

The Implementation of Random Trial of Statistical Distribution in Eviews

Yingzi Cao

International Business College, Shaanxi Normal University, Xi'an Shaanxi
Email: 793914288@qq.com

Received: Mar. 27th, 2018; accepted: Apr. 16th, 2018; published: Apr. 23rd, 2018

Abstract

Random simulation is an important way for people to understand the regularity of random phenomena. It has an important application in understanding and studying the distribution characteristics of statistics and estimates. Eviews software provides a platform for people to carry out random simulation. In this paper, a few examples are given to illustrate the strong support of Eviews's random simulation function to the research of statistics and estimation distribution.

Keywords

Random Simulated, Statistics, Estimators, Eviews

统计分布的随机试验在Eviews中的实现

曹樱子

陕西师范大学国际商学院, 陕西 西安
Email: 793914288@qq.com

收稿日期: 2018年3月27日; 录用日期: 2018年4月16日; 发布日期: 2018年4月23日

摘要

随机模拟是人们认识随机现象中的规律性的一种重要方式, 在认识和研究统计量和估计量的分布特征时具有重要应用。Eviews软件为人们进行随机模拟提供了一种平台。本文通过几方面的举例说明了该软件的随机模拟功能对统计量和估计量分布的研究提供的有力支持。

关键词

随机模拟, 统计量, 估计量, Eviews

Copyright © 2018 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

人们可以通过反复的观察认识随机现象中的规律性, 也可以通过主动的随机试验来认识随机现象中的规律性。它们都需要通过大量的重复进行才能获得对随机现象的认识。在实践中, 观察当然是被动的, 而随机试验有时费时费力, 有时难以进行。随机模拟是人们认识随机现象中的规律性的另外一种重要方式, 在研究统计量和估计量的分布特征时具有重要应用。统计量和估计量是统计过程中的重要概念, 它们是样本的函数, 随着样本的变化而变化, 具有一定的分布规律。人们通过解析的推理方法已经得出了很多统计量和估计量的分布特征, 还不清楚大量的统计量和估计量的分布特征。

随着统计软件的不推出, 随机模拟可以很容易地进行, 给人们认识随机现象提供了工具和手段。随机模拟是一种通过设定随机试验, 反复生成样本, 计算统计量和估计量, 进而研究其分布特征的方法。Eviews 软件除了提供各种经济计量分析功能外, 还提供了很多随机抽样函数(该软件有数据处理、作图、统计分析、建模分析、预测和模拟等六大类功能), 为人们进行随机模拟提供了一种平台。利用 Eviews 软件的随机模拟功能, 反复产生很多组样本, 能够探索统计量和估计量的分布规律。

2. Eviews 中的随机抽样函数

Eviews 中提供了 18 种随机变量的随机抽样函数, 都是以@r 开头, 加上相应的分布名称, 如表 1 所示(括号中是分布参数) [1] [2] [3]:

运用这些函数就可以从 18 种随机变量所服从的分布中抽取任意数目的随机样本, 进行随机试验或对统计量和估计量的分布特征进行经验研究。下面通过几方面的例子介绍随机试验在 Eviews 中的实现。

Table 1. Random sampling function in Eviews

表 1. Eviews 中的随机抽样函数

分布类型	抽样函数	分布类型	抽样函数
二项分布	@rbinom(n,p)	负二项分布	@rneqbin(n,p)
泊松分布	@rpoisson(m)	均匀分布	@runif(a,b)
标准正态分布	@rnorm 或 nrnd	对数正态分布	@rlognorm(m,s)
指数分布	@rexp(m)	威布尔分布	@rweib(m,a)
Beta 分布	@rbeta(a,b)	Gamma 分布	@rgamma(r,s)
Lodistic 分布	@rlogistic(x)	帕累托分布	@rpareto(a,k)
拉普拉斯分布	@rlaplace(r,s)	极值分布 I	@rextreme
广义误差分布	@rged(r)	卡方分布	@rchisq(v)
t 分布	@rtdist(v)	F 分布	@rfdist(v1,v2)

