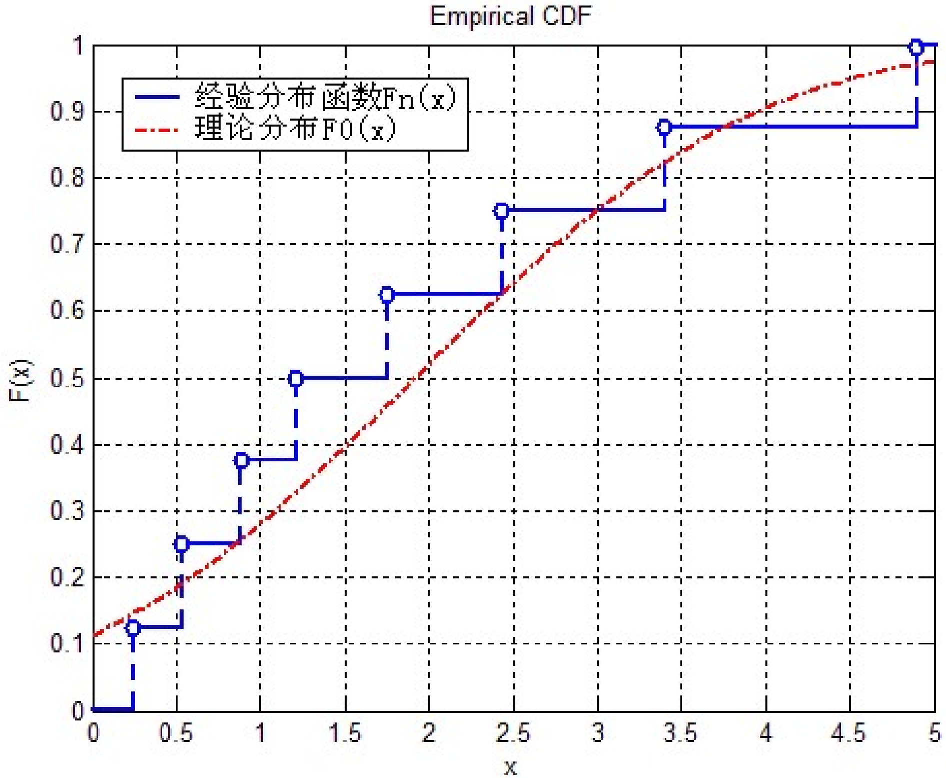
设n_i为$X_{(i)}$在样本中出现的频数, 则有

$$F_n(X_{(i)})=\frac{n_1+\cdots+n_{i-1}}{n},\quad i=1,2,\cdots,m$$

令 $$\boxed{d_i=\max\left\{\left|F_0(X_{(i)})-F_n(X_{(i)})\right|,\left|F_n(X_{(i+1)})-F_0(X_{(i)})\right|\right\},i=1,2,\cdots,m}\tag{3.28}$$

其中, 规定$F_n(X_{(m+1)})=1$. 此时,

柯尔莫哥洛夫统计量为: $$\boxed{D_n=\max\{d_1,d_2,\cdots,d_m\}}\tag{3.29}$$



2. H_0的拒绝域:

$$\boxed{D_n > D_{n,\alpha}}.$$

当 $n \le 100$ 时, $D_{n,\alpha}$ 查P420附表6;

当 $n > 100$ 时, $D_{n,\alpha} \approx \lambda_{1-\alpha} / \sqrt{n}$, 其中 $\lambda_{1-\alpha}$ 查P422附表7 $(Q(\lambda_{1-\alpha}) = 1-\alpha)$.

【★例3.18(P_{114})】对一台设备进行寿命试验，记录10次无故障工作时间，并从小到大排列得

420，500，920，1380，1510，1650，1760，2100，2300，2350

问此设备无故障工作时间X是否服从$\theta=1500$的指数分布($\alpha=0.05$)?

3. 柯尔莫哥洛夫检验的优缺点:

(1) 优点: 当总体为一维且理论分布完全已知时，柯尔莫哥洛夫检验优于Pearson χ^2检验.

(2) 缺点: 柯尔莫哥洛夫检验的适用范围不如χ^2检验广. 特别当理论分布含有未知参数时，目前只对正态分布和指数分布及I型极值分布作出了结果.

