非参数统计期末答辩

–两总体均值检验的功效比较问题

黄启岳 欧阳乾弘

导言

题目：

对两样本t检验和Wilcoxon秩和检验在均匀分布、正态分布、双指数分布、Logistic分布以及Cauchy分布下的检验功效进行模拟比较，再与两样本自助检验作比较。

报告内容：

1. 概述理论计算渐进相对效的过程，确定模拟应当控制的变量。(约3min)
2. 加入自助检验，利用蒙特卡罗方法模拟检验功效，分析结果。(约4min)

两样本t检验

针对总体$X:\{X_1,...,X_m\}$与$Y:\{Y_1,...,Y_n\}$，我们有以下假设：

$$H_0:\mu_X=\mu_Y \leftrightarrow H_1:\mu_X\neq\mu_Y$$

构造t检验统计量,若X与Y均服从正态分布，总体方差未知但相等，则原假设成立的条件下：

$$T=\frac{\overline{X}-\overline{Y}}{\sqrt{(\frac{1}{m}+\frac{1}{n})\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}}\sim t(m+n-2)$$

利用T值的大小判断总体X与Y之间的距离，以此检验两个总体X与Y的均值是否相等，T值过大则拒绝H_0。

两样本Wilcoxon秩和检验

针对总体$X:\{X_1,...,X_m\}$独立同分布，分布为$F(X)$，$Y:\{Y_1,...,Y_n\}$独立同分布，分布为$G(Y)$，$F(X)=G(X-\Delta)$,我们有以下假设：

$$H_0:\Delta=0 \ \leftrightarrow \ H_1:\Delta\neq 0$$

在H_0假设下，X与Y为独立且同分布的序列，通过将两总体混合得到序列$\{Z_a,a=1,...,m+n\}$，产生对应秩：$R=(Q_1,\cdots,Q_m;R_1,\cdots,R_n)$

构造Wilcoxon统计量

$$W^+=\sum_{a=1}^{n}R_a$$

R_a为Y_a的秩，W^+为Y的秩和。备择假设成立时W^+会较小(对应$Y<X$)或较大(对应$Y>X$)。双边置信α拒绝域可以写作：

$$\left[\omega(\frac{\alpha}{2},m,n),\omega(1-\frac{\alpha}{2},m,n)\right]$$

两样本Bootstrap检验

首先将两总体X与Y的均值修正到相等，即

$$\widetilde{X}_i = X_i - \bar{X} + \bar{Z}, \widetilde{Y}_i = Y_i - \bar{Y} + \bar{Z}$$

之后进行自助抽样，获得自助样本$\{\widetilde{X_i^b}\}, \{\widetilde{Y_i^b}\}$，一共进行B次，针对每一组$\{\widetilde{X_i^b}\}, \{\widetilde{Y_i^b}\}$计算：$T_b = \frac{\overline{X_b} - \overline{Y_b}}{\sqrt{\frac{S^2_{X_b}}{m} + \frac{S^2_{Y_b}}{n}}}$，并基于原样本计算：

$$T_{obs} = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S^2_X}{m} + \frac{S^2_Y}{n}}}$$

最后计算p值，p值为：

$$p = \frac{1}{B} \sum_{i=1}^{B} \mathbf{I}_{\{T_b > T_{obs}\}}$$

若p值小于给定的显著性水平α，则拒绝原假设H_0。

Theorem (两样本Wilcoxon秩和统计量渐进分布)

两样本情况下，样本$\{X_{1,m_1},\cdots,X_{m_i,m_i}\}$相互独立同分布，分布函数为连续函数$F(X)$;样本$\{Y_{1,n_1},\cdots,Y_{n_i,n_i}\}$相互独立同分布，分布函数为连续函数$F(X-\theta)$。检验：

$$H_0:\theta=0 \leftrightarrow H_1:\theta=\theta_i=\frac{c}{\sqrt{N_i}}, N_i=m_i+n_i$$

若有$\frac{m_i}{N_i}\to\lambda$，则：

$\sqrt{N_i}\left[\frac{1}{m_in_i}\sum_{l=1}^{m}\sum_{j=1}^{n}\psi(Y_{l,n_i}-X_{j,m_i})-P_{\theta_i}(X_{l,m_i}<Y_{l,n_i})\right]\to$
$N\left(0,\frac{1}{12\lambda(1-\lambda)}\right)$。