3. 估计圆周率的随机试验

对于圆周率 π ，人们提出了很多种估计方法。现在设计一个随机试验来估计 π 。构造一个单位正方形和内切圆，往正方形里面随机地投点。定义一个随机变量 X ，如果随机投点落在内切圆内则其值为 4，否则其值为 0。这样理论上来说 X 的期望为 π 。该试验在 Eviews 软件中实现的代码如下(代码可以在命令行逐条运行) [4]:

```
wfcreate mywork u 1 10000
    '建立工作文件，命名为 mywork，范围是 1~10000，即模拟 10000 次
series x=@runif(0,1)
    '建立序列 x，其值是从均匀分布中随机抽取，10000 个，也可用 rnd 函数得到
series y=@runif(0,1)
scat x y
    '查看散点图
genr z=4*((x-0.5)^2+(y-0.5)^2<0.5^2)
    '生成新序列 z，如果点(x,y)落在圆内，则取值为 4，否则取值为 0
z.stats
    '查看序列 z 的描述性统计量的值
```

图 1 是运行一次的结果。从图 1 可以看到，这次通过 10,000 个随机投点对 π 的估计结果为 3.1468。随着样本容量的增大，模拟结果就会越来越接近于 π 的真值。当然，以上代码也可以编写到程序窗口，通过点击程序窗口的 run 按钮，重复运行。

4. 对随机变量的函数分布的验证

统计理论指出，如果随机变量 X 服从标准正态分布，则 X^2 服从卡方分布；如果随机变量 Y 服从正态分布，则 e^Y 服从几何正态分布(或称作对数正态分布)。对于这些结论可以在 Eviews 软件中加以验证。模拟试验的代码如下:

```
wfcreate myfill u 1 1000
series x=@rnorm
    '创建序列 x，数值随机抽取于 N(0,1)
genr x2=x^2
    '生成新序列 x2，X2=X^2
x2.hist
    '绘制序列 x2 的直方图
series y=1+0.5*@rnorm
    '创建序列 y，数值来自于 N(1,0.5^2)
genr expy=exp(y)
    '生成新序列 expy，expy=exp(y)
expy.hist
    '绘制序列 expy 的直方图
```

下面是一次随机模拟的结果(图像经过了冻结、合并):

图 2 中的左图说明 X^2 服从的是卡方分布，右图说明 $\exp y$ 服从的是几何正态分布。通过该验证可以

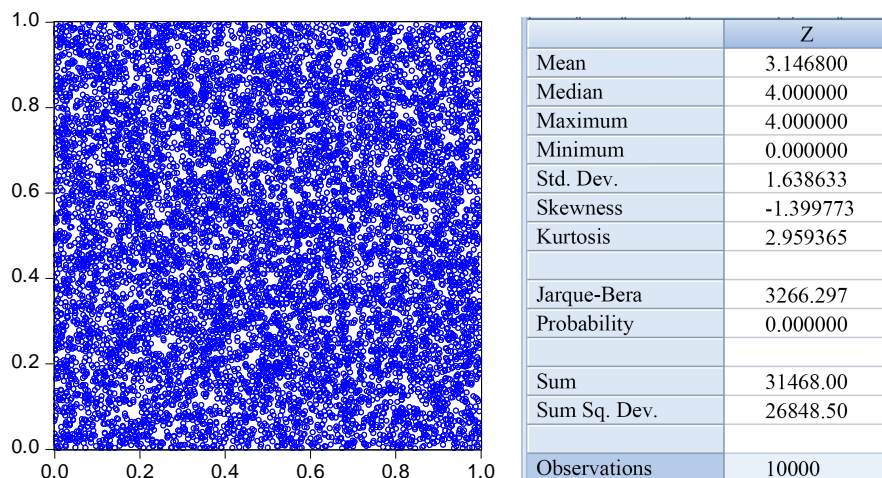


Figure 1. The distribution of 10,000 random sampling points and estimation of PI

图 1. 10,000 个随机抽样点的分布及其对圆周率的估计

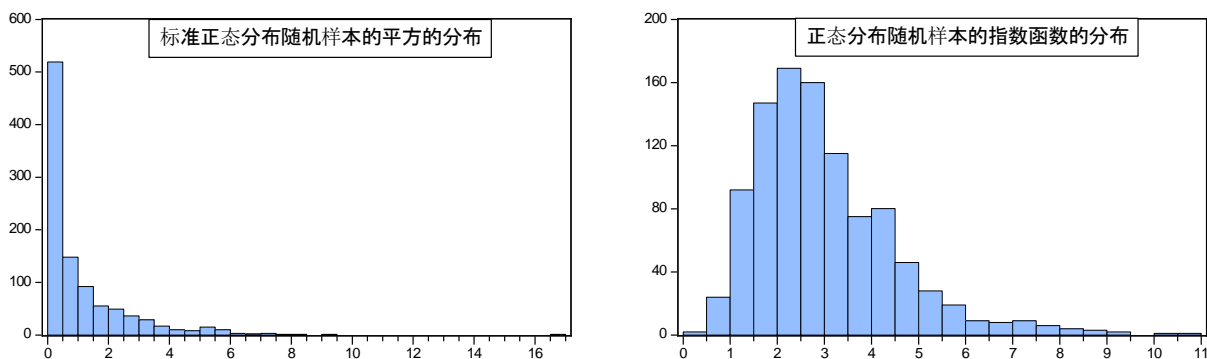


Figure 2. Empirical relations of normal, chi square and geometric normal distribution

