2014 年 7 月
株式会社　ライトストーン

大数の法則

前回はデータ (標本) の分布を知るためのモーメント (平均、分散、歪度、尖度) について紹介しました。実際に EViews のコマンドを利用して正規分布や対数正規分布の標本を、乱数機能でたくさん生成しました。今回は 1 次のモーメントである「平均」について少し丁寧に考えてみましょう。

我々の興味関心は、今、小学 6 年生の平均身長を推定するところにあるものとします。「日本全国」の小学 6 年生の平均身長に興味があるとすれば、母集団は「日本全国の小学 6 年生」ということになります。確率変数 X は全国の小学 6 年生の身長であり、その母平均を μ、母分散 σ^2 とします (分布は特定しません)。ただ、我々には費用と時間の制約ありますので、ランダムサンプリング[1]により、n 人の小学 6 年生を選び、身長を計測することにします。この時の n 人の平均値を $\bar{X}$ と書くことにします。

$$\bar{X} = \frac{x_1 + \cdots + x_n}{n} \tag{1}$$

n が小さいと、バラツキによって $\bar{X}$ と μ は少し離れた値になりそうですが、n が大きくなるほど、$\bar{X}$ が μ に近づくことが想像できます。手元に収集した標本 x_i の個数が増えると、一次のモーメントである平均 (推定値) は X の母平均 μ に限り無く近づきます。この事を「大数 (たいすう) の法則」(Law of Large Numbers) と呼び、式で示すと次のようになります。

$$\Pr\{|\bar{x} - \mu| > \epsilon\} \to 0 \quad (n \to \infty) \tag{2}$$

ϵ は $\epsilon > 0$ を満たすものとします。2 式は、「n が大きくなるにつれ、標本平均 $\bar{x}$ と母平均 μ の差の絶対値が ϵ より大きくなる確率は、ゼロに収束する」と事を意味しています。計量経済学の論文では 2 式を $\bar{x} \xrightarrow{P} \mu$ と書くこともあります。

それではここで EViews のプログラミング機能を利用して、大数の法則をシミュレーションしてみましょう。仮に小学 6 年生の身長 X は $N\left(145, 10^2\right)$ に従うものとします。

注意: ここではコンピュータによるシミュレーションの都合上、皆さんに馴染みのある正規分布を利用して標本を生成しますが、前述のように分布はどんなものであってもかまいません。

次のプログラムを記述してください。

’「標本数 n が増えると、標本平均 $\bar{X}$ は母平均 μ に収束する」ことを見て学ぶ
’550 個の標本を利用します。ワークファイル名は height にします
```
wfcreate(wf=height) u 550
```

’ 行列オブジェクト lln には 10, 50, 100, ..., 550 人の身長データが入ります
```
matrix(550,12) lln
```

’mean には 10, 50, 100, ..., 550 人の身長データの平均が入ります
```
matrix(12,2) mean
```

’ 乱数発生のキーを 12 にします。
```
rndseed 12
```

[1] ランダムサンプリングは容易ではありません.

'550 人分の身長データを x として一括作成します

```
series x=145+10*nrnd
```

'x の先頭から 10 行目までのデータを lln の 1 列目にコピーします

```
for !i=1 to 10
  lln(!i,1)=x(!i)
next
```

'1 列目の 11 行目から 550 行目まで NA(欠損値) を入力します。
' これを入力しないと最後のグラフ化がきれいにできません

```
for !i=11 to 550
  lln(!i,1)=na
next
```

' 先頭から $50, 100, \ldots, 550$ 行までのデータを lln の 2 列目から 12 列目まで x の値をコピーし、その下に NA を入力します

```
for !k=2 to 12
    for !i2=1 to 50*(!k-1)
    lln(!i2,!k)=x(!i2)
    next
    for !i2=50*(!k-1)+1 to 550
    lln(!i2,!k)=na
    next
next
```

'lln において、標本数が $10, 50, 100, \ldots, 550$ となっている各列のヒストグラムを作成します
' 標本数が増えるにしたがって、棒の数が増えます。
' プログラムの先頭からここまでの狙いは、標本数が徐々に増えていく処理を行う事です
' まだ、大数の法則のシミュレーションではありません

```
lln.distplot hist theory(dist=normal)
```

