10. 正規分布の再生性

1. 確率分布の再生性

● 定義（確率分布の再生性）

(1) 確率分布 F について，2 つの独立な確率変数 X, Y が F に従えば，それらの和

$$Z = X + Y$$

も常に F に従うとき，確率分布 F は再生性をもつという。

(2) すなわち，確率分布 F が再生性をもつとは，次が成立することである。

$$X \sim F,\ Y \sim F,\ X \text{ と } Y \text{ は独立} \implies X + Y \sim F$$

(1) 統計学には，代表的な確率分布が多数ある。二項分布 $B(n,\ p)$，ポアソン分布 $P_o(\lambda)$，正規分布 $N(\mu, \sigma^2)$，カイ 2 乗分布，$t-$分布，$F-$分布など，いろいろある。これらは，確率モデルでもある。

(2) 分野によって多少意味は異なるが，一般に「～モデル」という用語は，現実の複雑な現象を扱うために，その現象を抽象化して数学的に表現したものを指す。確率モデルとは，現実の母集団に適用される数学的に定義された確率分布のことである。そして，「この母集団は，二項分布モデルで扱うことができる」などという言い方をする。

(3) いま，二項分布や正規分布などの確率分布を，一般に F で表す。確率変数 X の確率分布が，F で定義されている確率分布になるとき，「X は確率分布 F に従う」といい，これを

$$X \sim F$$

で表す。X が正規分布 $N(\mu,\ \sigma^2)$ に従うことを

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

で表現したが，これと同じである。

(4) 確率分布の中には，上記の再生性という重要な性質をもつものがある。二項分布，ポアソン分布，正規分布などは，再生性をもつ。

2. 正規分布の再生性

● 定理（正規分布の再生性）

(1) 正規分布は，再生性をもつ。

(2) 確率変数 X, Y が独立で，$X \sim N(\mu_1, {\sigma_1}^2)$，$Y \sim N(\mu_2, {\sigma_2}^2)$ のとき，X, Y の 1 次結合 $aX + bY$（a, b は定数）は，正規分布に従う。すなわち，

$$aX + bY \sim N(a\mu_1 + b\mu_2,\ a^2{\sigma_1}^2 + b^2{\sigma_2}^2)$$

(1) 正規分布の再生性とは，2 つの独立な確率変数 X, Y がそれぞれ正規分布に従えば，それらの和 $Z = X + Y$ も常に正規分布に従うという性質である。

具体的には，X, Y が独立で，$X \sim N(\mu_1, {\sigma_1}^2)$，$Y \sim N(\mu_2, {\sigma_2}^2)$ ならば，必ず

$$X + Y \sim N(\mu_1 + \mu_2, {\sigma_1}^2 + {\sigma_2}^2)$$

となる。

(2) 一般に，変数 x, y に対して，

$ax + by$（a, b は定数）

の形の式を，x, y の 1 次結合という。従って，定理の(2)は，X, Y の 1 次結合は，常に正規分布に従うことを意味している。

(3) 定理(2)を確認すると，$X \sim N(\mu_1, {\sigma_1}^2)$，$Y \sim N(\mu_2, {\sigma_2}^2)$ であるから，

X の平均は $E(X) = \mu_1$，　X の分散は $V(X) = {\sigma_1}^2$
Y の平均は $E(Y) = \mu_2$，　Y の分散は $V(Y) = {\sigma_2}^2$

このとき，1 次結合 $aX + bY$ の平均は

$$E(aX + bY) = aE(X) + bE(Y) = a\mu_1 + b\mu_2$$

また，X, Y が独立であるから，aX, bY も独立であり，

$$V(aX + bY) = V(aX) + V(bY) = a^2V(X) + b^2V(Y)$$
$$= a^2{\sigma_1}^2 + b^2{\sigma_2}^2$$

さらに，1 次変換 aX や bY も正規分布をなすので，再生性より，

$$aX + bY$$

も正規分布をなす。従って，

$$aX + bY \sim N(E(aX + bY), V(aX + bY))$$

よって，

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2{\sigma_1}^2 + b^2{\sigma_2}^2)$$