(二) 参数未知的单总体的检验

一) 正态性检验(Lilliefors)

利用样本$X_1,\cdots,X_n$检验有关总体X分布的假设:

$$H_0: X \sim N(\mu,\sigma^2),\qquad \text{实际是}\ \boxed{H_0: X \sim N(\hat{\mu},\hat{\sigma}^2)}.$$

其中 $\hat{\mu}=\bar{X},\ \hat{\sigma}^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 是无偏估计量.

1. 构造检验统计量:

$$\boxed{\hat{D}_n=\sup_{-\infty<x<\infty}\left|F_n(x)-F_0(x;\hat{\mu},\hat{\sigma}^2)\right|}.$$

2. H_0的拒绝域:

$$\boxed{\hat{D}_n>\hat{D}_{n,\alpha}}.$$

$\hat{D}_{n,\alpha}$ 查P_{423}附表8.

【★例3.19(P_{116})】对8个产品进行强度试验, 所得数据取自然对数后为

0.25, 0.53, 0.88, 1.22, 1.76, 2.44, 3.41, 4.90

问这批强度数据是否来自对数正态分布($\alpha = 0.2$)?

【★例3.20(P_{117})】在20天内, 从维尼纶正常生产时生产报表上看到的维尼纶纤度(表示纤维粗细程度的量)的情况, 有如下100个数据:

1.36,	1.49,	1.43,	1.41,	1.37,	1.40,	1.32,	1.42,	1.47,	1.39
1.41,	1.36,	1.40,	1.34,	1.42,	1.42,	1.45,	1.35,	1.42,	1.39
1.44,	1.42,	1.39,	1.42,	1.42,	1.30,	1.34,	1.42,	1.37,	1.36
1.37,	1.34,	1.37,	1.37,	1.44,	1.45,	1.32,	1.48,	1.40,	1.45
1.39,	1.46,	1.39,	1.53,	1.36,	1.48,	1.40,	1.39,	1.38,	1.40
1.36,	1.45,	1.50,	1.43,	1.38,	1.43,	1.41,	1.48,	1.39,	1.45
1.37,	1.37,	1.39,	1.45,	1.31,	1.41,	1.44,	1.44,	1.42,	1.47
1.35,	1.36,	1.39,	1.40,	1.38,	1.35,	1.42,	1.43,	1.42,	1.42
1.42,	1.40,	1.41,	1.37,	1.46,	1.36,	1.37,	1.27,	1.37,	1.38
1.42,	1.34,	1.43,	1.42,	1.41,	1.41,	1.44,	1.48,	1.55,	1.37

要求在显著性水平$\alpha = 0.01$下检验假设

$$H_0: X \sim N(\mu, \sigma^2).$$

二) 指数分布的检验(Finklestein & Schafer)

利用样本$X_1,\cdots,X_n$检验有关总体X分布的假设:

$$H_0: F(x) = 1 - e^{-x/\theta},\ x > 0.$$

实际是 $\boxed{H_0: F(x) = 1 - e^{-x/\hat{\theta}},\ x > 0}$.

其中$\hat{\theta} = \bar{X}$是极大似然估计量,也是无偏估计量.

1. 构造检验统计量:

检验统计量1: $\boxed{D_n^* = \sup_{0\le x<\infty} |F_n(x) - F_0(x;\hat{\theta})| = \max_{1\le i\le n} d_i}$;

检验统计量2(Finklestein & Schafer): $\boxed{S_n^* = \sum_{i=1}^{n} d_i}$.

2. H_0的拒绝域:

$$\boxed{S_n^* > S_{n,\alpha}^*}.$$

$S_{n,\alpha}^*$查P_{424}附表9.

【例3.21(P_{120})】记录一台计算机的无故障工作时间七次, 数据如下:

530, 450, 120, 530, 600, 650, 460

能否认为此台计算机的无故障工作时间X服从指数分布$(\alpha = 0.05)$?

四、斯米尔诺夫检验

检验原理: 类似柯尔莫哥洛夫检验的原理.

