Pure Mathematics 理论数学, 2020, 10(9), 852-861
Published Online September 2020 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2020.109099

Hans汉斯

t分布的渐进正态性证明及其特点分析

杨欣童

东北大学秦皇岛分校数学与统计学院，河北 秦皇岛
Email: 15105419855@163.com

收稿日期：2020年8月24日；录用日期：2020年9月15日；发布日期：2020年9月22日

摘 要

前人的研究已经证明t分布函数具有渐进正态的特性。我们改进了前人的研究，应用差函数的特性再次证明了t分布的渐进正态性，并结合MATLAB，直观验证了t分布的这一特性。之后，我们借助直观图形进一步分析了t函数的特性。

关键词

t分布，正态性，证明，MATLAB

Proof of Asymptotic Normality of t Distribution and Analysis of Its Characteristics

Xintong Yang

School of Mathematics and Statistics, Northeastern University at Qinhuangdao, Qinhuangdao Hebei
Email: 15105419855@163.com

Received: Aug. 24th, 2020; accepted: Sep. 15th, 2020; published: Sep. 22nd, 2020

Abstract

Previous studies have proved that the t distribution function has the characteristics of asymptotic normality. This paper applies the characteristics of the difference function to prove the asymptotic normality of the t distribution again, improving the previous research. And with the use of MATLAB software, the asymptotic normality is verified intuitively. After that, we further analyze the characteristics of the t function with the help of intuitive graphics.

文章引用: 杨欣童. t 分布的渐进正态性证明及其特点分析[J]. 理论数学, 2020, 10(9): 852-861.
DOI: 10.12677/pm.2020.109099

Keywords

t Distribution, Normality, Proof, MATLAB

1. 引言

在概率统计学中，t分布函数是一种十分重要的分布函数，这种分布函数具有渐进正态性，即当自由度n趋于无穷大时，t分布函数趋近于标准正态分布。前人已经用求极限的方法证明了此结论[1]-[9]，但是证明过程比较复杂，并且前人没有给出这个过程的细致描述。因此，对于这个逼近过程具有什么样的具体特点不是很清楚。本文拟引进t分布函数与标准正态分布的差函数，首先分析差函数的性质，利用差函数的性质简明地证明t分布的极限分布为标准正态分布，然后，通过利用MATLAB绘制其函数图像来细致分析其逼近过程的具体特点。

2. t 分布及其性质

2.1. t 分布的定义

若两个独立的随机变量 $\xi \sim N(0,1)$，$\zeta \sim \chi^2(n)$，那么随机变量 $X = \dfrac{\xi}{\sqrt{\zeta/n}}$ 服从自由度为n的t分布[10]。

其概率密度函数为

$$f_t(x,n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad (-\infty < x < +\infty) \tag{1}$$

2.2. t 分布的性质

性质 1 (t 分布的等价定义)：若 $n+1$ 个独立的随机变量 $\xi \sim N(0,1)$，$\zeta_1 \sim N(0,1), \zeta_2 \sim N(0,1), \cdots, \zeta_n \sim N(0,1)$，那么随机变量 $X = \dfrac{\xi}{\sqrt{\sum_{i=1}^{n}\zeta_i \Big/ n}}$，服从自由度为 n 的 t 分布；

性质 2 (对称性)：由 t 分布的概率密度函数可得 $f_t(x,n) = -f(-x,n)$ 成立，因此 t 分布的概率密度函数关于 x 轴对称；

性质 3 (均值和方差)：若随机变量 $X \sim t(n)$，则 X 的方差与均值为 $E(X)=0$，$D(X) = \dfrac{n}{n-2}$；

性质 4 (有界性)：若随机变量 $X \sim t(n)$，因为 $D(X) = \dfrac{n}{n-2}$，所以当 $n>2$ 时 X 的方差有界，因此 X 的概率密度函数有界。

3. 利用差函数分析证明渐进正态性

t 分布函数与标准正态分布的差函数的意义为两者在 y 轴方向的距离，若当 n 趋于无穷时，差函数在

其各极值点的取值均趋于 0，则可间接证明当 n 趋于无穷时，t 分布趋于标准正态分布，即

若 $\lim\limits_{n\to\infty}\sup\varphi(x,n)=0$，则 $t(n)$ 趋近标准正态分布 $N(0,1)$。

因此，下面首先分析差函数的性质，再利用差函数的性质证明 t 分布的极限分布为标准正态分布。

3.1. 差函数的定义

设随机变量的取值为 x，则有标准正态分布的概率密度函数为 $f_N(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$，自由度为 n 的 t 分布的概率密度函数为

$$f_t(x,n)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \tag{2}$$

这样，自由度为 n 的 t 分布与标准正态分布的差函数表达式为：

$$\varphi(x,n)=f_t(x,n)-f_N(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}-\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \tag{3}$$