なお，$a = 1$，$b = -1$ とおけば，

$$X - Y \sim N(\mu_1 - \mu_2, {\sigma_1}^2 + {\sigma_2}^2)$$

となることに注意する。

■ 例 1

X, Y が独立で，$X \sim N(5, 2)$，$Y \sim N(7, 3)$ のとき，正規分布の再生性より，確率変数 $2X - Y$ は正規分布に従い，

$$2X - Y \sim N(\bigcirc, \triangle)$$

となる。○と△の値を求めよ。

＜解＞

$$\bigcirc = E(2X - Y) = 2E(X) - E(Y) = 2 \times 5 - 7 = 3$$

$$\triangle = V(2X - Y) = V(2X) + V(-Y) = 4V(X) + V(Y) = 4 \times 2 + 3 = 11$$

■ 例２

ある大学の男子学生の体重は平均 60kg，標準偏差 8kg の正規分布に従い，女子学生の体重は平均 52 kg，標準偏差 5kg の正規分布に従うことが知られている。いま，1 人の男子学生と 1 人の女子学生を無作為に選ぶとき，2 人の体重の合計が 100kg を超える確率を求めよ。

＜解＞

男子学生の体重を X，女子学生の体重を Y とすると，

$$X \sim N(60, 8^2), \quad Y \sim N(52, 5^2)$$

また，男子学生から 1 人を選ぶという試行と，女子学生から 1 人を選ぶという試行は，互いに独立であるから，X, Y は独立である。

$S = X + Y$ とおくと，

S の平均　$E(S) = E(X) + E(Y) = 60 + 52 = 112$

S の分散　$V(S) = V(X) + V(Y) = 8^2 + 5^2 = 89$

S の標準偏差　$\sigma(S) = \sqrt{V(S)} = \sqrt{89}$

従って，正規分布の再生性より，

$$S \sim N(112, 89)$$

よって，S を標準化して，$Z = \dfrac{S - 112}{\sqrt{89}}$ とおくと，$Z \sim N(0, 1)$ である。

ここで，$P(S > 100)$ を求めればよい。

$$P(S > 100) = P\left(Z > \frac{100 - 112}{\sqrt{89}}\right) = P(Z > -1.27)$$

$$= P(Z < 1.27) = 0.898 \text{（答）}$$

確率は約 0.9 であるから，大学生全体から，「男子を 1 人，女子を 1 人選ぶ」という試行を 10 回行えば，体重の合計が 100kg を超えるのは 9 回程度である。従って，1 回の試行で，ほぼ確実に超えることがわかる。

3. 正規分布の再生性（n 変数の場合）

● 定理

確率変数 $X_1, X_2, \cdots, X_n$ が独立で，各 X_i が正規分布 $N(\mu_i, \sigma_i^2)$ に従うとき，

確率変数 $\sum_{i=1}^{n} a_i X_i$ は，正規分布 $N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$ に従う。（a_i は任意の定数）

(1) 一般に，n 個の変数 $x_1, x_2, \cdots, x_n$ に対して，

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n \quad （各 a_i は定数）$$

の形の式を，$x_1, x_2, \cdots, x_n$ の 1 次結合という。

(2) 定理を，Σ を使わずに書くと，次のようになる。まず，次の 2 つの条件が仮定である。

① $X_1, X_2, \cdots, X_n$ は独立

② $X_i \sim N(\mu_i, \sigma_i^2)$ （$i = 1, 2, \cdots, n$ ）

この 2 つが成立するとき，各 a_i がどんな定数であっても，$X_1, X_2, \cdots, X_n$ の 1 次結合

$$a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

は，正規分布に従うことを定理は主張している。

ここで，プリント p.58 の定理より，

$$\begin{aligned} E(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) &= a_1 E(X_1) + a_2 E(X_2) + \cdots + a_n E(X_n) \\ &= a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n \end{aligned}$$

$$\begin{aligned} V(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) &= a_1^2 V(X_1) + a_2^2 V(X_2) + \cdots + a_n^2 V(X_n) \\ &= a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2 \end{aligned}$$

従って，$a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$ は，正規分布

$$N(a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2)$$

に従うことになる。