其中

$$\frac{1}{m_in_i}\sum_{l=1}^{m}\sum_{j=1}^{n}\psi(Y_{l,n_i}-X_{j,m_i})$$

为$\theta=P(X<Y)$的U统计量。

Theorem (Noether定理)

对于假设检验问题

$$H_0: \theta = \theta_0 \quad H_1: \theta = \theta_i$$

其中

$$\theta_i \neq \theta_0, \lim_{i \to \infty} \theta_i = \theta_0$$

S_{n_i}, T_{m_i}为两个不同的检验统计量，若存在数列$\{\mu_{S_{n_i}}(\theta)\}, \{\mu_{T_{m_i}}(\theta)\}, \{\sigma_{S_{n_i}}(\theta)\}, \{\sigma_{T_{m_i}}(\theta)\}$满足：

(1)$\frac{S_{n_i} - \mu_{S_{n_i}}(\theta)}{\sigma_{S_{n_i}}(\theta)}$与$\frac{T_{m_i} - \mu_{T_{m_i}}(\theta)}{\sigma_{T_{m_i}}(\theta)}$在$\theta = \theta_i$为真时拥有相同的取值范围的相同的连续型的极限分布。

*(2)(1)*中结论对$\theta = \theta_0$也成立。

*(3)*标准差在θ_0处连续，即：

$$\lim_{i \to \infty} \frac{\sigma_{S_{n_i}}(\theta_i)}{\sigma_{S_{n_i}}(\theta_0)} = 1, \lim_{i \to \infty} \frac{\sigma_{T_{m_i}}(\theta_i)}{\sigma_{T_{m_i}}(\theta_0)} = 1$$

Theorem (Noether定理)

(4)$\frac{\mathrm{d}}{\mathrm{d}\theta}\mu_{S_{n_i}}(\theta), \frac{\mathrm{d}}{\mathrm{d}\theta}\mu_{T_{m_i}}(\theta)$存在，且在$\theta = \theta_0$的某一个闭区域内连续。且$\mu_{S_{n_i}}(\theta), \mu_{T_{m_i}}(\theta)$在$\theta_0$处导数不为0。

(5)μ在θ_0处光滑，即：

$$\lim_{i \to \infty} \frac{\mu'_{S_n}(\theta_i)}{\mu'_{S_{n_i}}(\theta_0)} = \lim_{i \to \infty} \frac{\mu'_{T_{m_i}}(\theta_i)}{\mu' a_{T_{m_i}}(\theta_0)} = 1$$

(6)

$$\lim_{n \to \infty} \frac{\mu'_{S_{n_i}}(\theta_0)}{\sqrt{n\sigma^2_{S_n}(\theta_0)}} = \kappa_S, \lim_{n \to \infty} \frac{\mu'_{T_m}(\theta_0)}{\sqrt{n\sigma^2_{T_m}(\theta_0)}} = \kappa_T$$

则：κ_S与κ_T为S与T的效力因子，且：

$$ARE(S, T) = \left(\frac{\kappa_S}{\kappa_T}\right)^2$$

首先对于假设检验，设计$\{X_1, \cdots, X_m\}$与$\{Y_1, \cdots, Y_n\}$分别独立同分布服从$F(X)$和$F(X-\theta)$,$F(t)$连续且关于原点对称，且存在密度函数$f(t)$，除了可数点外连续，并且在0处连续。假设为：

$$H_0 : \theta = 0 \leftrightarrow H_1 : \theta = \theta_i = \frac{c}{\sqrt{N_i}}, N_i = m_i + n_i, n_i \to +\infty$$