3.2. 差函数的基本性质

性质 1 (对称性)：已知标准正态分布和 t 分布的概率密度函数均关于 x 轴对称，因此差函数也关于 x 轴对称；

性质 2 (极值点)：对 $\varphi(x,n)$ 关于 x 求导数，得到

$$\frac{\mathrm{d}\varphi(x,n)}{\mathrm{d}x}=x\left[\frac{n+1}{n}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+3}{2}}-\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}\right] \tag{4}$$

令 $\frac{\mathrm{d}\varphi(x,n)}{\mathrm{d}x}=0$ 得到极值点 $x_0=0$，另有极值点 x_k 满足方程

$$\frac{n+1}{n}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+3}{2}}=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}} \tag{5}$$

利用上式方程得到差函数的极值

$$\varphi(x_k,n)=\frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x_k}{n}\right)^{-\frac{n+1}{2}}\left[1-\frac{n+1}{n}\left(1+\frac{x_k^2}{n}\right)^{-1}\right] \tag{6}$$

3.3. t 分布的正态渐进性证明

引理 1 (瓦里斯公式推论)：

$$\lim_{n\to\infty}\frac{\left[\frac{(2n-1)!!}{(2n)!!}\right]}{\sqrt{n\pi}}=1 \quad (n\in N) \tag{7}$$

证明：

已知瓦里斯公式为

$$\prod_{n=1}^{\infty}\frac{2n}{2n-1}\times\frac{2n}{2n+1}=\frac{\pi}{2} \quad (n\in N) \tag{8}$$

公式左端变形后为

$$\lim_{n\to\infty}\left[\frac{(2n)!!}{(2n-1)!!}\right]^2\frac{1}{2n-1}=\frac{\pi}{2} \tag{9}$$

再次变形得到

$$\begin{gathered}\lim_{n\to\infty}\left[\frac{(2n)!!}{(2n-1)!!}\right]\sqrt{\frac{2}{(2n-1)\pi}}=1\\ \therefore \lim_{n\to\infty}\left[\frac{(2n)!!}{(2n-1)!!}\right]\sqrt{\frac{1}{n\pi}}=1\end{gathered} \tag{10}$$

证毕

定理 1：当 $n\to\infty$ 时，差函数的极值均趋于 0。

证明：

首先，带入极值点 $x_0=0$ 得到

$$\lim_{n\to\infty}\varphi(x_0,n)=\frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}-\frac{1}{\sqrt{2\pi}} \tag{11}$$

当 n 为偶数时，不妨令 $n=2k$

$$\begin{aligned}\lim_{n\to\infty}\varphi(x_0,n)&=\frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}-\frac{1}{\sqrt{2\pi}}\\&=\lim_{n\to\infty}\frac{\frac{(2k-1)!!}{2^k}}{\sqrt{2k}\,(k-1)!}-\frac{1}{\sqrt{2\pi}}\\&=\lim_{n\to\infty}\frac{1}{\sqrt{2\pi}}\frac{\sqrt{k\pi}\,(2k-1)!!}{(2k)!!}-\frac{1}{\sqrt{2\pi}}\\&=\frac{1}{\sqrt{2\pi}}-\frac{1}{\sqrt{2\pi}}\\&=0\end{aligned} \tag{12}$$

当 n 为奇数时同理可证明 $\lim\limits_{k\to\infty}\varphi(x_0,n)=0$。

下面，考虑其他极值点的极值

$$\lim_{n\to\infty}\varphi(x_k,n)=\lim_{n\to\infty}\frac{\Gamma\left(\frac{n+1}{n}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x_k}{n}\right)^{-\frac{n+1}{2}}\left[1-\frac{n+1}{n}\left(1+\frac{x_k^2}{n}\right)^{-1}\right]$$
$$=\lim_{n\to\infty}f_t(x_k,n)\lim_{n\to\infty}\left[1-\frac{n+1}{n}\left(1+\frac{x_k^2}{n}\right)^{-1}\right] \tag{13}$$

