応用計量経済分析 TA セッション
第 5 回：標本平均の分布

TA：北村友宏 *

2015 年 11 月 10 日

1　標本分布

- 母集団から個体 n 個 $(X_1, X_2, \ldots, X_n)$ を無作為抽出して標本を得る。
  - ⋆ e.g., 日本全国の大学生の 1 か月間の携帯電話利用料金の特徴を調べるため、その中から 100 人分の利用料金 $(X_1, X_2, \ldots, X_{100})$ を無作為に抽出する。
- どの n 個を抽出するかにより、各個体 $X_1, X_2, \ldots, X_n$ の値は異なる。
  - ⋆ e.g., どの大学生 100 人を抽出するか、また、1 人目が誰で、2 人目が誰で、・・・、100 人目が誰になるかによって、1〜100 人目の各大学生の料金という数値が異なる。

  ⇒どの値が実現するかは確実にはわからないが、とりうる値それぞれについて、その値が出る確率が振られている。

  ⇒ $X_1, X_2, \ldots, X_n$ は確率変数。

  ⇒統計量である **標本平均**（sample mean）$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ なども確率変数。
- 無作為抽出ならば、$X_1, X_2, \ldots, X_n$ は互いに独立。
- 統計量の確率分布を**標本分布**（sample distribution）とよぶ。
- 標本平均の確率分布の特徴（期待値や分散）を知りたい。

  ⇒ $E\left(\overline{X}\right) = E\left(\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n\right)$ や $V\left(\overline{X}\right) = V\left(\frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n\right)$ を計算する。

2　多次元の確率変数の期待値と分散

- 確率変数 X, Y と定数 a, b について、

$$
\begin{aligned}
E(aX + bY) &= aE(X) + bE(Y), \\
V(aX + bY) &= a^2V(X) + b^2V(Y) \underbrace{+2abCov(X,Y)}_{\textbf{注意！}}
\end{aligned}
$$

  が成り立つ（証明略）。

---

* Email: kitamu.tom@gmail.com　URL: http://tomkitamura.html.xdomain.jp

- X と Y が独立ならば X と Y は無相関、つまり $Cov(X,Y)=0$ となる。このとき、$aX+bY$ の分散は、

$$V(aX+bY)=a^2V(X)+b^2V(Y)$$

となる（証明略）。

- 確率変数 n 個、定数 n 個の場合も同様に考えることができる。互いに独立な確率変数 $X_1,X_2,\ldots,X_n$ と定数 $w_1,w_2,\ldots,w_n$ について、

$$\begin{aligned}E(w_1X_1+w_2X_2+\cdots+w_nX_n)&=w_1E(X_1)+w_2E(X_2)+\cdots+w_nE(X_n),\\ V(w_1X_1+w_2X_2+\cdots+w_nX_n)&=w_1^2V(X_1)+w_2^2V(X_2)+\cdots+w_n^2V(X_n)\end{aligned}$$

が成り立つ（証明略）。

  ⋆ 無作為抽出された標本 $(X_1,X_2,\ldots,X_n)$ であれば、上記の分散の性質を利用できる。

例題 1.　平均 μ、分散 σ^2 の母集団から抽出した無作為標本 $(X_1,X_2,\ldots,X_n)$ を考える。この標本は、$E(X_1)=E(X_2)=\cdots=E(X_n)=\mu, V(X_1)=V(X_2)=\cdots=V(X_n)=\sigma^2$ を満たす。このとき、$100\sum_{i=1}^{n}X_i$ の期待値と分散を求めなさい。

$100\sum_{i=1}^{n}X_i$ の期待値は、

$$\begin{aligned}E\left(100\sum_{i=1}^{n}X_i\right)&=E[100(X_1+X_2+\cdots+X_n)]\\&=E(100X_1+100X_2+\cdots+100X_n)\\&=100E(X_1)+100E(X_2)+\cdots+100E(X_n)\\&=100\mu+100\mu+\cdots+100\mu\\&=100(\underbrace{\mu+\mu+\cdots+\mu}_{n\text{ 個}})\\&=100n\mu.\end{aligned}$$

また、$100\sum_{i=1}^{n}X_i$ の分散は、

$$\begin{aligned}V\left(100\sum_{i=1}^{n}X_i\right)&=V[100(X_1+X_2+\cdots+X_n)]\\&=V(100X_1+100X_2+\cdots+100X_n).\end{aligned}$$

ここで、$(X_1,X_2,\ldots,X_n)$ は無作為標本なので、$X_1,X_2,\ldots,X_n$ は互いに独立である。よって、

$$\begin{aligned}V(100X_1+100X_2+\cdots+100X_n)&=100^2V(X_1)+100^2V(X_2)+\cdots+100^2V(X_n)\\&=10000\sigma^2+10000\sigma^2+\cdots+10000\sigma^2\\&=10000\left(\underbrace{\sigma^2+\sigma^2+\cdots+\sigma^2}_{n\text{ 個}}\right)\\&=10000n\sigma^2.\end{aligned}$$