这样的假设有利于求解秩和统计量的渐进分布。

t检验的效力因子

针对t检验，由t分布的性质可知其渐进服从正态分布。设定$S_X = S_Y \to \sigma$,则:

$$
\begin{aligned}
\mu_{T_n}(\theta_i) &= E(T) \\
&= \frac{E(\overline{Y} - \overline{X})}{\sqrt{\left(\frac{m_i+n_i}{m_i n_i}\right)\frac{(m_i-1)\sigma^2+(n_i-1)\sigma^2}{m_i+n_i-2}}} \\
&= \frac{\sqrt{N_i}\theta_i}{\sqrt{\frac{N_i^2}{n_i m_i}\sigma^2}} \\
&= \frac{\sqrt{N_i}\theta_i}{\sqrt{\frac{1}{\lambda(1-\lambda)}\sigma^2}} \\
&= \sqrt{\lambda(1-\lambda)}\frac{\sqrt{N_i}}{\sigma}\theta_i
\end{aligned}
$$

显然T_n经过标准化，因此：

$$\sigma_{T_n}(\theta_i) = 1$$

因此：

$$\frac{T_n - \mu_{T_n}(\theta_i)}{\sigma_{T_n}(\theta_i)} \to N(0,1)$$

容易验证Noether定理的前五个条件满足。

$$\begin{aligned} eff(T) &= \frac{\frac{\mathbf{d}}{\mathbf{d}\theta_i}\sqrt{\lambda(1-\lambda)}\frac{\sqrt{N_i}}{\sigma}\theta_i}{\sqrt{N_i}} \\ &= \frac{\sqrt{\lambda(1-\lambda)}}{\sigma} \end{aligned}$$

可以看出eff(T)与两总体的样本容量以及这两个总体的方差都有关系，样本容量相等时功效相对最大，方差越大则功效越低。

Wilcoxon秩和检验的效力因子

为了得出Wilcoxon秩和的渐进分布，将其改写成U统计量的形式。

Wilcoxon秩和可以写作：

$$W_n^+ = \sum_{i=1}^{n_i} \sum_{j=1}^{m_i} \psi(Y_{l,n_i} - X_{j,m_i})$$

容易看出Y_{l,n_i}和X_{j,m_i}均为1可估，由两样本U统计量的定义，

$P(X < Y)$的U统计量为：

$$\begin{aligned} V_n^+ &= \frac{1}{\binom{m_1}{1}\binom{n_i}{1}} \sum_{i=1}^{n_i} \sum_{j=1}^{m_i} \psi(Y_{l,n_i} - X_{j,m_i}) \\ &= \frac{1}{m_i n_i} \sum_{i=1}^{n_i} \sum_{j=1}^{m_i} \psi(Y_{l,n_i} - X_{j,m_i}) \end{aligned}$$

其均值为$\mu_{W_{n_i}}(\theta_i) = P_{\theta_i}(Y_{l,n_i} > X_{j,m_i})$。由于备择假设的设定根据秩和统计量的渐进分布，可以得到:

$$\sigma_{W_{n_i}}(\theta_i) = \frac{1}{\sqrt{12\lambda(1-\lambda)N_i^2}}$$

$$\begin{aligned}
\mu_{W_{n_i}}(\theta_i) &= P_{\theta_i}(X_{j,m_i} < Y_{l,n_i}) \\
&= \int_{-\infty}^{+\infty} \int_{X}^{+\infty} \mathbf{d}F_Y \mathbf{d}F_X \\
&= \int_{-\infty}^{+\infty} (\int_{X}^{+\infty} \mathbf{d}F_Y) \mathbf{d}F_X \\
&= \int_{-\infty}^{+\infty} 1 - F_Y(X) \mathbf{d}F_X \\
&= \int_{-\infty}^{+\infty} 1 - F_X(X - \theta_i) \mathbf{d}F_X
\end{aligned}$$

因为我们假定了F有密度f，作为密度函数$f \in L^1(R)$，当$\theta_i \to \theta_0 = 0$时，$f_{\theta_i} \to f_{\theta_0}$，而且由于$f$的可积性很容易构造出其控制函数，因此由Lebesgue控制收敛定理：

$$\frac{\mathbf{d}}{\mathbf{d}\theta_i}\mu w_{n_i}(\theta_i) = \int_{-\infty}^{+\infty} f(X-\theta_i)\mathbf{d}F_X \quad a.e$$

代入$\theta_i = \frac{c}{\sqrt{N_i}}$，改写为：

$$\frac{\mathbf{d}}{\mathbf{d}\theta_i}\mu w_{n_i}(\theta_i) = \int_{-\infty}^{+\infty} f(X-\frac{c}{\sqrt{N_i}})\mathbf{d}F_X$$