图 2. 正态分布和卡方分布、几何正态分布的经验关系

增加对正态分布和卡方分布、几何正态分布之间关系的认识，同时获得对卡方分布、几何正态分布的分布特征的认识。还有其他的分布结论也可以验证，例如 t 分布、F 分布等。

5. 对样本统计量的随机模拟

统计量是统计数据处理中的重要概念，是对样本特征的总结或概括。统计量的分布是通过样本特征推断总体特征的理论基础，研究统计量的分布是统计理论中的一项重要内容。对于像均值这样的统计量所服从的分布，理论上已经有了确定的结果，通过 Eviews 软件可以进行验证，从而对其特征获得感性的、直观的认识。而对于偏度、峰度这些统计量所服从的分布，理论上还没有确定的结论，通过 Eviews 软件的随机模拟可以获得一定的经验认识。

5.1. 对样本均值分布的随机模拟

统计理论指出，如果总体服从正态分布，即 $X \sim N(\mu, \sigma^2)$ ，则容量是 n 的样本均值也服从正态分布，且 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ ；如果总体不服从正态分布，当样本容量很大时，样本均值渐近地服从正态分布。现在假定一个总体服从的分布为 $N(1, 2^2)$ ，另一个总体服从参数为 3 的指数分布。在 Eviews 中进行随机模拟，

重复抽样 1000 次，每次 200 个样本数，观察样本均值的分布。模拟的代码如下(循环语句必须输入到程序窗口，通过命令 `program` 打开程序窗口，运行大约 2 分钟):

```
wfcreate mywork u 1 1000
for !k=1 to 1000
    '定义控制变量!k，建立循环，循环 1000 次
series x{!k}
    '建立序列 x{!k}
series y{!k}
smpl 1 200
    '调整样本范围到 1~200
x{!k}=1+2*@rnorm
y{!k}=@rexp(3)
    '序列 y{!k}的值来自参数是 3 的指数分布
scalar xm=@mean(x{!k})
    '计算序列 x{!k}的样本均值
scalar ym=@mean(y{!k})
smpl 1 1000
    '恢复工作文件范围到 1~1000
matrix(1000,2) mat
    '建立矩阵 mat(1000×2)
mat(!k,1)=@mean(x{!k})
    '将序列 x{!k}的均值存放于矩阵 mat 第一列中
mat(!k,2)=@mean(y{!k})
next
    '循环结束
delete x?? x??? y?? y???
    '删除一些序列(?是通配符)
mtos(mat,g)
    '将矩阵 mat 转化为两个序列(默认名称是 ser01 和 ser02)
ser01.hist
    '查看序列 ser01 的直方图
ser02.hist
```

下面的图 3 和图 4 是一次运行的结果:

从图 3 可以看出，样本均值的均值接近于 1，样本均值的标准差接近于 $\sqrt{2}/10$ ，并且 JB 统计量说明不能拒绝样本均值服从于正态分布的原假设。

从图 4 可以看出，样本均值的均值接近于 3，样本均值的标准差接近于 $\sqrt{9/200}$ ，同样 JB 统计量说明也不能拒绝样本均值服从于正态分布的原假设，表明试验结果与理论结论一致。

对于该随机模拟，可以更改模拟次数、样本容量、总体的分布类型等对样本均值的分布特征一一加以验证。

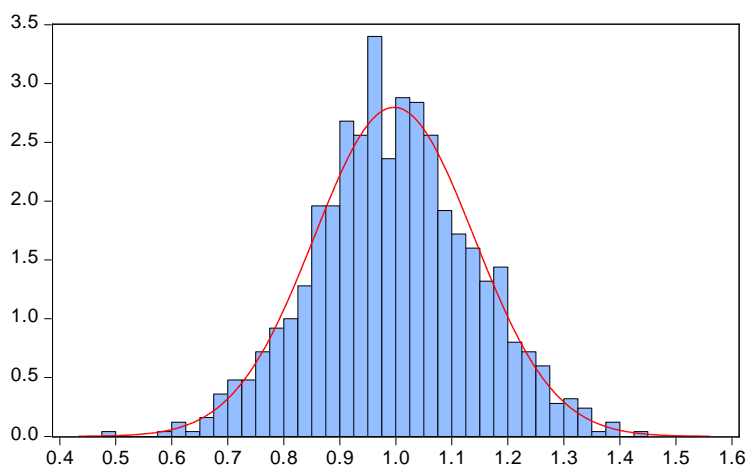


Figure 3. Distribution of sample mean of normal distribution (sample volume is 200)