又由于 t 分布的概率密度函数的有界性得到

$$\lim_{n\to\infty}f_t(x_k,n)=p\text{，}(p\text{ 为一有限数}) \tag{14}$$

另外因为

$$\lim_{n\to\infty}\left[1-\frac{n+1}{n}\left(1+\frac{x_k^2}{n}\right)^{-1}\right]=0 \tag{15}$$

所以

$$\lim_{n\to\infty}\varphi(x_k,n)=0 \tag{16}$$

综上，当 $n\to\infty$ 时，差函数的极值趋于 0。

因此，t 分布的极限分布为标准正态分布。

4. t 分布函数的渐进正态性直观验证

前文已经证明了 t 分布函数具有渐进正态性，但没有给出直观描述。下面首先利用 MATLAB 对其渐进正态性进行分析。在 MATLAB 中，输入下列代码，绘制标准正态分布以及自由度从 1 到 100 的 t 分布的概率密度函数。

```
clear all;clc;
x=-4:0.1:4;
n=linspace(1,100,100);
axis([-4 4 0 0.41]);
ylabel('$t(n)$','interpreter','latex', 'FontSize', 18);
xlabel('x')
for i=1:100
A(i,:)=tpdf(x,n(i));
end
plot(x,A);
hold on;
z=normpdf(x,0,1)
plot(x,z,'color','r','linewidth',2.3);
title('自由度从 1 到 100 的 t 分布密度函数和标准正态分布');
legend('n 从 1 变化到 100');
```

结果如图 1 所示。有图 1 可知，当 t 分布函数的自由度 n 增大的时，其概率分布函数在 0 附近的部分上升，其余两边的部分下降，总体趋近于标准正态分布曲线。t 分布函数确实具有渐进正态性。

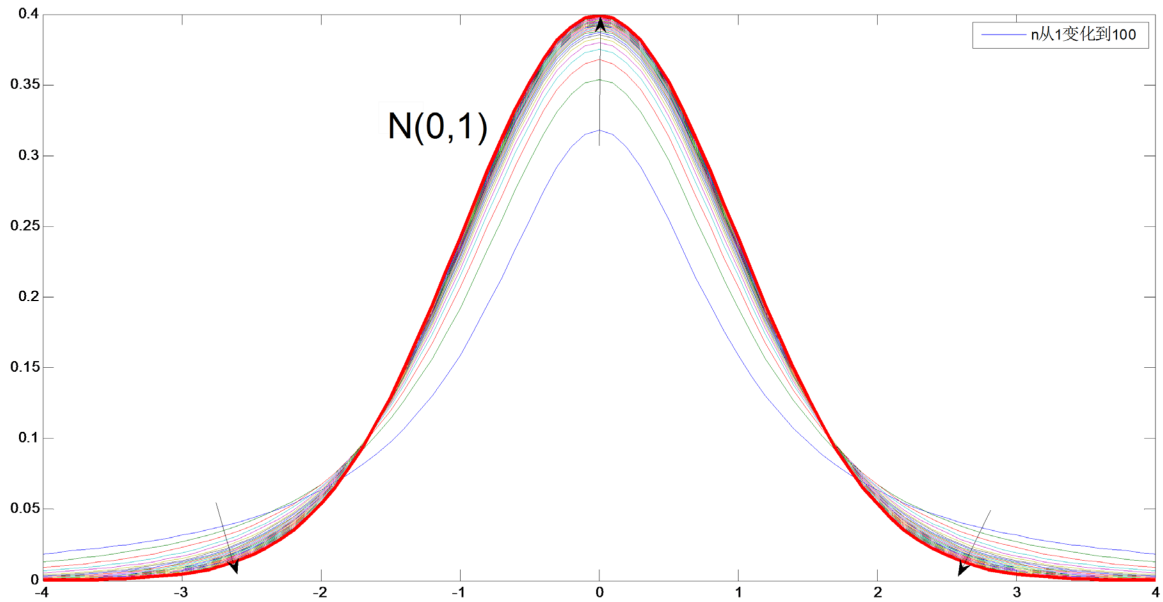

Figure 1. Asymptotic normality of t distribution
图 1. t 分布的渐进正态性

5. 利用差函数分析其具体特点

5.1. 差函数的变化图像

输入如下代码，绘制此函数当参数 n 从 1 变化到 100 的函数图像。

```
clc;clear all;
x=-4:0.01:4;
n=linspace(1,100);
axis([-4 4 -0.2 0.41]);
ylabel('$t(n)-N(0,1)$','interpreter','latex', 'FontSize', 18);
xlabel('x')
z=normpdf(x,0,1)
for i=1:100
A(i,:)=tpdf(x,n(i))-z;
end
plot(x,A);
title('自由度从 1 到 100 的 t 分布密度函数与标准正态分布的差函数');
legend('n 从 1 变化到 100');
```

