中心極限定理について

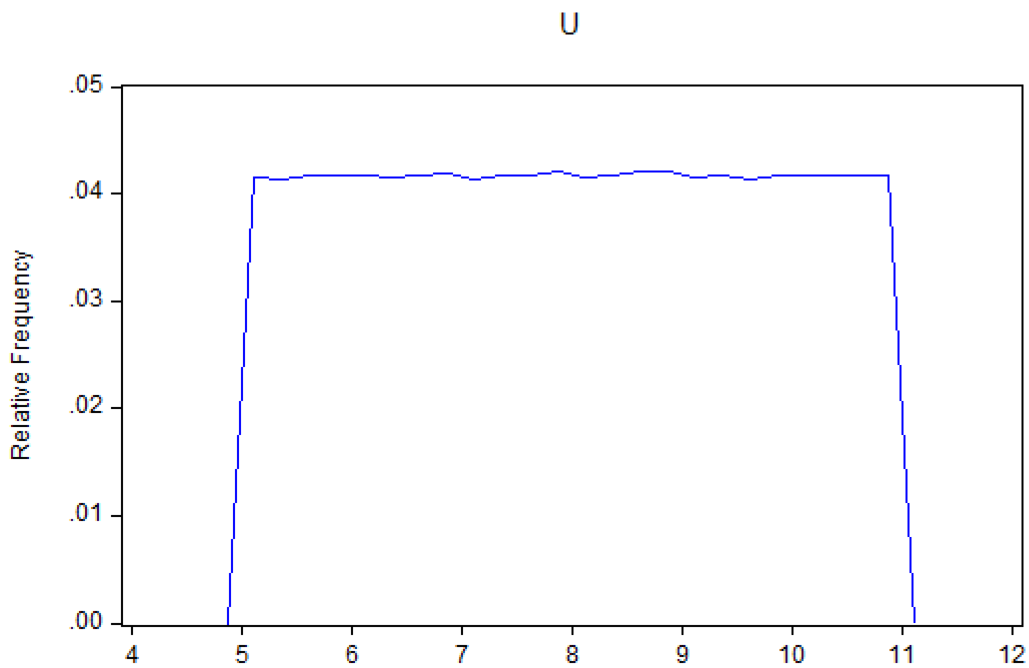
株式会社ライトストーン

前回は大数の法則の働きを、EViews 上でプログラムを走らせることにより、実際に目で見て確認しました。今回は統計学において大数の法則と並んで重要である中心極限定理の働きを、簡単なプログラムを使って確認したいと思います。

1 中心極限定理とは

確率分布から乱数を大量に生成し、それらの乱数をヒストグラムとして描画してみると、真の確率分布に近いグラフが描画されるのは直感的にご理解いただけるかと思います。たとえば、以下のコマンドを順に実行して確認してみます。

```
rndseed 12345
wfcreate(wf=CLT) u 1000000
series u =5+(11-5)*rnd
u.distplot(s) freqpoly(anchor=0, scale=relfreq)
```



これは乱数を生成した分布である、5 から 11 までの一様分布の特徴に非常によく似たグラフとなっています。

次に、一様分布から乱数を N 個生成し、その標本平均を取ることを考えてみます。このような操作を何セットも繰り返すと、標本平均がいくつも得られます。確率変数の平均なので、標本平均も確率変数となるわけですが、では標本平均はどのような分布に従うでしょうか。直感的には、標本平均の分布の平均値は、母平均と等しくなりそうです。一方でばらつきに関しては、平均を取る過程で個々の乱数の影響が中和されますので、より小さくなりそうです。中心極限定理を用いると、これらの直感をフォーマルに表現することが可能になります。

中心極限定理　$\mathrm{X}_i(\mathrm{i} = 1, 2, 3, \cdots \mathrm{N})$ が互いに独立に同一の分布 (平均 μ, 分散 σ^2) に従うとします。また、標本平均が以下で定義されるとします。

$$\overline{X}_N = \frac{1}{N}(X_1 + X_2 + \cdots + X_N) \tag{1}$$

中心極限定理によると、N が無限に大きくなっていく場合に、標本平均について以下の性質が成立します。

$$\overline{X}_N \sim N(\mu, \frac{\sigma^2}{N}) \tag{2}$$

すなわち、標本平均は平均が μ, 分散が $\frac{\sigma^2}{N}$ の正規分布に漸近的に従います。N が大きくなればなるほど、分散である $\frac{\sigma^2}{N}$ は小さくなっていきます。つまり、N が増えれば増えるほど、山の中心部分が非常に steep な分布になっていきます。一方で平均は母平均と変わりません。
また、確率分布の性質から以下の表現は (2) と同値です。

$$\frac{\overline{X}_N - \mu}{\frac{\sigma}{\sqrt{N}}} \sim N(0, 1) \tag{3}$$

つまり、(3) の統計量は標準正規分布に従います。以下では、(3) に基づいて、中心極限定理が本当に成立しているのかを EViews を用いて確認してみます。

2　EViews による確認

細かい説明は後回しにし、まずはプログラムを実際に動かしてみたいと思います。EViews を立ち上げ、新しい EViews プログラムを作成し、以下のコマンドを記述してください。ワークファイルは自動で新規作成されますので、特に開く必要はありません。プログラムファイルは、`File>New>Program` より作成できます[*1]。また、弊社で作成したプログラムファイル clt.prg をダウンロードしてお使いいただくことも可能です。

```
!n = 10000
```