图 3. 正态分布的样本均值的分布(样本容量为 200)

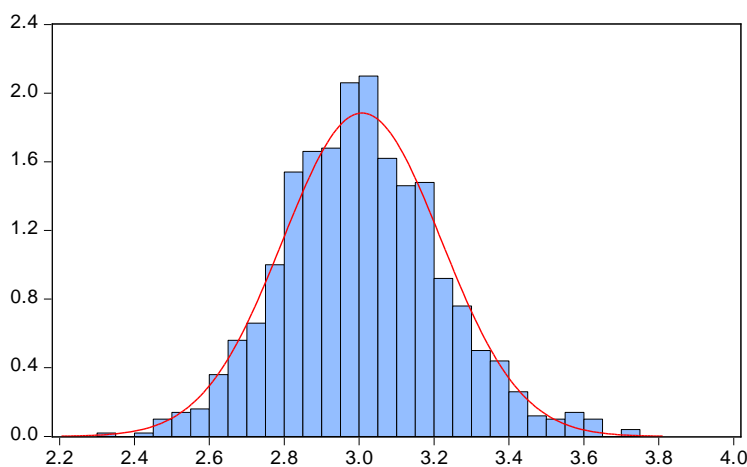


Figure 4. The distribution of sample mean of exponential distribution (sample volume is 200)

图 4. 指数分布的样本均值的分布(样本容量为 200)

5.2. 对正态分布的样本偏度和峰度分布的随机模拟

统计理论指出, 如果总体服从于正态分布, 则容量是 n 的样本偏度和样本峰度渐近地服从均值为 0, 方差分别为 $6/n$ 和 $24/n$ 的正态分布。(检验样本是否来自于正态分布的 JB 统计量正是在此基础上构造出来的。)该结论的证明比较困难, 但是可以通过随机模拟获得直观认识。现在假定一个总体服从的分布为 $N(1, 2^2)$, 从中重复抽样 2000 次, 每次 300 个样本数, 观察样本偏度和峰度的分布。模拟的代码如下:

```
wfcreate myfile u 1 300
for !k=1 to 2000
series x{!k}
x{!k}=1+2*@rnorm
scalar xs=@skew(x{!k})
    '计算序列 x{!k} 的样本偏度
```

```

scalar xk=@kurt(x{!k})
    '计算序列 x{!k} 的样本峰度
matrix(2000,2) mat1
mat1(!k,1)=@skew(x{!k})
mat1(!k,2)=@kurt(x{!k})
next
delete x?? x???
range 1 2000
    '扩大工作文件范围为 1~2000
mtos(mat1,g)
ser01.qqplot
    '查看序列 ser01 的 qq 图(分位数-分位数图)
ser02.qqplot

```

下面的图 5 是一次运行的结果(将两个图像冻结、合并):

从图 5 可以看出, 样本偏度服从正态分布, 而样本峰度与正态分布还有一定差距, 可以增大样本容量和增加模拟次数来进一步验证。

类似的, 还可以通过随机模拟来研究来自于其他分布的各种统计量(比如偏度、峰度、中位数、极小值、极大值等等)的分布特征。

5.3. 对相关系数的分布的模拟

统计理论指出, 当两个总体的相关系数 ρ 等于 0 时, 样本相关系数 r 服从于自由度为 $n-2$, 标准差为 $\sqrt{\frac{1-r^2}{n-2}}$ 的 t 分布。现在从两个独立的标准正态分布中每次抽取 500 个样本数, 重复 10,000 次来验证该

结论。在 Eviews 中进行随机模拟的代码如下:

```

wfcreate xishu u 1 10000
series rxy
for !a=1 to 10000
    smpl 1 500
    series x=@rnorm
    series y=@rnorm
    scalar r=@cor(x,y)
    '计算样本相关系数 r
    rxy(!a)=r
    '将样本相关系数存放在序列 rxy 中
next
smpl 1 10000
rxy.hist
rxy.distplot cdf
    '查看序列 rxy 的经验分布函数图

