2014 年 5 月
株式会社　ライトストーン

確率分布の復習

1　確率変数

今回は確率分布の基本的な事柄を復習します。教科書にあるような記述が多くなりますが、EViews のコマンドを利用して数学の記号や意味を理解していきましょう。

1.1　離散型確率変数

まず、確率変数 X がとりうる値が離散的な (サイコロの目のように決まった値しかでない) 場合の離散型確率変数について次の 2 点を確認しておきます。

$$f_X\left(x\right) = \mathrm{P}\left(X = x\right) \tag{1}$$

$f_X\left(\cdot\right)$ のことを X の確率関数や確率分布と呼び、次の式を満たします。

$$\sum_{k=1}^{\infty} f_X\left(x_k\right) = 1 \tag{2}$$

1.2　連続型確率変数

一方、投資収益率や経済成長率のような連続型確率変数の場合、確率変数 X と確率には次のような関係があります。

$$\mathrm{P}\left(a \le X \le b\right) = \int_a^b f_X\left(x\right) dx \tag{3}$$

$X \le x$ という事であれば、

$$\mathrm{P}\left(X \le x\right) = F_X\left(x\right) = \int_{-\infty}^{x} f_X\left(x\right) dx \tag{4}$$

$F\left(x\right)$ のことは累積分布関数と呼びます。

離散型の場合は $f_X\left(\cdot\right)$ のことを X の確率関数と呼びましたが、連続型の場合は、これを確率密度関数と呼びます。2 式に対応する式として、次の式が成立します。

$$\int_{-\infty}^{\infty} f_X\left(x\right) dx = 1 \tag{5}$$

2　標準正規分布

ここからは EViews の関数を利用して連続変数確率変数 X のプロットを作成してみましょう。今、$X \sim N\left(\mu, \sigma^2\right)$ とします。この時、正規分布の密度関数は次のようになります。

$$f_X = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right) \tag{6}$$

したがって、

$$\mathrm{P}\left(X \le x\right) = F_X\left(x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right) dx \tag{7}$$

となります。EViews の関数で用意されている`@dnorm()`は $N\left(0,1\right)$ の時の標準正規分布で、密度関数は次のようになります。

$$f_X = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{8}$$

$N\left(\mu,\sigma^2\right)$ は正規分布の分布を示す記法で、μ は平均、σ^2 は分散です。よって、$N\left(0,1\right)$ は平均 0、分散 1 の正規分布を示します。$N\left(0,1\right)$ の累積分布関数は 4 式にならって、

$$\mathrm{P}\left(X \le x\right) = F_X\left(x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \tag{9}$$

となります。

それではここで、`@dnorm()` を利用して標準正規分布のグラフを作成してみましょう。まずは次のコマンドをプログラムファイルに記述して、8 式の標準正規分布を描いてみましょう。$-8 \le X \le 8$ とします。前回までの資料をご覧いただいていれば、コマンドは特に難しいものはありませんので解説はいたしません。

```
wfcreate(wf=dist) u 1601
series x=-8
smpl 2 @last
x=x(-1)+0.01
smpl @all
series fn=@dnorm(x)
group group01 x fn
show group01.xyline
```