[*1] 恐れ入りますが、学生版にはプログラムの作成・実行機能はございません。

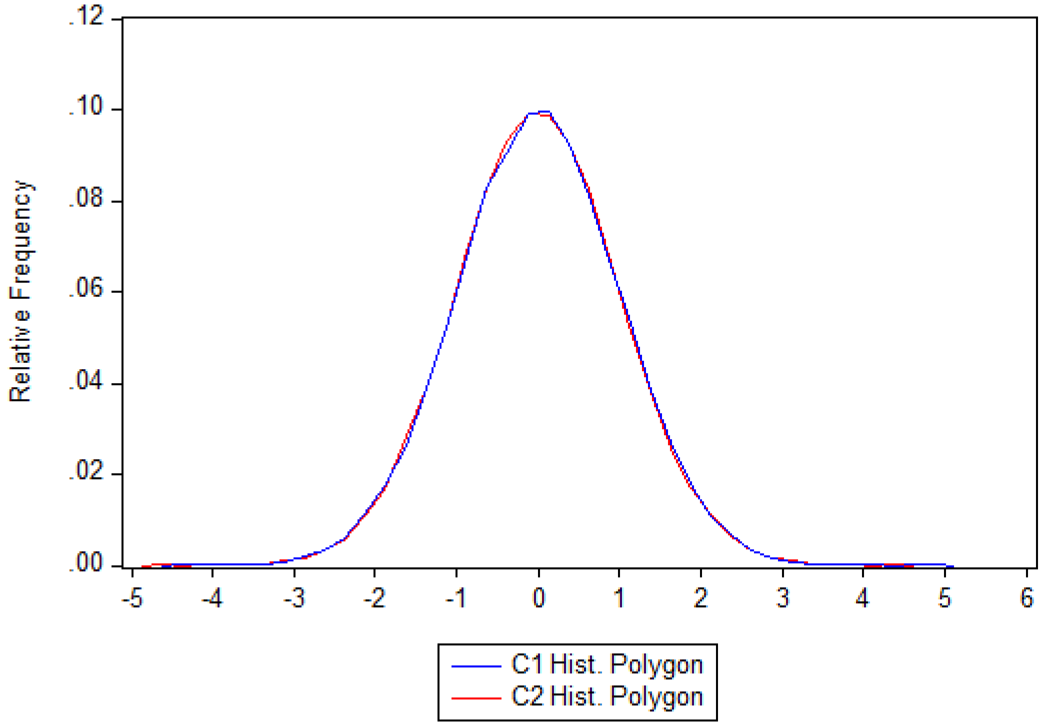
```
rndseed 12345
wfcreate(wf=CLT) u !n
!ex = (11+5)/2
!vx = (11-5)^2/12
!temp = !vx/!n
!denom = @sqrt(!temp)
!rep = 100000
matrix(!rep,2) stats
for !i=1 to !rep
    series u =5+(11-5)*rnd
    scalar mean = @mean(u)
    scalar stat = (mean - !ex)/!denom
    stats(!i,1)=stat
    stats(!i,2)= @nrnd
next
stats.distplot(s) freqpoly(anchor=0, scale=relfreq)
```

記述が終わりましたら、まずはプログラムを保存し、Run ボタンを押してみてください。実行にはそれなりの時間がかかりますので、しばらくそのままお待ちください。
プログラムが上手く動作しない場合、記述を見直して修正してください。典型的には、以下のような問題が考えられます。

- 関数名やシリーズ名の打ち間違い
- 必要な位置に半角スペースが挿入されていない/入力してはいけない位置に半角スペースが入力されている（意味のまとまり毎に半角スペースを入れてください）
- 括弧が閉じていない
- 全角スペースや全角括弧などの全角文字が入力されている

これらをチェックしても動かないようであれば、弊社で作成した clt.prg をダウンロードし、実行してください。

ワークファイルが作成され、以下のようなグラフが表示されます。



青が統計量の分布、赤が標準正規分布の分布です。中心極限定理の主張の通り、元々一様分布から発生させた乱数を使っているにも関わらず、(3) の統計量は標準正規分布とほぼ同じように分布していることが分かります。中心極限定理は確かに成立しているようです。

以下では、上記のプログラムの内容について、パーツ毎に簡単な解説を行います。

```
!n = 10000
rndseed 12345
wfcreate(wf=CLT) u !n
```

- サンプルサイズを 10000 にセットします
- 乱数の発生に使うシード値を 12345 に設定します（本質的には不要ですが、本資料の画像と、操作時の結果を一致させるために設定しています）
- CLT という名前でワークファイルを作ります。サンプルサイズを!n(すなわち、10000) に指定しています

```
!ex = (11+5)/2
!vx = (11-5)^2/12
!temp = !vx/!n
!denom = @sqrt(!temp)
```