' それぞれのヒストグラムに対応する平均値 (推定値)$\bar{X}$ を求めます
' 最初に先頭 10 人分の平均値を求め、行列 mean の 1 行 2 列目にコピーします

```
smpl 1 10
mean(1,1)=10
mean(1,2)=@mean(x)
```

' 同じ処理を $50, 100, \ldots, 550$ 人で行います。

```
for !j=2 to 12
    !end=50*(!j-1)
    smpl 1 !end
    mean(!j,1)=!end
```

```
    mean(!j,2)=@mean(x)
next
```

’求めた 12 個の標本平均の値をグラフ化します。
’標本数が増えるにしたがって、徐々に $\mu\,(=145)$ に近づいていますか?
’ここのところが、大数の法則の視覚化です

```
    mean.xyline(o=reverse)
    smpl @all
```

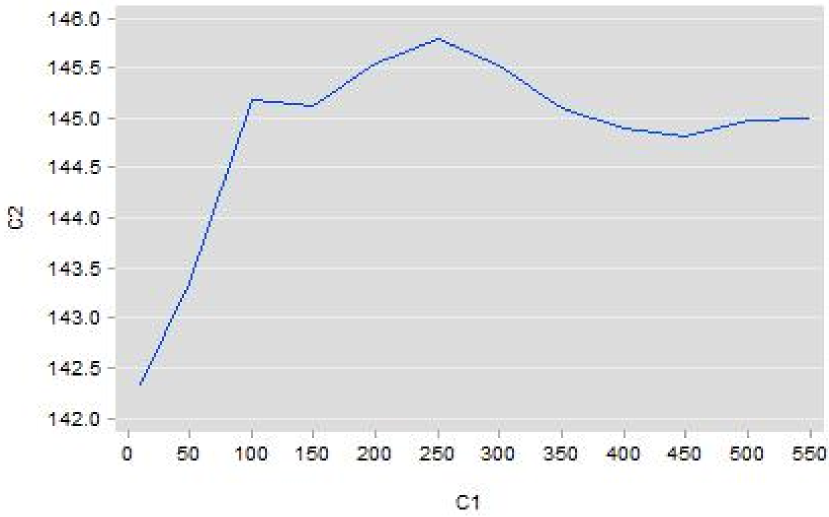

図 1.12 個の $\bar{X}$ のグラフ. 横軸は標本数

標本数が増えるごとに $\bar{X}$ が 145 に近づく様子が見てとれます。標本数が 10 の時の平均は 142.5 より若干小さい程度になっていますが、この値を生成しているのは、行列オブジェクト LLN の左上のグラフです。