```

下面的图 6 是一次运行的结果(将直方图和经验分布函数图冻结、合并)。

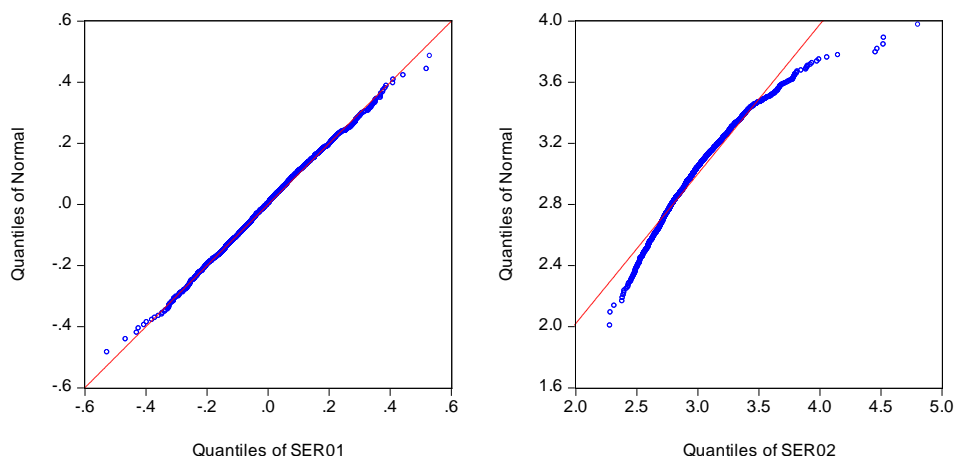


Figure 5. qq diagram of the skewness and kurtosis of random sample (sample volume is 300)

图 5. 正态分布随机样本的偏度和峰度的 qq 图(样本容量为 300)

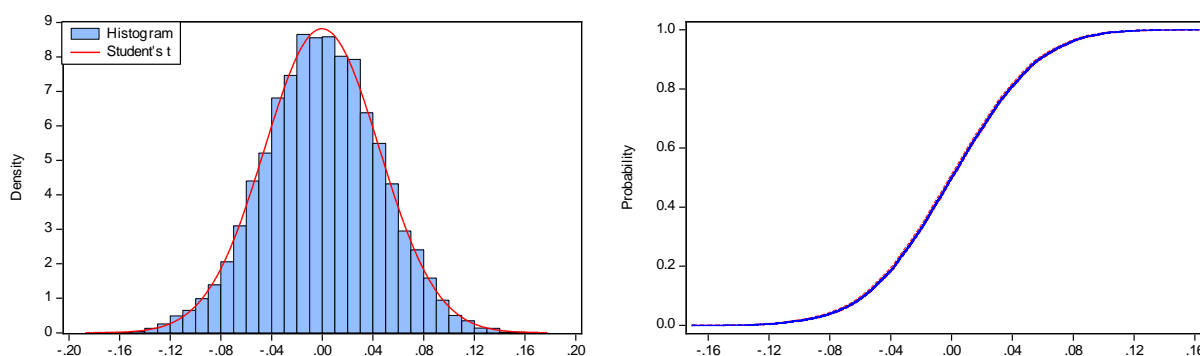


Figure 6. Distribution diagram of sample correlation coefficient (sample volume is 500)

图 6. 样本相关系数的分布图(样本容量为 500)

从图 6 左边的直方图可以看出, 样本相关系数的分布与 t 分布一致; 右边的经验分布函数图说明样本相关系数的分布与标准正态分布相当接近(样本容量很大)。

当总体相关系数 $\rho \neq 0$ 时, 因为 $-1 < r < 1$, 所以样本相关系数将会是一个有偏的分布。理论上, 经过 Fisher 变换: $z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$, z 将渐进地服从均值为 $\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$, 标准差为 $\frac{1}{\sqrt{n-3}}$ 的正态分布。这时获

得样本数据的方法是首先生成两个无关的 u, v 的随机取值, 然后令 $X = u, Y = \rho u + \sqrt{1-\rho^2}v$, 则 X, Y 的相关系数从理论上讲为 ρ , 通过反复生成 u, v 的取值, 就可以模拟样本相关系数的分布。