- 後で 5 から 11 までの一様分布から乱数を生成しますので、一様分布の平均と分散の公式より、順番に母平均、母分散、(3) の分母を計算しています

```
!rep = 100000
matrix(!rep,2) stats
```

- 計算回数!rep を 100000 回に設定します
- アウトプットを格納する行列 stats(10 万行 ×2 列）を作ります

```
for !i=1 to !rep
    series u =5+(11-5)*rnd
    scalar mean = @mean(u)
    scalar stat = (mean - !ex)/!denom
    stats(!i,1)=stat
    stats(!i,2)= @nrnd
next
```

- !rep 回だけ、同じ計算を行い、行列 stats に格納していきます
- 5 から 11 までの一様分布から乱数を生成します。一回の計算毎に、!n 個の乱数が生成されることになります。
- 生成した乱数の標本平均を計算します
- 標本平均から母平均を引き、!denom で割ることで、(3) の統計値が得られます
- 行列の (!i, 1) 要素に先ほど計算した統計値を格納します
- 同時に、行列の (!i, 2) 要素に N(0, 1) の分布に従う乱数値を格納します

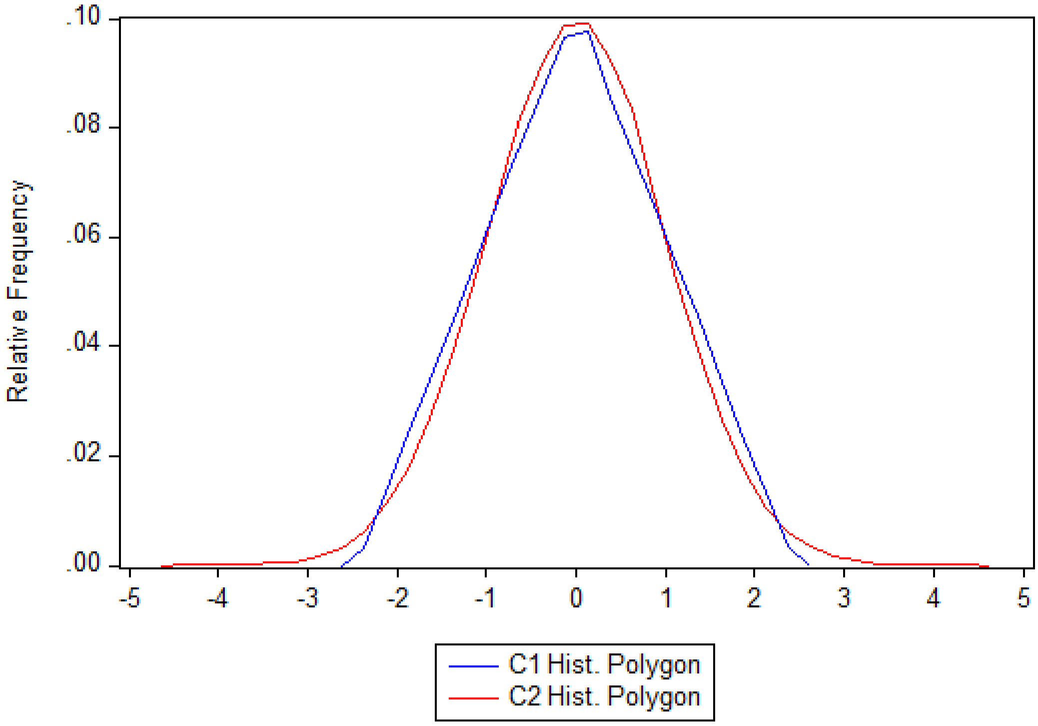
```
stats.distplot(s) freqpoly(anchor=0, scale=relfreq)
```

- ここまでの処理で行列 stats の 1 列目には統計値、2 列目には正規分布に従う乱数が格納されています。それらを分布として一つのグラフに表示します。

プログラムの中身の説明は以上の通りです。統計学の教本などを基に、プログラムをアレンジして実行してみると、統計学や EViews への理解が深まると思います。ここではその一例として、サンプルサイズを変更してみたいと思います。中心極限定理はサンプルサイズが大きくなっていくときに漸近的に成立する定理ですので、サンプルサイズが小さいと上手く機能しないことが予想されます。最後にこの点を確かめるため、少々極端ですが、プログラムの最初で

```
!n = 2
```

とし、プログラムを実行してみます。



統計量の分布が正規分布からかけ離れている様子が確認できます。

今回は EViews のプログラミング機能を用いて中心極限定理について検証しました。本文書ではデータの生成に一様分布を用いましたが、中心極限定理は一様分布以外でももちろん成立します。一様分布以外からデータを生成し、プログラムを走らせることは良い演習になります。その際は、乱数の生成部分だけでなく、母平均や母分散の計算部分も変更しなければならないことにお気を付けください。

内容は未定ですが、来月からも引き続き EViews の技術資料を公開して参ります。メールマガジンと合わせてご覧いただけると幸いです。