経済統計学講義ノート No.5

連続確率分布

蛭川雅之

2022 年 9 月 12 日

目　次

1. 連続確率変数
2. 複数の連続確率変数
3. 正規分布

1　連続確率変数

1.1　連続確率変数の性質

- 連続確率変数はある区間（例：$[0,1]$, $(-\infty,\infty)$）内の任意の実数をとる。
- 連続確率変数が特定の一点をとる確率をゼロとする。
  - なぜ？
  - 区間内の全ての実数値に離散確率変数と同様の確率を定義すると、全事象の確率が 1 を超えてしまう。
- 連続確率変数の確率は区間に対して定義される。

1.2 分布関数

- 連続確率変数 X に対して関数 $F_X(x)$ を $F_X(x) = \Pr(X \leq x)$ と定義する。
  - X が区間 $[a, b]$ $(a \leq b)$ に入る確率は次のように表現される。

$$\begin{aligned} \Pr(a \leq X \leq b) &= \Pr(a < X \leq b) \\ &= \Pr(X \leq b) - \Pr(X \leq a) \\ &= F_X(b) - F_X(a) \end{aligned}$$

  - $F_X(x)$ は連続確率変数 X の（累積）分布関数と呼ばれる。
- 分布関数は以下の性質を持つ。
  1. $0 \leq F_X(x) \leq 1$
  2. $F_X(-\infty) = 0, F_X(\infty) = 1$
  3. $x < x^*$ を満たす x, x^* に対し、$F_X(x) \leq F_X(x^*)$

1.3　確率密度関数

- 離散確率変数の確率関数に相当するものとして、分布関数 $F_X(x)$ の微小な増加分を表す確率密度関数を

$$f_X(x) = \frac{d}{dx}F_X(x) = \lim_{dx \to 0} \frac{F_X(x+dx) - F_X(x)}{dx}$$

と定義する。

- 確率密度関数は以下の性質を持つ。
  1. $f_X(x) \geq 0$
  2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

1.4 分布関数と確率密度関数との対応

- 確率密度関数の定義から

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$$

が成り立つ。

- ところで、

$$\Pr(a < X \le b) = F_X(b) - F_X(a)$$

であった。

  – この確率を確率密度関数を用いて表現すると

$$\Pr(a < X \le b) = \int_{-\infty}^{b} f_X(x)\,dx - \int_{-\infty}^{a} f_X(x)\,dx$$

となる。

1.5　期待値・分散・標準偏差

- 連続確率変数 X の期待値（平均値）および分散はそれぞれ

$$E(X) = \mu = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$Var(X) = \sigma^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx$$

と定義される。

- 連続確率変数 X の標準偏差はこの分散の正の平方根

$$\sigma = \sqrt{Var(X)}$$

である。

2　複数の連続確率変数

2.1　同時分布・周辺分布

- 連続確率変数 X と Y の同時分布関数は
$$F_{X,Y}(x,y) = \Pr(X \le x, Y \le y)$$
と定義される。
- 連続確率変数 X と Y の同時密度関数は
$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$
と定義される。
  – 同時分布関数と同時密度関数は以下のように対応する。
$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du$$

- X および Y の周辺密度関数はそれぞれ

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

である。

  – X および Y の周辺分布関数はそれぞれ

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(u,y)\,dydu$$
$$F_Y(y) = \int_{-\infty}^{y} f_Y(v)\,dv = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f_{X,Y}(x,v)\,dxdv$$

  で与えられる。

2.2 共分散・相関係数・独立性

- 連続確率変数 X と Y の期待値を以下のように表記する。

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f_X(x)\, dx$$
$$E(Y) = \mu_Y = \int_{-\infty}^{\infty} y f_Y(y)\, dy$$

- このとき、連続確率変数 X と Y の共分散および相関係数はそれぞれ以下のように定義される。

$$\begin{aligned} Cov(X,Y) &= E\{(X-\mu_X)(Y-\mu_Y)\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f_{X,Y}(x,y)\, dx dy \end{aligned}$$

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

- 連続確率変数 X と Y について、

$$F_{X,Y}(x,y) = F_X(x) F_Y(y)$$

  が成り立つとき、X と Y は互いに独立である（もしくは独立に分布している）という。
  - X と Y が独立ならば

    $$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

    も成り立つ。
  - X と Y が無相関であっても独立でない場合はあるが、X と Y が独立ならば必ず無相関である。

2.3　条件付き確率密度関数・条件付き期待値

- 連続確率変数 X と Y について、$X = x$ を与えたときの Y の条件付き確率密度関数は以下のように定義される。

$$f_{Y|X}\left(y|\,x\right) = \frac{f_{X,Y}\left(x,y\right)}{f_X\left(x\right)}$$