当$N_i \to +\infty$时，原式近似为：

$$\frac{\mathbf{d}}{\mathbf{d}\theta_i}\mu w_{n_i}(\theta_i) \to \int_{-\infty}^{+\infty} f(X)\mathbf{d}F_X = \int_{-\infty}^{+\infty} f^2(X)\mathbf{d}X$$

至此Noether定理的前五个条件均易于验证。计算出秩和检验的效力因子，即：

$$
\begin{aligned}
eff(W_n^+) &= \lim_{i\to+\infty} \frac{\int_{-\infty}^{+\infty} f^2(X)\mathbf{d}X}{\sqrt{N_i \frac{1}{12\lambda(1-\lambda)N_i}}} \\
&= \sqrt{12\lambda(1-\lambda)} \int_{-\infty}^{+\infty} f^2(X)\mathbf{d}X
\end{aligned}
$$

$$
\begin{aligned}
ARE(W,T) &= \left(\frac{eff(W_n^+)}{eff(T_n)}\right)^2 \\
&= 12\sigma^2 \left(\int_{-\infty}^{+\infty} f^2(X)\mathbf{d}X\right)^2 \\
&= J(\sigma, f)
\end{aligned}
$$

理论计算结果

分布	ARE(W,T)
均匀分布U$(-\sqrt{3},\sqrt{3})$	1
正态分布N(0,1)	$\frac{3}{\pi}$
双指数分布La$(0,\frac{\sqrt{2}}{2})$	$\frac{3}{2}$
Logistic分布Logis$(0,\frac{\pi}{\sqrt{3}})$	$\frac{\pi^2}{9}$
Cauchy分布	∞

通过计算发现，即使是t检验假定的正态分布情形下，Wilcoxon秩和检验的功效也达到了t检验的接近95.5%，而在其他分布下Wilcoxon秩和检验的功效均不低于甚至大幅超过t检验。Cauchy分布下功效比为∞的原因是t检验在Cauchy分布下失效。其他文献中有时会使用自由度较小的t分布近似柯西分布，并且将方差调整为1，$t(3)$功效比约为1.24，$t(5)$功效比约为1.90。因此，Wilcoxon秩和检验整体上优于t检验。

模拟中应注意的问题

(1)注意控制方差。可以看出渐进相对效是方差的函数，这是由t检验导致的。研究选择的分布中除了柯西分布均控制了方差为1。

(2)样本量较大的情况下X与Y的样本量之比会影响各自的功效，但对功效比没有影响。模拟实验选择使功效最大的情形，即样本量相同。

(3)由于理论推导是基于渐进分布的，所以单次模拟生成X与Y的样本量不能太小。

(4)Bootstrap方法不容易理论计算其功效，应控制B的大小，保证其抽取次数和其余两种检验相同。

模拟实验

使用R语言进行模拟，分别模拟均匀分布U($-\sqrt{3},\sqrt{3}$)、正态分布N(0,1)、双指数分布La($0,\frac{\sqrt{2}}{2}$)、Logistic分布Logis($0,\frac{\pi}{\sqrt{3}}$)、Cauchy分布下三种方法的功效。

前文已经理论推导出一系列渐进相对效，相对直观的体现了不同检验的检验效率。接下来将利用R软件编程，使用蒙特卡罗方法产生多组数据，比较几种检验正确拒绝原假设的频率，并以此频率作为功效的估计，探讨检验在各种实际情况下的性能。

控制因素

要求总体X与总体Y除了位置参数不能有任何不同。主要控制以下几点：

(1)两总体相互独立；

(2)两总体的分布仅有位置参数的差异

	总体一	总体二
均匀分布	$U(-\sqrt{3},\sqrt{3})$	$U(\mu-\sqrt{3},\mu+\sqrt{3})$
正态分布	$N(0,1)$	$N(\mu,1)$
双指数分布	$La(0,\frac{\sqrt{2}}{2})$	$La(\mu,\frac{\sqrt{2}}{2})$
logistic分布	$Logis(0,\frac{\pi}{\sqrt{3}})$	$Logis(\mu,\frac{\pi}{\sqrt{3}})$
Cauchy分布	$C(0,1)$	$C(location,1)$