结果如图 2 所示。从图 2 中可以看出，当参数 n 固定时，差函数是一个关于 y 轴对称的函数，这表明标准正态分布和 t 分布都是关于 y 轴对称的函数。

另外，当参数 n 固定时，很容易看出，一条差函数曲线具有 5 个极值点。由于其对称性，其中一个极值点为 $x=0$，但它不是最值点。因此，虽然 t 分布和标准正态分布的最大值点都在 $x=0$ 处取得，但是他们的差的最大值并不是在 $x=0$ 处取得，而是在其他的极值点处取得。

当参数 n 从 1 增加到 100 时，差函数图像逐渐趋于 x 轴，整体变得平阔，函数范围越来越小，这直观地反映了当 n 增加时，t 分布逐渐趋近于标准正态分布。

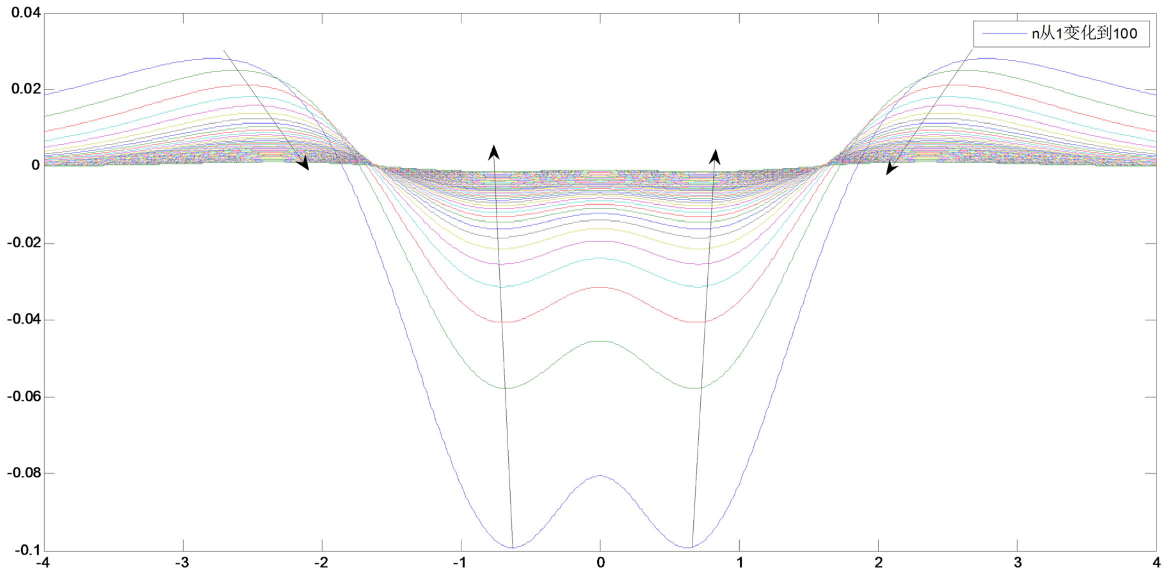

Figure 2. The image of difference function
图 2. 差函数的变化图像

5.2. 差函数的最值

为了进一步分析差函数的性质，输入如下代码，求其当参数 n 从 1 变化到 100 时的差函数的最值。

```
clc;clear all;
MAX=[];MIN=[];a=zeros(1,5)
x=-4:0.01:4;
n=linspace(1,100);
axis([-4 4 -0.2 0.41]);
z=normpdf(x,0,1)
for i=1:100
A(i,:)=tpdf(x,n(i))-z;
MAX(i)=max(A(i,:));
MIN(i)=min(A(i,:));
end
ma=vpa(MAX,3);
ma1=[ma a]
ma2=reshape(ma1,7,[])
ma3=vpa(ma2.',4)
mi=vpa(MIN,5)
mi1=[mi a]
mi2=reshape(mi1,7,[])
mi3=vpa(mi2.',3)
```

结果为：

从表 1、表 2 所得数据可以看出，当差函数的参数从 1 变化到 100 时，差函数最大值从 0.028099 变化到 0.001139，最小值从−0.099264 变化到−0.0013572，显然，他们的绝对值都减小了很多。