この時のグラフは次のようになります ($x = 2$ の所の縦線はつきません)。

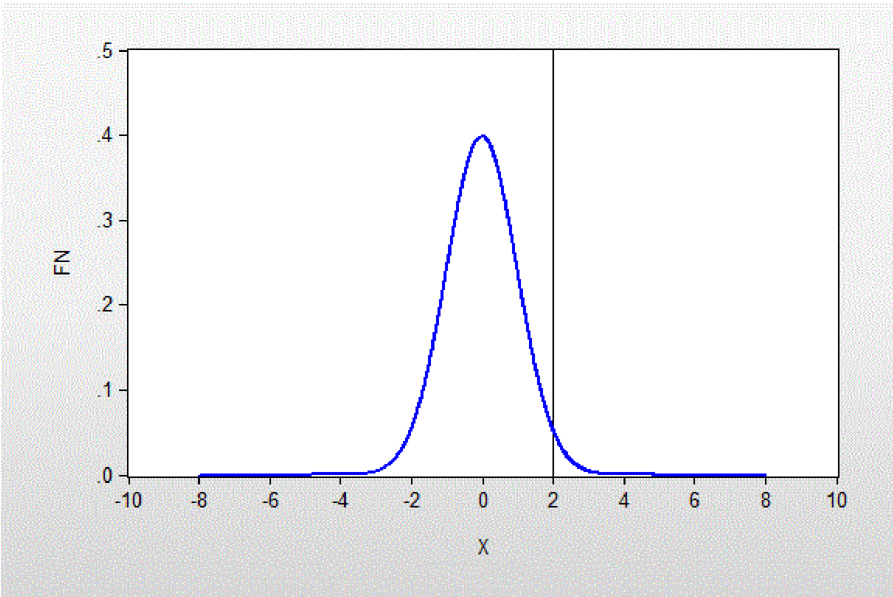

図 1. 標準正規分布

グラフに示した $x = 2$ の所に線から**右側**の部分 (曲線下) の面積はいくつになるでしょうか。つまり、

$$\mathrm{P}\left(X \ge 2\right) = \int_{2}^{\infty} f_X\left(x\right) dx$$

を計算したいと思います。標準正規分布の曲線が左右対称であることを利用して、次のように計算できます。

$$\mathrm{P}\left(X \ge 2\right) = \mathrm{P}\left(X \le 2\right) = F_X\left(2\right) = \int_{-\infty}^{-2} f_X\left(x\right) dx$$

8 式と `@dnorm()` の対応関係と同じく面積 (確率) については、7 式に対応する `@cnorm` というコマンドが用意されています。これを利用すれば、上式の面積はコマンドウィンドウを利用して次のように求めることができます。

```
show @cnorm(-2)
```

3 標準化

3.1 グラフの作成

それでは次に、$X \sim N\left(2,3^2\right)$ の場合について考えてみましょう。これをグラフ化したいのですが、どのようにしたら良いでしょうか?コマンド`@dnorm`をうまく利用して $X \sim N\left(2,3^2\right)$ 図をプロットしたいと思います。まずは、6 式に $\mu=2, \sigma^2=3^2$ を入力してみましょう。

$$f_X=\frac{1}{3} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-2)^2}{2 \cdot 3^2}\right) \tag{10}$$

となりますが、exp () についてうまく整理する必要があります。そこで、次の式を利用して変数変換を行います。

$$z=\frac{x-\mu}{\sigma} \tag{11}$$

このようにする事で、変数 X を $N(0,1)$ の変数 Z に変換できます。この変換のことを標準化 (または正規化) と呼びます。標準化により 6 式は次のように書き換えられます。

$$f_Z=\frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{12}$$

これで`@dnorm()`が利用できる形になりました。よって、次のようにプログラムを一部、変更します。

```
wfcreate(wf=dist) u 1601
series x=-8
smpl 2 @last
x=x(-1)+0.01
smpl @all
series fn=@dnorm(x)
’次のコマンドを追加します
series z=(x-2)/3
series fz=(1/3)*@dnorm(z)
’グループメンバに fz を加えます
group group01 x fn fz
show group01.xyline
```

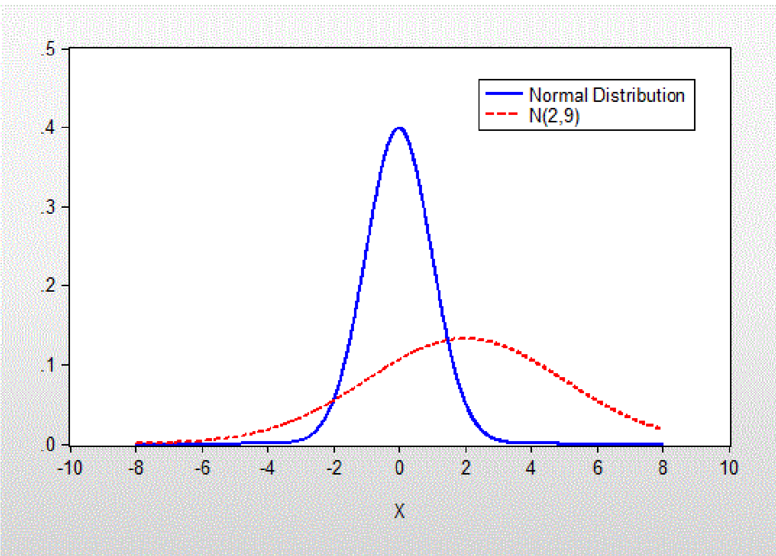

図 2.N(0,1) と N(2,9) の密度関数のプロット

3.2　確率の計算

さて、$X \sim N\left(2, 3^2\right)$ の正規分布について確率 $P\left(X \leq 5\right)$ を求めてみましょう。つまり、

$$F_X\left(5\right) = \int_{-\infty}^{5} \frac{1}{\sqrt{2\pi 3^2}} \exp\left(-\frac{\left(x-2\right)^2}{2 \cdot 3^2}\right) dx \tag{13}$$