(3)两样本的样本量大小相等；

(4)检验的显著性水平均设定为为0.05，重复实验次数均为500次；

(5)自助检验的自助重复次数等于样本量。

变化因素与检验指标

变化因素

(1)三种检验：t检验、Wilcoxon检验和自助检验；

(2)总体服从五种分布；

(3)两总体位置参数差从0.1至1.5变化，步长为0.05；

(4)两样本的样本量有三个档次：10、30和50。

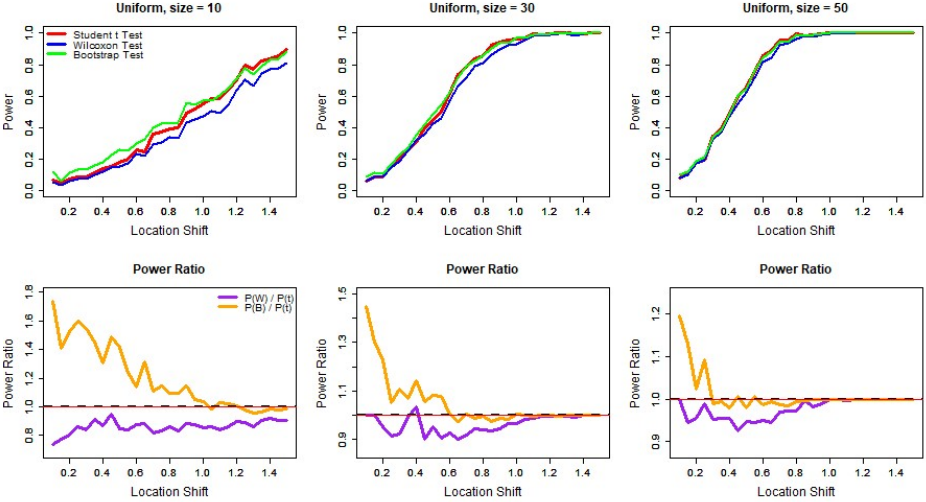
检验指标

(1)三种检验的功效受到各种因素影响的规律；

(2)Wilcoxon检验和t检验的实际功效比和理论渐进相对效的区别；

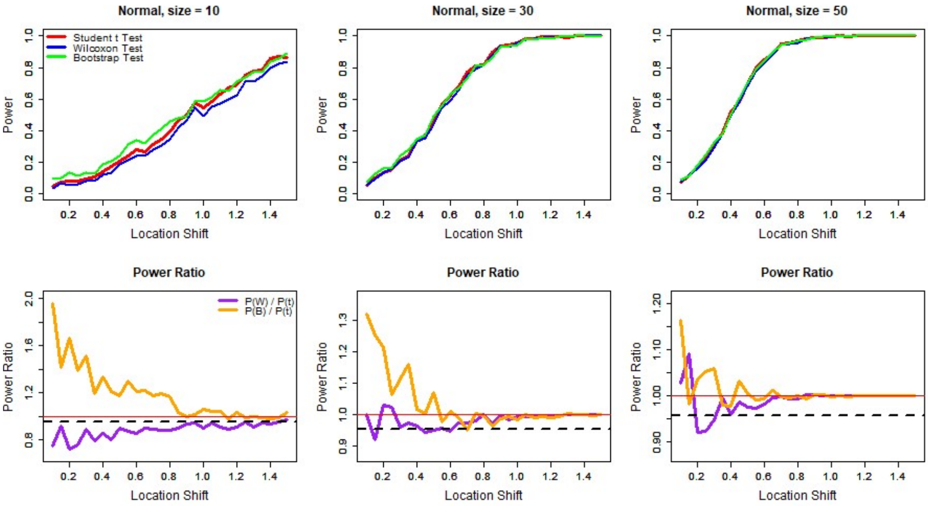
(3)同分布下不同检验的功效比，和不同分布下同检验的功效比。

均匀分布

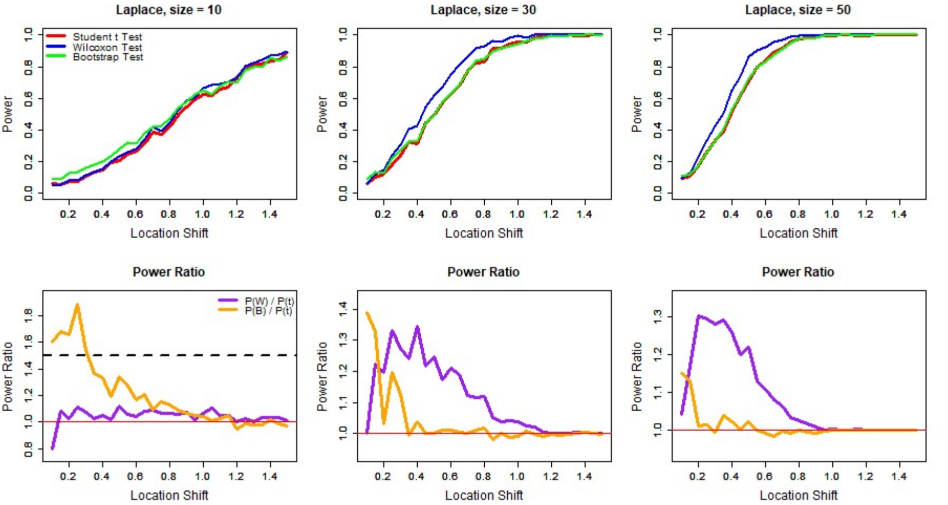