设 $X_1, X_2, \cdots, X_{n_1}$ 是从分布函数为F(x)的**连续型**总体中抽取的样本, $Y_1, Y_2, \cdots, Y_{n_2}$ 是从分布函数为 G(x)的**连续型**总体中抽取的样本, 且F(x)与G(x)未知, 并假定这两个样本相互独立. 要检验假设:

$$\boxed{H_0 : F(x) = G(x); \quad H_1 : F(x) \neq G(x)}$$

1. 构造检验统计量:

设 $F_{n_1}(x)$和$G_{n_2}(x)$分别是这两个样本所对应的经验分布函数.

$$\boxed{D_{n_1, n_2} = \sup_{-\infty < x < \infty} |F_{n_1}(x) - G_{n_2}(x)|}$$

且$\sqrt{n} D_{n_1, n_2}$ 的极限分布见教材P121.

2. H_0的拒绝域:

$$\boxed{D_{n_1,n_2} > D_{n,\alpha}}$$

当 $n=\left[\dfrac{n_1 n_2}{n_1+n_2}\right]\le 100$时, $D_{n,\alpha}$查P420附表6;

当 $n=\left[\dfrac{n_1 n_2}{n_1+n_2}\right]> 100$时, $D_{n,\alpha}\approx \lambda_{1-\alpha}/\sqrt{n}$, 其中 $\lambda_{1-\alpha}$查P422附表7.

说明5: 斯米尔诺夫检验的困难在于求统计量D_{n_1,n_2}的观测值.

【例3.22(P122)】某自动车床加工一种零件, 一位工人刚接班时, 抽取n_1=150只零件作为第一个样本. 在自动车床工作了四小时后, 他又抽了n_2=100只零件作为第二个样本. 测定每个零件的尺寸与标准尺寸的偏差 (单位: μm) 范围如下表所示:

偏差范围	组中值	组频数n_{1i}	组频数n_{2i}
[-12.5, -7.5)	-10	10	0
[-7.5, -2.5)	-5	27	7
[-2.5, 2.5)	0	43	17
[2.5, 7.5)	5	38	30
[7.5, 12.5)	10	23	29
[12.5, 17.5)	15	8	15
[17.5, 22.5)	20	1	1
[22.5, 27.5)	25	0	1

试问在显著性水平$\alpha = 0.01$下, 能否认为这批零件尺寸的分布相同?

五、Shapiro-Wilk W 检验和D'Agostino D检验（不讲）

W检验和D检验都是正态检验, 用于检验一批随机数是否来自同一正态分布. 检验问题为

H_0: 总体服从正态分布; H_1: 总体不服从正态分布

(一) W检验 $(3 \le n \le 50)$

1. 构造检验统计量:

(1) 先将观测值按非降次序排列成: $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$

(2) 检验统计量:

$$W = \frac{\left\{\sum_{k=1}^{[\frac{n}{2}]} a_k(W)[X_{(n+1-k)} - X_{(k)}]\right\}^2}{\sum_{k=1}^{n}(X_{(k)} - \bar{X})^2} \tag{3.32}$$

2. H_0的拒绝域:

$$\boxed{W < W_\alpha}$$

其中, W_α 查P_{427}附表11.

【例3.23(P_{125})】抽查用克矽平治疗的矽肺患者10名, 得他们治疗前后血红蛋白的差(g%)如下:

2.7, -1.2, -1.0, 0, 0.7, 2.0, 3.7, -0.6, 0.8, -0.3

试检验治疗前后血红蛋白的差是否服从正态分布($\alpha = 0.05$).

(二) D检验($50 \le n \le 1000$)

1. 构造检验统计量:

(1) 先将观测值按非降次序排列成:

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

(2) 检验统计量

$$D=\frac{\sum_{k=1}^{n}\left(k-\frac{n+1}{2}\right)X_{(k)}}{(\sqrt{n})^3\sqrt{\sum_{k=1}^{n}(X_{(k)}-\bar{X})^2}},$$

$$\boxed{Y=\frac{D-0.282}{0.02999/\sqrt{n}}\ \underset{H_0\text{为真}}{\overset{n\to\infty}{\sim\sim\sim\sim}}\ N(0,1)}$$

2. H_0的拒绝域:

$$\boxed{Y<Z_{\alpha/2}\ \text{或}\ Y>Z_{1-\alpha/2}}$$

其中, $Z_{\alpha/2}, Z_{1-\alpha/2}$查$P_{428}$附表12.