- この条件付き確率密度関数を用いて、$X = x$ を与えたときの Y の条件付き期待値

$$E\left(Y|\,X = x\right) = \int_{-\infty}^{\infty} y f_{Y|X}\left(y|\,x\right) dy$$

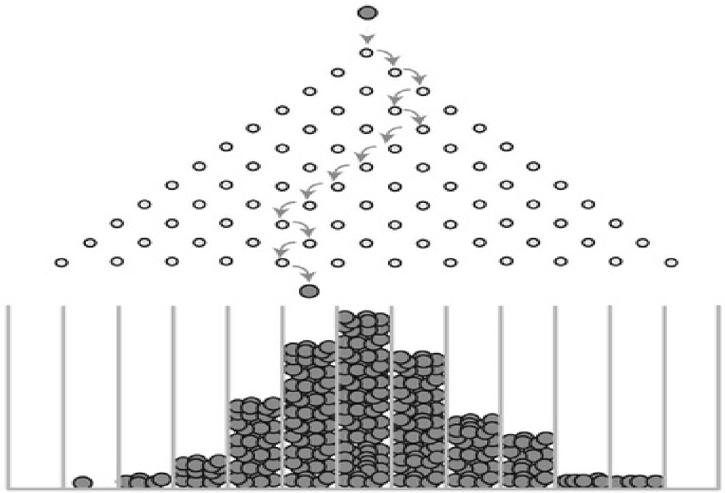
が得られる。

  – 期待値繰り返しの法則 $E\left(Y\right) = E\left\{E\left(Y|\,X\right)\right\}$ も成り立つ。

3　正規分布

3.1　最も重要な分布

- 正規分布は理論的にも実用的にも最も重要な分布である。
  - 中心極限定理
  - Galton Board（下図参照）



3.2 正規分布の確率密度関数

- 確率変数 X が平均 μ と分散 σ^2 の正規分布に従うとき、X の確率密度関数は

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

で与えられる。

  – 確率変数 X が平均 μ と分散 σ^2 の正規分布に従うことを

$$X \sim N\left(\mu, \sigma^2\right)$$

と表記する。

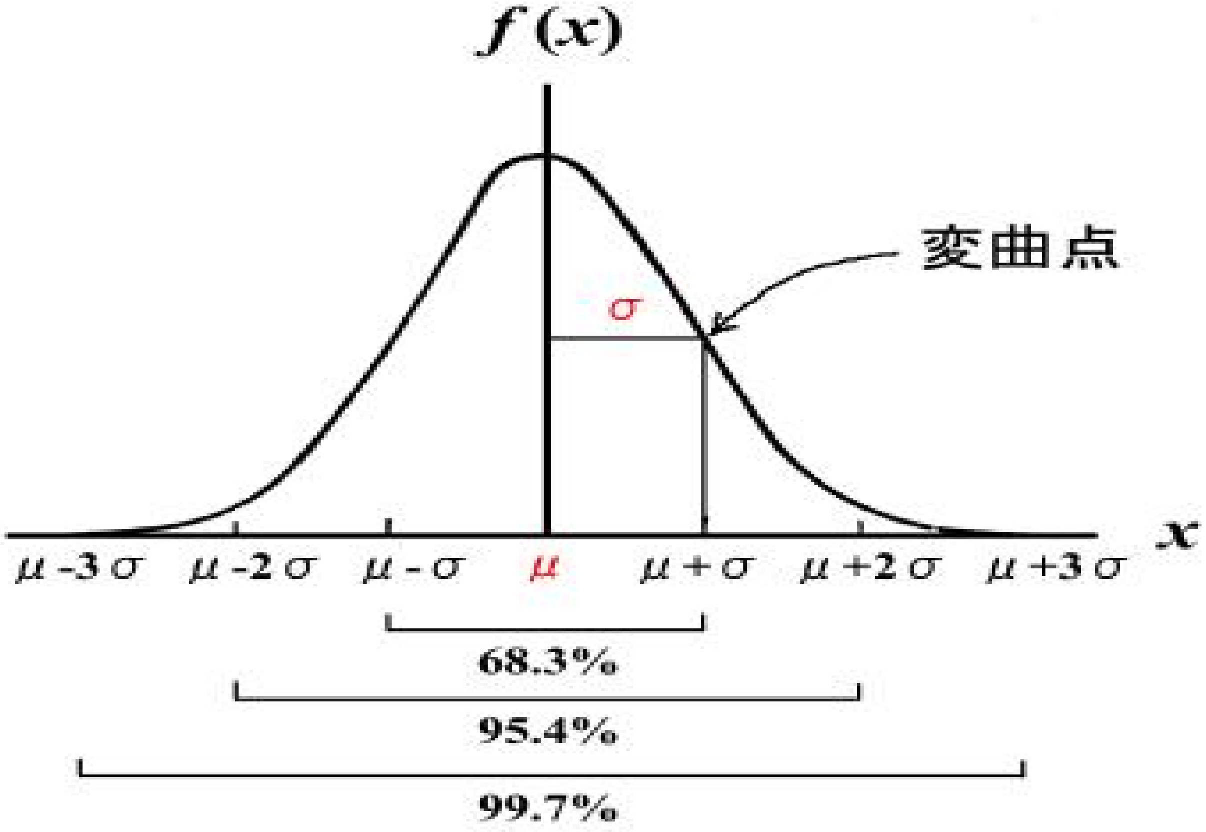
  – 当然のことながら

$$E(X) = \mu, Var(X) = \sigma^2$$

である。

3.3　正規分布の形状

- 平均 μ に関して左右対称の '釣鐘型' をしている。



3.4 標準正規分布

- 平均 $\mu = 0$ と分散 $\sigma^2 = 1$ の正規分布 $N(0,1)$ を**標準正規分布**という。
  - しばしば標準正規分布に従う確率変数を Z、さらに、Z の確率密度関数および分布関数を以下のように表記する。

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$\Phi(z) = \Pr(Z \leq z) = \int_{-\infty}^{z} \phi(u)\, du$$

- $Z \sim N(0,1)$ に対し、以下が成り立つ（ので記憶しておくとよい）。

$$\Pr(-1.65 \leq Z \leq 1.65) \approx 0.90$$

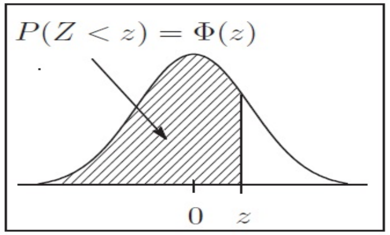
$$\Pr(-1.96 \leq Z \leq 1.96) \approx 0.95$$

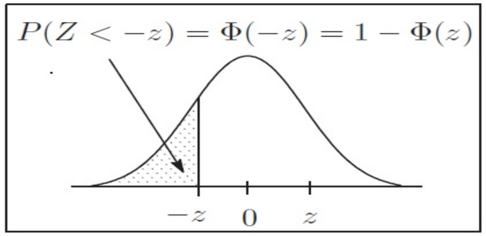
$$\Pr(-2.58 \leq Z \leq 2.58) \approx 0.99$$

3.5 正規分布の確率計算

■直接的な確率計算