Uniform, size = 10
Uniform, size = 30
Uniform, size = 50
Student t Test
Wilcoxon Test
Bootstrap Test
Power
Location Shift
Power Ratio
P(W) / P(t)
P(B) / P(t)
Power Ratio
Location Shift

正态分布

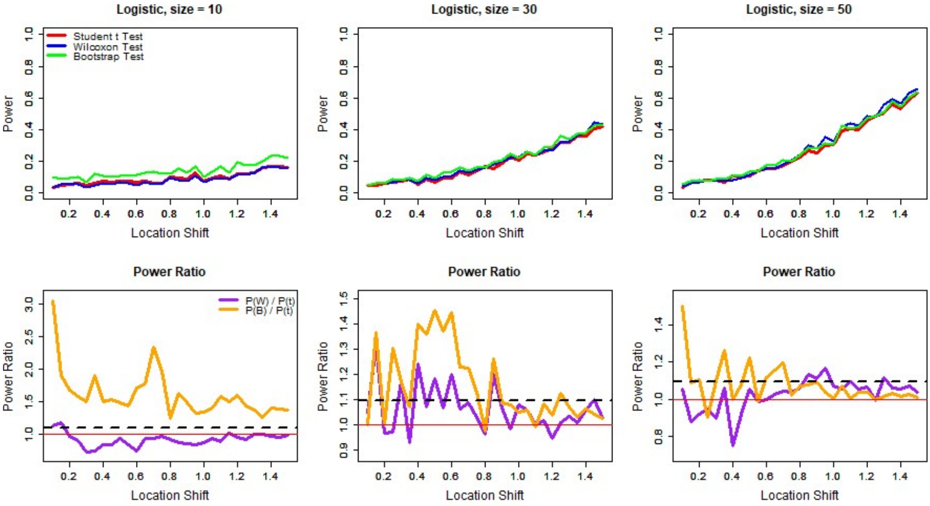

Normal, size = 10
Normal, size = 30
Normal, size = 50
Student t Test
Wilcoxon Test
Bootstrap Test
Power
Location Shift
Power Ratio
P(W) / P(t)
P(B) / P(t)
Power Ratio
Location Shift