【例3.24(P_{127})】上海中心气象台独立测定的上海市九十九年(1884~1982) 的年降雨量的数据如下(单位: mm):

1184.4,	1113.4,	1203.9,	1170.7,	975.4,	1462.3,	947.8,	1416.0
709.2,	1147.5,	935.0,	1016.3,	1031.6,	1105.7,	849.9,	1233.4
1008.6,	1063.8,	1004.9,	1086.2,	1022.5,	1330.9,	1439.4,	1236.5
1088.1,	1288.7,	1115.8,	1217.5,	1320.7,	1078.1,	1203.4,	1480.0
1269.9,	1049.2,	1318.4,	1192.0,	1016.0,	1508.2,	1159.6,	1021.3
986.1,	794.7,	1318.3,	1171.2,	1161.7,	791.2,	1143.8,	1602.0
951.4,	1003.2,	840.4,	1061.4,	958.0,	1025.2,	1265.0,	1196.5
1120.7,	1659.3,	942.7,	1123.9,	910.2,	1398.5,	1208.6,	1305.5
1242.3,	1572.3,	1416.9,	1256.1,	1285.9,	984.8,	1390.3,	1062.2
1287.3,	1477.0,	1017.9,	1217.7,	1197.1,	1143.0,	1018.8,	1243.7
909.3,	1030.3,	1124.4,	811.4,	820.9,	1184.1,	1107.5,	991.4
901.7,	1176.5,	1113.5,	1272.9,	1200.3,	1508.7,	772.3,	813.0
1392.3,	1006.2,	1108.8					

试问年降雨量是否服从正态分布(α=0.05).

六、秩和 (威尔柯克斯 – Wilcoxon) 检验

检验原理: 通过将混合排序后两个样本的秩和进行比较, 推断这两个样本是否来自相同的分布.

设 $X_1,X_2,\cdots,X_{n_1}$ 是从分布函数为F(x)的连续型总体中抽取的样本, $Y_1,Y_2,\cdots,Y_{n_2}$ 是从分布函数为G(x)的连续型总体中抽取的样本, 且F(x)与G(x)未知, 并假定这两个样本相互独立. 要检验假设:

$$\boxed{H_0: F(x)=G(x);\quad H_1: F(x)\neq G(x).}$$

1. 构造检验统计量:

(1) 将 $X_1,\cdots,X_{n_1}$ 和 $Y_1,\cdots,Y_{n_2}$ 混合排列:

$$Z_1\le Z_2\le\cdots\le Z_{n_1+n_2}$$

如果 $X_k=Z_i$, 则记 $r(X_k)=i$, 称为 X_k 在混合样本中的秩.

$$T_1 = \sum_{k=1}^{n_1} r(X_k) \text{——} X_1, \cdots, X_{n_1} \text{的秩和}$$

$$T_2 = \sum_{k=1}^{n_2} r(Y_k) \text{——} Y_1, \cdots, Y_{n_2} \text{的秩和}$$

(2) 检验统计量

$$\boxed{T = \begin{cases} T_1, & n_1 \le n_2 \\ T_2, & n_1 > n_2 \end{cases}}$$

2. H_0的拒绝域:

$$\boxed{T < T_\alpha^{(1)} \text{ 或 } T > T_\alpha^{(2)}}$$

其中, $T_\alpha^{(1)}, T_\alpha^{(2)}$通过查“秩和检验表”得到.

【★例3.25(P_{128})】以下是两个地区所种小麦的蛋白质含量检验数据:

地区1: 12.6, 13.4, 11.9, 12.8, 13.0

地区2: 13.1, 13.4, 12.8, 13.5, 13.3, 12.7, 12.4

问两地区小麦的蛋白质含量有无显著性差异($\alpha = 0.05$)?

3. 当 $n_1, n_2 \geq 10$ 时的近似计算:

当 $n_2 \geq n_1$ 时, $ET \xlongequal{H_0\text{成立}} \dfrac{n_1(n_1+n_2+1)}{2}$, $DT \xlongequal{H_0\text{成立}} \dfrac{n_1 n_2(n_1+n_2+1)}{12}$, 则

(1)检验统计量: $\boxed{u=\dfrac{T-ET}{\sqrt{DT}} \overset{n_2 \geq n_1 > 7}{\underset{H_0\text{为真}}{\sim\sim\sim\sim\sim\sim\sim}} N(0,1)}$.

(2) H_0的拒绝域: $\boxed{|u| > u_{1-\alpha/2}}$.