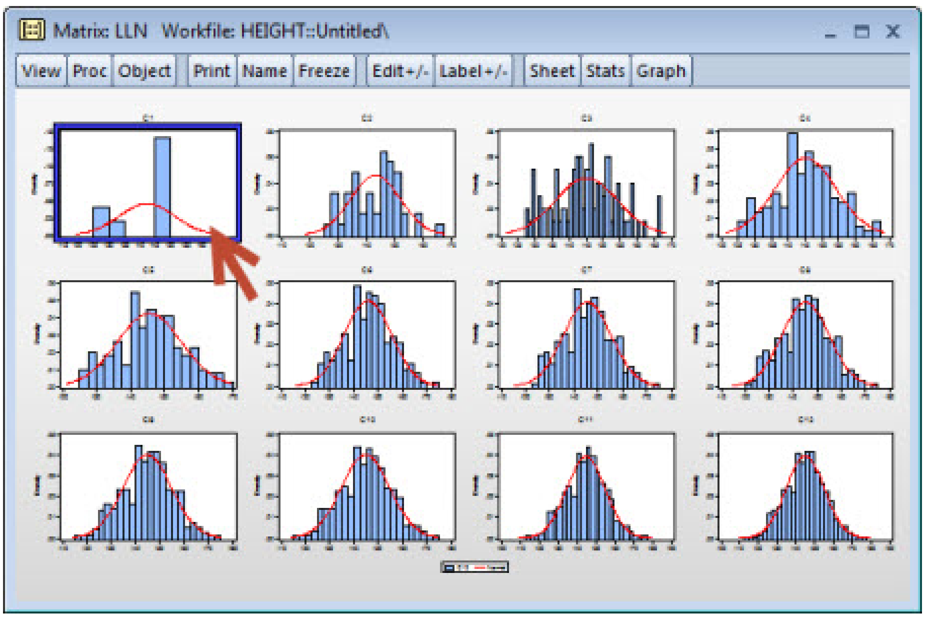

図 2.10, . . . , 550 個の標本によるヒストグラム

12 個のグラフを左上から、右下に向かって 1 行ごとに眺めていくと、標本数が増え、ヒストグラムの形状が分布曲線に近くなり、ヒストグラムの平均値が母平均に近くなっています。最後に EViews の機能を一つご紹介します。行列オブジェクトによる図 2 のグラフや、一般的なグループオブジェクトで同様のグラフを作成すると、グラフの個数が多すぎて、個々のグラフが非常に小さくなってしまう場合があります。そのような場合は、図 2 で目的のグラフをクリックし、サブメニューから Extract selected graph コマンドを実行します。

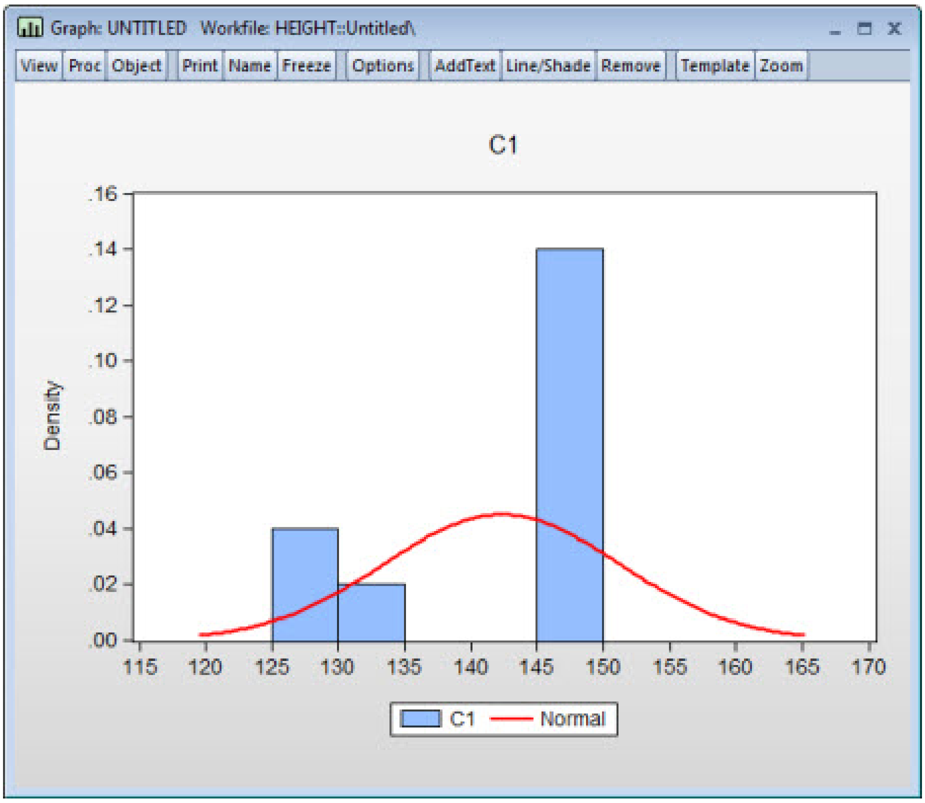

図 3.Extract selected graph コマンドで取り出したグラフ

このコマンドを利用すると、目的のグラフだけを単独の Graph オブジェクトとして取り出すことができます。

前回まで、賃金や株価など観測したデータについての分布についてモーメントという角度から、EViews を利用して学んできました。お話を確率分布の一つである t 分布に進める前に、今回は平均値 (一次のモーメント) について、非常に重要な特徴を解説しました。「データの分布」から「推定値の分布」にトピックを移していきます。

次回は平均値 (推定値) の分布に関する中心極限定理について、やはり EViews のプログラミング機能を活用して、ご解説したいと思います。

参考文献：東京大学教養学部統計学教室 編 「統計学入門」　東京大学出版会