- 標準正規分布に従う確率変数 Z が区間 $[a, b]$ $(a \leq b)$ に入る確率 $\Pr(a \leq Z \leq b)$ を求めるには、正規分布表もしくは Excel の関数 `normsdist`（以下のタイプの正規分布表に対応）を利用すればよい。





$Z \sim N(0,1),\ \ P(Z < z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133

- 一般の正規分布の場合はどうするか？
  - $X \sim N\left(\mu, \sigma^2\right)$ を標準化すると

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

となることを利用する。

  - 具体的には、$X \sim N\left(\mu, \sigma^2\right)$ であるとき、

$$\begin{aligned}\Pr(a \leq X \leq b) &= \Pr\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\end{aligned}$$

となることから、この右辺に正規分布表もしくは `normsdist` を適用する。

■分位点の導出

- しばしば、$X \sim N\left(\mu, \sigma^2\right)$ および 0 以上 1 以下の定数 q に対し、

$$\Pr\left(X \leq c\right) = q$$

  を満たす c を求める必要が生じる。
  - この点 c は $N\left(\mu, \sigma^2\right)$ の 100q% 分位点と呼ばれる。
- c を求めるには以下の手順に従う。
  1. 正規分布表もしくは `normsinv` を利用して、q に対応する標準正規分布上の 100q% 分位点 c' を求める。
  2. 左辺が $\Phi\left(\frac{c-\mu}{\sigma}\right)$ と変形できることに注目し、

$$\frac{c-\mu}{\sigma} = c' \Rightarrow c = \mu + c'\sigma$$

  を得る。

問題 1 $X \sim N(78, 64)$ であるとき、次の確率を求めよ。

1. $\Pr(60 \leq X \leq 72)$
2. $\Pr(X \geq 88)$

問題 2 ある公務員の採用数が 120 名のところ、1500 名が採用試験を受験した。この公務員採用試験は 600 点満点で、受験者の得点分布は平均 365 点、標準偏差 45 点の正規分布とみなして差し支えないことが分かった。このとき以下の問いに答えよ。

1. 400 点以上得点した受験者は何名か。
2. 合格最低点は何点か。