【例3.26(P_{129})】甲、乙两人分析同一气体的CO_2含量, 数据如下:

甲: 14.6, 15.1, 15.4, 14.7, 15.2, 14.7, 14.8, 14.6, 15.2, 15.0,
14.6, 14.6, 14.8, 15.3, 14.7, 14.6, 14.8, 14.9, 15.2, 15.0

乙: 14.7, 15.0, 15.2, 14.8, 15.5, 14.6, 14.9, 14.8, 15.1, 15.0,
14.7, 14.8, 14.7, 15.0, 14.9, 14.9, 15.2, 14.7, 15.4, 15.3

问两人分析结果有无显著差异($\alpha = 0.05$)?

七、偏度、峰度检验 (用于检验总体的正态性)

1. 总体X的偏度、峰度定义:

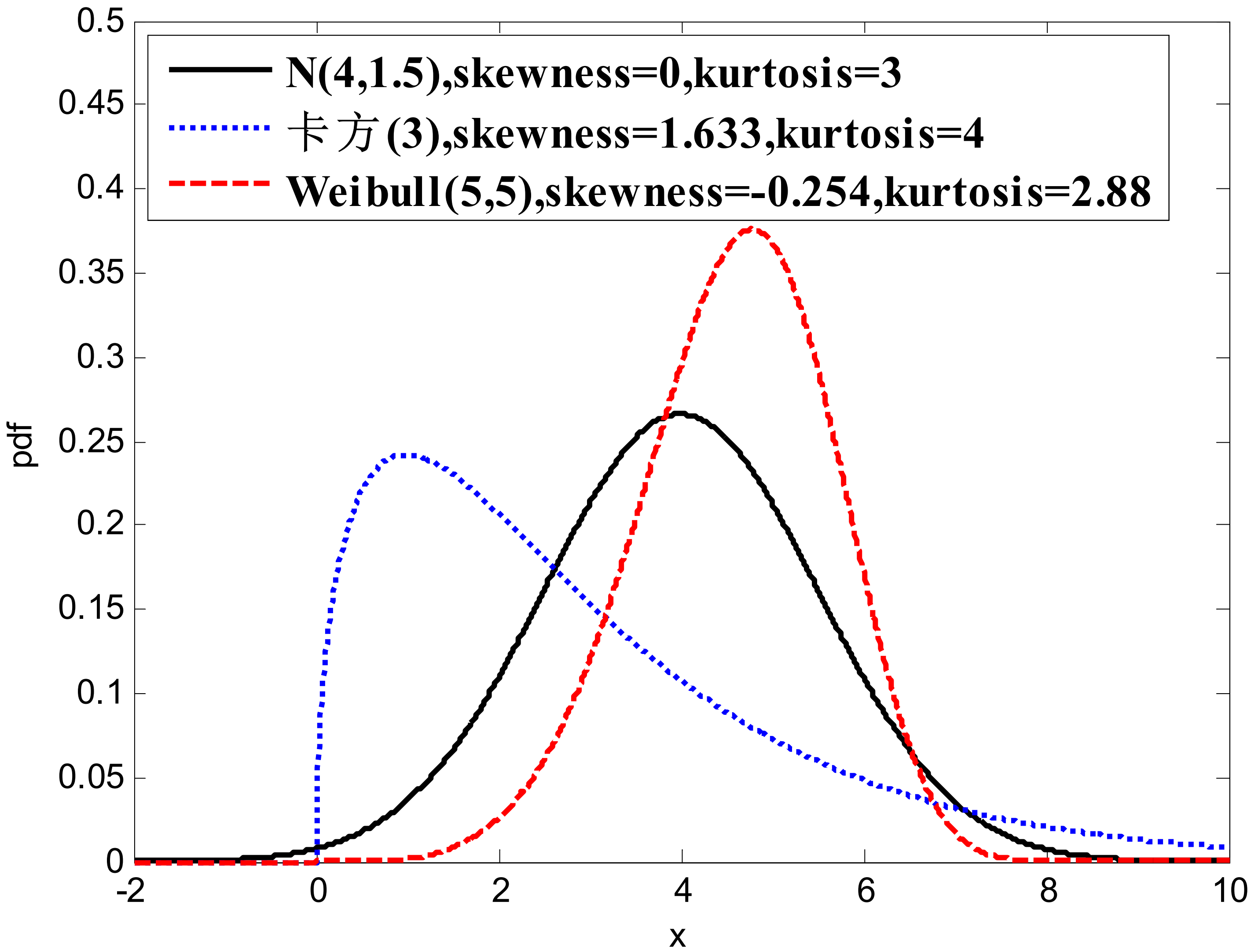
$$v_1 = E\left(\frac{X - E(X)}{\sqrt{D(X)}}\right)^3 \quad \text{—X的偏度}$$