双指数分布



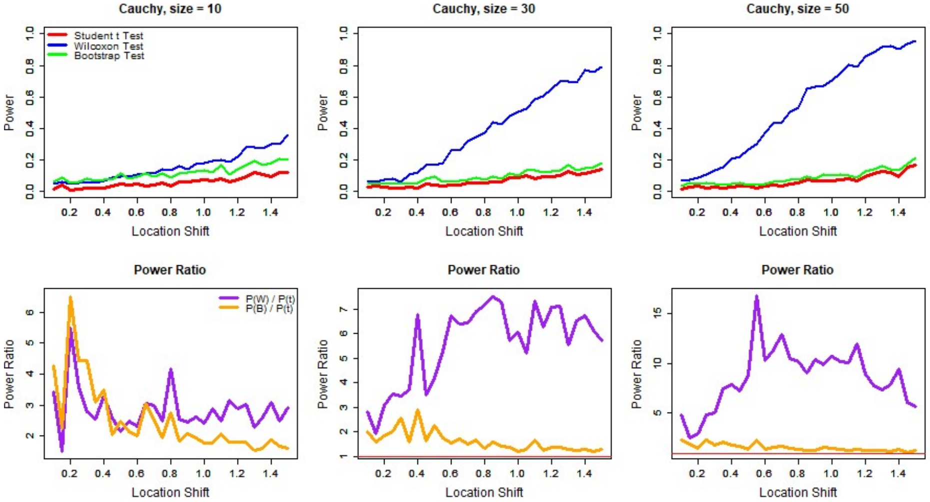
Laplace, size = 10
Laplace, size = 30
Laplace, size = 50
Student t Test
Wilcoxon Test
Bootstrap Test
Power
Location Shift
Power Ratio
P(W) / P(t)
P(B) / P(t)

Logistic分布



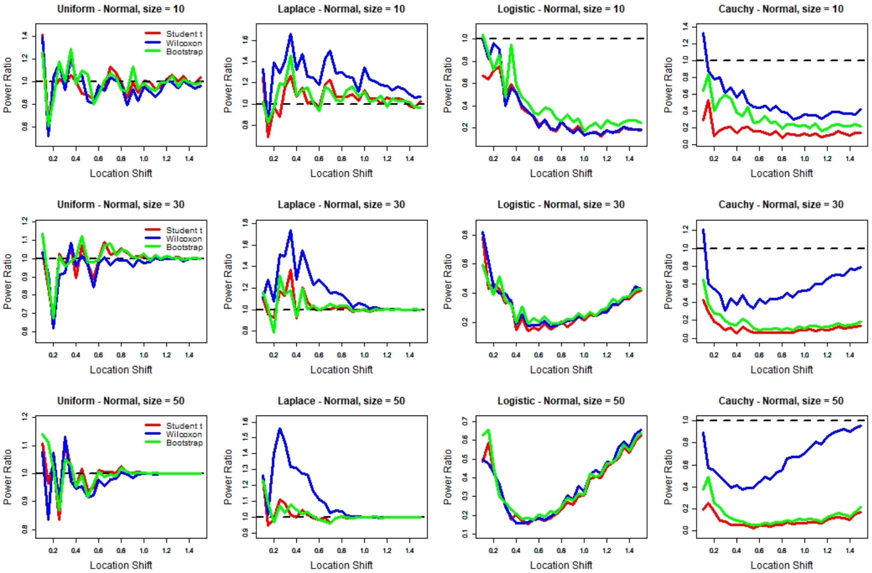
Logistic, size = 10
Logistic, size = 30
Logistic, size = 50
Student t Test
Wilcoxon Test
Bootstrap Test
Power
Location Shift
Power Ratio
P(W) / P(t)
P(B) / P(t)

Cauchy分布



Cauchy, size = 10
Cauchy, size = 30
Cauchy, size = 50
Student t Test
Wilcoxon Test
Bootstrap Test
Power
Location Shift
Power Ratio
P(W) / P(t)
P(B) / P(t)

同种检验不同分布功效比



结论

(1)三种检验的功效随着样本量增加和位置差增加而上升，但在样本来自Logistic分布和Cauchy分布的总体时功效会大打折扣。

(2)Wilcoxon秩和检验在总体服从Cauchy分布时有着另外两种分布无法比拟的优势，并且在双指数分布总体下也有稍高的功效。但对于均匀分布和正态分布总体其功效是最低的。

(3)自助检验在小样本、两总体差距难以区分时表现得更好，其他情况下和t检验的功效几乎没有差异。

(4)t 检验在正态和均匀总体下表现稍微优于Wilcoxon检验，在较大样本和总体位置差异时功效和自助检验相近。

(5)t检验与Wilcoxon秩和检验的功效比较中Pitman渐进相对效往往高于模拟计算的功效比。

Thanks!