Table 1. The maximum of difference functions
表 1. 差函数最大值

[0.0281, 0.025, 0.02118, 0.01815, 0.0158, 0.01397, 0.0125]
[0.01131, 0.01032, 0.009487, 0.008778, 0.008167, 0.007634, 0.007167]
[0.006753, 0.006384, 0.006053, 0.005755, 0.005484, 0.005238, 0.005013]
[0.004806, 0.004616, 0.00444, 0.004277, 0.004126, 0.003985, 0.003853]
[0.00373, 0.003614, 0.003505, 0.003403, 0.003306, 0.003215, 0.003129]
[0.003047, 0.002969, 0.002895, 0.002825, 0.002758, 0.002694, 0.002633]
[0.002575, 0.002519, 0.002466, 0.002415, 0.002366, 0.002319, 0.002274]
[0.00223, 0.002188, 0.002148, 0.002109, 0.002071, 0.002035, 0.002]
[0.001966, 0.001934, 0.001902, 0.001872, 0.001842, 0.001813, 0.001785]
[0.001758, 0.001732, 0.001707, 0.001682, 0.001658, 0.001635, 0.001612]
[0.00159, 0.001569, 0.001548, 0.001528, 0.001508, 0.001489, 0.00147]
[0.001451, 0.001434, 0.001416, 0.001399, 0.001382, 0.001366, 0.00135]
[0.001335, 0.00132, 0.001305, 0.00129, 0.001276, 0.001263, 0.001249]
[0.001236, 0.001223, 0.00121, 0.001198, 0.001185, 0.001173, 0.001162]
[0.00115, 0.001139]

Table 2. The minimum of difference functions
表 2. 差函数最小值

[−0.0993, −0.0578, −0.0407, −0.0313, −0.0255, −0.0215, −0.0186]
[−0.0163, −0.0146, −0.0132, −0.012, −0.011, −0.0102, −0.0095]
[−0.00888, −0.00834, −0.00786, −0.00743, −0.00704, −0.0067, −0.00638]
[−0.0061, −0.00584, −0.0056, −0.00538, −0.00517, −0.00498, −0.00481]
[−0.00464, −0.00449, −0.00435, −0.00421, −0.00409, −0.00397, −0.00385]
[−0.00375, −0.00365, −0.00355, −0.00346, −0.00338, −0.00329, −0.00322]
[−0.00314, −0.00307, −0.003, −0.00294, −0.00288, −0.00282, −0.00276]
[−0.00271, −0.00265, −0.0026, −0.00255, −0.00251, −0.00246, −0.00242]
[−0.00238, −0.00233, −0.0023, −0.00226, −0.00222, −0.00218, −0.00215]
[−0.00212, −0.00208, −0.00205, −0.00202, −0.00199, −0.00196, −0.00194]
[−0.00191, −0.00188, −0.00186, −0.00183, −0.00181, −0.00178, −0.00176]
[−0.00174, −0.00172, −0.0017, −0.00167, −0.00165, −0.00163, −0.00161]
[−0.0016, −0.00158, −0.00156, −0.00154, −0.00152, −0.00151, −0.00149]
[−0.00147, −0.00146, −0.00144, −0.00143, −0.00141, −0.0014, −0.00138]
[−0.00137, −0.00136]

5.3. 差函数最值的直观表示

为了使差函数最值随 n 的变化更直观，输入如下代码，绘制其当 n 从 1 变化到 100 函数的图像。

```
clc;clear all;
syms E F;
X1=[];X2=[];SUP=[];INF=[];
x=-4:0.001:4;
n=linspace(1,100);
```

```
z=normpdf(x,0,1)
for i=1:100
A(i,:)=tpdf(x,n(i))-z;
SUP(i)=max(A(i,:));
INF(i)=min(A(i,:));
E=find(A(i,:)==SUP(i))
X1(i)=(E(:,2)-4000)*0.01
F=find(A(i,:)==INF(i))
X2(i)=(F(:,2)-4000)*0.01
end
plot(n,SUP,'o-','color','r','linewidth',2);
hold on;
plot(n,INF,'s-','color','b','linewidth',2);
title('参数从 1 到 100 的的差函数最值');
legend('差函数最大值','差函数最小值');
ylabel('$t(n)-N(0,1)最值$','interpreter','latex', 'FontSize', 18);
xlabel('n')
grid on;
```