$$v_2 = E\left(\frac{X - E(X)}{\sqrt{D(X)}}\right)^4 \quad \text{—X的峰度}$$

偏度(skewness) ——表征X的概率密度曲线相对于均值不对称程度的特征数. >0称为右偏态, <0称为左偏态.

峰度(kurtosis) ——表征X标准化后的密度函数在0周围的峰峭性. >3表示$\dfrac{X - E(X)}{\sqrt{D(X)}}$的密度函数比$N(0,1)$更尖峭, <3表示比$N(0,1)$更平坦.

特别地, 当$X \sim N(\mu,\sigma^2)$时, $v_1 = 0, v_2 = 3$.


N(4,1.5),skewness=0,kurtosis=3
卡方(3),skewness=1.633,kurtosis=4
Weibull(5,5),skewness=-0.254,kurtosis=2.88
pdf
x
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-2
0
2
4
6
8
10

2. X的样本偏度与样本峰度定义:

$$G_1=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{S}\right)^3$$ —样本偏度

$$G_2=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{S}\right)^4$$ —样本峰度

其中, $S=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$.

3. 样本偏度与样本峰度的结论:

(1) $\hat{v}_{1(\text{矩})}=G_1, \hat{v}_{2(\text{矩})}=G_2$

(2) 当 $X\sim N(\mu,\sigma^2)$时,

$$\boxed{G_1 \overset{n\to\infty}{\sim\sim\sim} N\left(0,\frac{6(n-2)}{(n+1)(n+3)}\right)\triangleq N(0,\sigma_1^2)}$$

$$\boxed{G_2 \overset{n\to\infty}{\sim\sim\sim} N\left(3-\frac{6}{n+1}, \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}\right) \triangleq N(\mu_2, \sigma_2^2)}$$

4. 检验过程:

(1) 关于总体的假设:

H_0: 总体X是正态总体.

(2) 构造检验统计量:

$$\boxed{U_1 = \frac{G_1}{\sigma_1} \underset{H_0为真}{\overset{n\to\infty}{\sim\sim\sim\sim\sim}} N(0,1),\ U_2 = \frac{G_2-\mu_2}{\sigma_2} \underset{H_0为真}{\overset{n\to\infty}{\sim\sim\sim\sim\sim}} N(0,1)}$$

(3) H_0的拒绝域:

$$\boxed{|u_1| \ge u_{1-\alpha/4} \quad 或 \quad |u_2| \ge u_{1-\alpha/4}}$$

注意3:偏度、峰度检验法要求$n>100$.

小结:

<table>
<tr><th>检验法</th><th>功能</th><th>总体X类型</th><th>总体参数</th><th>优点</th><th>缺点</th></tr>
<tr><td>正态概率纸检验</td><td>判断单总体 $X\sim N(\mu,\sigma^2)$</td><td>一维,连续型</td><td>未知</td><td>粗略估计总体的某些数字特征</td><td>定性分析而非定量分析</td></tr>
<tr><td rowspan="2">皮尔逊 χ^2 拟合检验</td><td>检验单总体 $X\sim F_0(x;\theta)$</td><td>一维或多维,离散或连续型</td><td>已知或未知</td><td rowspan="2">利用观测频数与理论频数的差异构造检验统计量.
可用于全样本,也可用于截尾样本,还可用于成群数据.</td><td rowspan="2">由于分组处理样本的观测值,因而易犯第II类错误.</td></tr>
<tr><td>检验两个总体的独立性</td><td>离散型</td><td>未知</td></tr>
<tr><td rowspan="2">柯尔莫哥洛夫检验</td><td rowspan="2">检验单总体 $X\sim F_0(x;\theta)$</td><td>一维,连续型</td><td>已知</td><td rowspan="2">利用经验分布函数与理论分布函数的差异构造检验统计量.
与Pearson χ^2 检验相比,当总体为一维且理论分布 $F_0(x;\theta)$ 完全已知时,该检验法优于 χ^2 检验.</td><td rowspan="2">柯尔莫哥洛夫检验的适用范围不如 χ^2 检验广.且当理论分布含未知参数时,目前只对正态、指数、I型极值分布给出了结果.</td></tr>
<tr><td>正态,指数,I型极值</td><td>未知</td></tr>
</table>