6. 对估计量的随机模拟

经济计量学理论指出, 在满足高斯-马尔科夫假定条件下, 普通最小二乘估计量是最佳线性无偏估计量。由于在实际中总体模型总是未知的, 并且只有一组样本数据, 所以该特征难以展示。我们可以指定一个总体模型, 然后在 Eviews 中反复抽样进行模型的估计, 由多次得到的估计结果就可以展示参数估计量的无偏性[5]。

假定的总体模型为: $Y = 2 + 0.2X + \mu, \mu \sim N(0, 3^2)$ 。随机模拟的代码如下:

```
wfcreate my u 1 2000
series x
```



```

x(1)=-10
    '序列 x 的第 1 个值赋值为-10
smpl 2 21
x=x(-1)+1
    '序列 x 的前一个值加 1 为下一个值
smpl 1 21
for !k=1 to 2000
series y{!k}
y{!k}=2+0.2*x+3*@rnorm
equation eq{!k}.ls y{!k} c x
    '建立方程对象, 名称为 eq{!k}, 用最小二乘法估计
matrix(2000,3) ma
ma({!k},1)=c(1)
ma({!k},2)=c(2)
ma({!k},3)=@se
    '矩阵 ma 的第三列存放每次估计的标准误差
next
smpl 1 2000
mtos(ma,g)
freeze(g1) ser01.hist
    '冻结 ser01 的直方图, 名称为 g1
freeze(g2) ser02.hist
freeze(g3) ser03.hist
graph graph01.merge g1 g2 g3
    '合并三个直方图,名称为 graph01
graph01.align(3,1,1)
    '将三个直方图横向排列

```

图 7 是一次运行的结果。从图 7 可以看出, 参数估计量具有无偏性, 并且也服从正态分布。这就展示或者说验证了理论结论。在经济计量学理论中, 还有很多类似结论都是对重复样本而言的, 通过 Eviews 软件进行重复抽样就可以直观化这些结论。

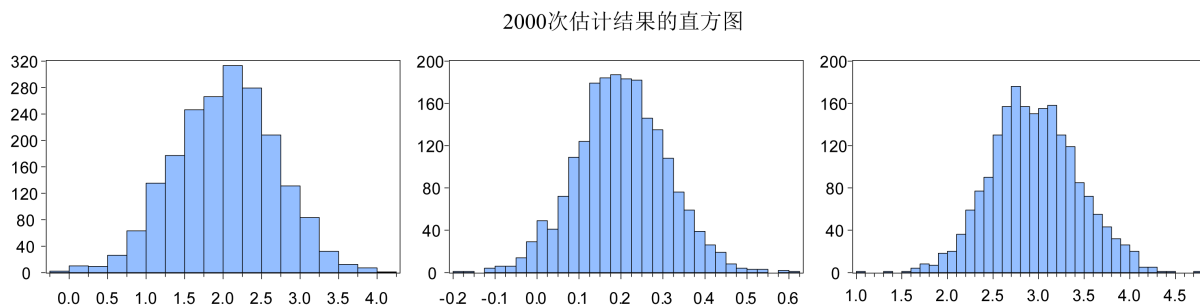


Figure 7. The unbiased estimator of the least squares (intercept, slope, and standard error)

图 7. 最小二乘法估计量的无偏性(从左到右分别是截距、斜率和标准误差)

7. 结语

在实践中，人们首先是对某些随机现象中的规律性获得经验认识，然后经过理论分析才上升到理性认识。大量的随机现象中的规律性是难以通过这种方式来认识 and 研究的。随机模拟是人们获得经验认识的一种重要方式。Eviews 软件提供了强大的随机模拟功能。本文通过几方面的举例说明了 Eviews 的随机模拟功能对统计量和估计量分布的研究提供的有力支持。对于统计活动和经济计量活动中的很多随机模拟问题，可以类似地进行研究。

参考文献

- [1] 陈灯塔. 应用经济计量学——EViews 高级讲义[M]. 北京: 北京大学出版社, 2012.
- [2] 高铁梅. 计量经济分析方法与建模: EViews 应用及实例[M]. 第 2 版. 北京: 清华大学出版社, 2009.
- [3] 张晓彤. Eviews 使用指南与案例[M]. 北京: 机械工业出版社, 2007.
- [4] 莫达隆. 概率统计中随机模拟实验的设计——兼谈 eviews 随机数发生器的使用[J]. 福建教育学院学报: 高教, 2012(4): 119-121.
- [5] 王佐仁, 徐生霞. 蒙特卡罗方法下线性模型的异方差性检验方法[J]. 统计与信息论坛, 2016(11): 33-37.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sa@hanspub.org