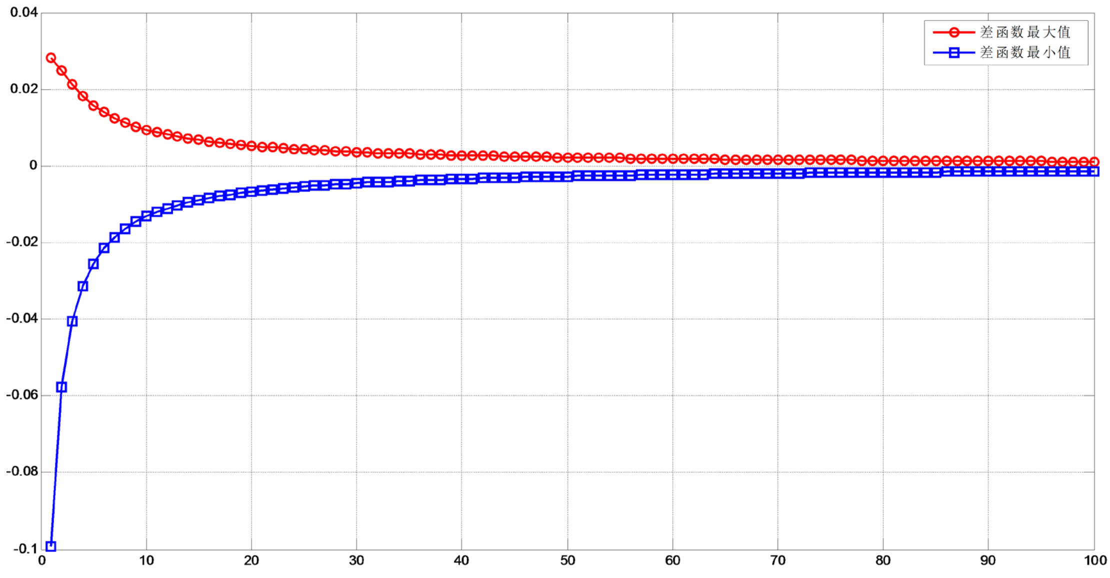

Figure 3. The maximum value of the difference function varies with n
图 3. 差函数最值随 n 的变化

结果如图 3 所示。从图 3 可以看出差函数的最大值和最小值随参数 n 的增大都趋于 0。当 n 增加时，函数的变化速度减慢，因此看出，自由度 n 增加时，t 分布渐进于标准正态分布的速度减慢。

6. 结论

1) 当自由度 n 趋于无穷的时，t 分布总体趋近于标准正态分布曲线，可通过构造的差函数比较简单地证明；

2) 趋近方式为其概率分布函数在 0 附近的部分上升，其余两边的部分下降；
3) 当自由度 *n* 趋于无穷的时，t 分布趋近于标准正态分布，但是速度越来越慢；
4) 当自由度 *n* 趋于无穷的时，t 分布在不同点与标准正态分布的差值不相同；
5) 当自由度 *n* 趋于无穷的时，t 分布不同点趋近于标准正态分布的速度不相同。

参考文献

[1] 杨洁, 李兆庚. 关于 t 分布的极限分布为标准正态分布的证明[J]. 石油化工高等学校学报, 1994(3): 78-80.

[2] 严琴, 唐贺, 李开灿. 中心多元 t 分布的一致渐近正态性[J]. 湖北文理学院学报, 2013, 34(2): 9-13.

[3] 李开灿, 孟赵玲. x~2 分布、t 分布和 F 分布的一致渐近正态性[J]. 北京印刷学院学报, 2004(3): 30-33.

[4] 梁琼. X~2·T·F 分布的正态逼近[J]. 湖北民族学院学报(自然科学版), 1995(2): 33-38.

[5] 王娟. t 分布密度函数之性质[J]. 淮阴工学院学报, 2007(5): 15-21.

[6] 彭定忠, 张映辉, 刘朝才. t 分布概率密度的分析[J]. 湖南城市学院学报(自然科学版), 2012, 21(4): 50-52.

[7] 张光远. 关于多元 t 分布的一些讨论[J]. 新疆大学学报(自然科学版), 1996(3): 33-38.

[8] Li, R. and Saralees, N. (2020) A Review of Student's t Distribution and Its Generalizations. *Empirical Economics*, **58**, 1461-1490. https://doi.org/10.1007/s00181-018-1570-0

[9] Chen, C.P. (2018) Complete Asymptotic Expansions for the Density Function of t -Distribution. *Statistics and Probability Letters*, 141. https://doi.org/10.1016/j.spl.2018.05.014

[10] 魏宗舒. 概率论与数理统计教程[M]. 北京: 高等教育出版社, 2008: 145-146.