を求めます。この場合、標準化した z を利用して計算を行います。最初に 11 式で $x = 5$ を変換します。

$$z = \frac{5-2}{3} = 1$$

よって次の式を計算すればよいということになります。

$$\begin{aligned} F_Z\left(1\right) &= \int_{-\infty}^{1} \frac{1}{3} \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \cdot 3dz \\ &= \int_{-\infty}^{1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \end{aligned} \tag{14}$$

よって、次のコマンドで目的の値を知ることができます。

```
show @cnorm(1)
```

`@dnorm()` と `@cnorm()` はEViews でプログラミングを行う際に必ず覚えておきたいコマンドです。分布関数の `@cnorm()` を利用する際は、$\mathrm{P}\left(X \leq x\right) = F_X\left(x\right)$ という関係から点 x より左側を積分するという、イメージを思い浮かべるようにしてください。

4　対数正規分布

さて、最後に前回紹介した歪度 (わいど) と尖度 (せんど) について今一度、確認しておきましょう。計算式は前回示しましたので、ここでは具体的な数値を利用して、この 2 つの統計量を確認します。つぎのコマンドをコマンドウィンドウに入力します。

```
' 乱数キー 123 に設定します。
  rndseed 123
' 対数正規分布の密度関数を利用して乱数を作成します。
' 乱数キーを設定していますので、いつでも同じ結果を再現できます。
' 平均 m = 6, 標準偏差 s = 0.4 とし、賃金 rent のサンプルデータを作成します。

  series rent=@rlognorm(6,0.4)
  show rent.hist
```

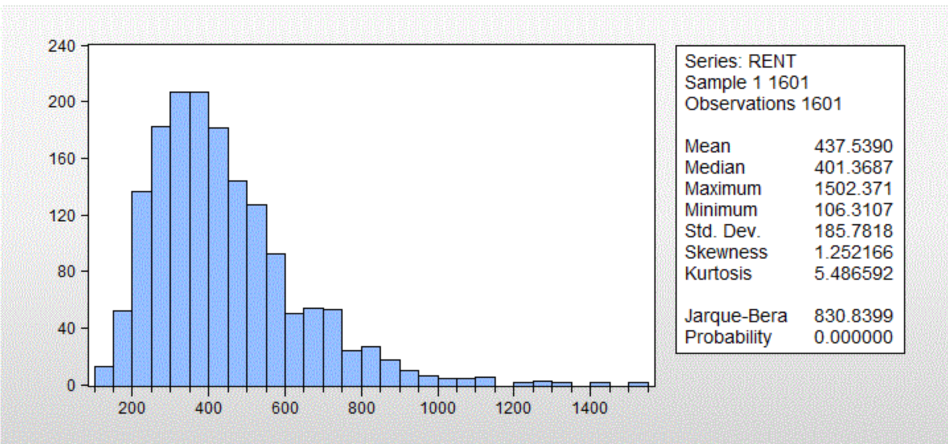

図 3. 仮想賃金のヒストグラム

平均 $6 = \log(\mathrm{rent})$ ですから、実際の賃金は $e^6 \simeq 403$ 万円程度になります。グラフの右側にある統計値を見ますと、賃金平均が 437 万円で、標準偏差が 185 万円となっていることが分かります。見るからに標準正規分布とは異なりますが、平均と標準偏差の情報だけでは、この図のような分布は想像できません。正規分布では歪度は 0、右裾が長くなると正、逆に左裾が長い場合は負の値を取ります。確率変数 X の歪度は、前回紹介した式を期待値オペレータ E を使って書くと次のようになります。

$$\mathrm{Skewness} = \frac{\mathrm{E}\left[\{X - \mathrm{E}(X)\}^3\right]}{\{\mathrm{V}(X)\}^{3/2}} \tag{15}$$

一方、分布の尖り具合を示す尖度の定義は、

$$\mathrm{Kurtosis} = \frac{\mathrm{E}\left[\{X - \mathrm{E}(X)\}^4\right]}{\{\mathrm{V}(X)\}^{2}} \tag{16}$$

となり、正規分布では 3 となります。この仮想賃金データの場合、尖度は 5.48 ですから、ピークの周辺により多くのデータが集まっていることが分かります。ご存知のように平均は $\mathrm{E}(X)$, 分散は $\mathrm{V}(X) = \mathrm{E}\left[(X - \mathrm{E}(X))^2\right]$ と表現することができます。よって、平均の事を 1 次のモーメント、分散は 2 次、歪度は 3 次、尖度は 4 次のモーメントと呼びます。これらの情報を利用すると確率分布の概略が凡そ理解できます。

それでは最後に rent の対数を取ったグラフを作成し、正規分布することを確認しましょう。

```
show log(rent).hist
```

グラフの作成は皆様におまかせします。次回は最後にでてきました「モーメント」という言葉について、基礎知識を確認しましょう。

■