偏度峰度检验	检验单总体 $X \sim N(\mu,\sigma^2)$	一维,连续型	未知	适用于$n \geq 100$	
W 检验	检验单总体 $X \sim N(\mu,\sigma^2)$	一维,连续型	未知	适用于小样本$(3 \leq n \leq 50)$	
D检验	检验单总体 $X \sim N(\mu,\sigma^2)$	一维,连续型	未知	适用于大样本$(50 \leq n \leq 1000)$	
斯米尔诺夫检验	检验两个总体是否同分布	一维,连续型	未知	采用与柯尔莫哥洛夫检验类似的方法,借助经验分布函数构造检验统计量	
秩和检验	检验两个总体是否同分布	一维,连续型	未知	利用样本混合排序后的秩和构造检验统计量,方法简单	只利用了样本数据的排序,而没有利用样本数据本身

说明:上表中的所有内容仅来自于吴翊等编写的《应用数理统计》与盛骤等编写的《概率论与数理统计》.

作业: (P131-134) 3.4, 3.6, 3.12, 3.13, 3.15, 3.19, 3.21, 3.22

问题:

1、斯米尔诺夫检验与秩和检验均是检验两个总体是否同分布，谁的效果更好? (斯米尔诺夫检验通常效果更好)

2、D检验与偏度峰度检验都是针对大样本的正态检验，谁的效果更好?

求证: 在秩和检验中, 当 $H_0: F(x)=G(x)$ 成立且 $n_2 \geq n_1$ 时, 有 $ET=\dfrac{n_1(n_1+n_2+1)}{2}$, $DT=\dfrac{n_1 n_2(n_1+n_2+1)}{12}$.

证明: 设 X_i 的秩次为 $rank(X_i)$, 则 $rank(X_i)$ 不独立但同 $[1,2,\cdots,n_1+n_2]$ 上的离散型均匀分布, 且

$$E[rank(X_i)]=\frac{n_1+n_2+1}{2},\quad D[rank(X_i)]=\frac{(n_1+n_2)^2-1}{12}$$

又 $T=\sum\limits_{i=1}^{n_1} rank(X_i)$, 则

(1) 由期望的性质有:

$$ET=E\left[\sum_{i=1}^{n_1} rank(X_i)\right]=\sum_{i=1}^{n_1} E\left[rank(X_i)\right]=\frac{n_1(n_1+n_2+1)}{2}.$$

(2) 由方差的性质有:

$$DT=D\left[\sum_{i=1}^{n_1} rank(X_i)\right]=\sum_{i=1}^{n_1} D\left[rank(X_i)\right]+\sum_{i\neq j} Cov\left[rank(X_i),rank(X_j)\right]\triangleq A+B$$

$$A=n_1\cdot\frac{(n_1+n_2)^2-1}{12}$$

$$\begin{aligned}
B&=(n_1^2-n_1)Cov\left[rank(X_i),rank(X_j)\right]\\
&=(n_1^2-n_1)\left\{E\left[rank(X_i)\cdot rank(X_j)\right]-E\left[rank(X_i)\right]E\left[rank(X_j)\right]\right\}\\
&=(n_1^2-n_1)\left\{\frac{[1+2+\cdots+(n_1+n_2)][1+2+\cdots+(n_1+n_2)]-[1^2+2^2+\cdots+(n_1+n_2)^2]}{A_{n_1+n_2}^2}-\left(\frac{n_1+n_2+1}{2}\right)^2\right\}\\
&=(n_1^2-n_1)\left\{\frac{\left[\dfrac{(n_1+n_2)(n_1+n_2+1)}{2}\right]^2-\dfrac{1}{6}(n_1+n_2)(n_1+n_2+1)(2n_1+2n_2+1)}{(n_1+n_2)(n_1+n_2-1)}-\left(\frac{n_1+n_2+1}{2}\right)^2\right\}
\end{aligned}$$

化简得 $DT=A+B=\dfrac{n_1 n_2(n_1+n_2+1)}{12}$.