正規分布の再生性について

浜田　宏

正規分布にしたがう確率変数同士の和や差によって新しい確率変数をつくると，その新しい確率変数がやはり正規分布に従うことはよく知られています．これを分布の再生性といいます．一般に，すべての分布が再生性を持つわけではありませんが，2 項分布やポアソン分布などは再生性を持っています．

この再生性は自明な性質ではないので，以下に初等的な証明を与えます[1]．

証明には変数変換定理を使うので，まず内容を確認します．

命題 1 (変数変換定理)**.** 確率変数の組 (X, Y) と (U, V) に

$$\begin{cases} X &= g(U, V) \\ Y &= h(U, V) \end{cases}$$

という関係があるとき, (x, y) 上の確率分布 $p(x, y)$ は，(u, v) 上の確率分布 $q(u, v)$

$$\begin{aligned} q(u, v) &= p(x, y) \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \\ &= p(g(u, v), h(u, v)) \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \end{aligned}$$

に変換される (小針 1973).

命題 2 (相関のある正規分布の和と差の分布)**.** 確率変数の組 (X, Y) が 2 次元正規分布

$$\text{Normal}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$$

に従うと仮定する． $Z = X - Y$ および $W = X + Y$ と定義すると，それぞれの分布は

[1] このノートは浜田宏．2020『その問題，やっぱり数理モデルが解決します』第 11 章の補足資料です．ver.1.0.; 2021 年 10 月 24 日公開. ver.1.0.1.; 誤字修正.

$$
\begin{aligned}
Z \sim & \mathrm{Normal}(\mu_1-\mu_2, \sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2) \\
W \sim & \mathrm{Normal}(\mu_1+\mu_2, \sigma_1^2+2\rho\sigma_1\sigma_2+\sigma_2^2)
\end{aligned}
$$

である．ここで ρ は X, Y の相関係数である．

証明. 相関のある確率変数 X, Y の合成を次のように考える．

$$
\begin{cases}
x & = z+u \\
y & = u
\end{cases}
$$

すると $X-Y=Z$ である．確率変数 Z の密度は変数変換定理より

$$
\begin{aligned}
h(z,u) &= f(x,y)\begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial u} \end{vmatrix} \\
&= f(x,y)\begin{vmatrix} 1 & 1 \\ 0 & 0 \end{vmatrix} = f(x,y)\cdot 1 = f(z+u,u)
\end{aligned}
$$

だから，合成積の定理 を使い

$$
g(z) = \int_{-\infty}^{\infty} h(z,u)du = \int_{-\infty}^{\infty} f(z+u,u)du
$$

と表すことができる．被積分関数の $f(z+u,u)$ を明示的に書くと，2 次元正規分布の確率密度関数の定義から

$$
\begin{aligned}
f(z+u,u) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{\alpha} \\
\alpha &= -\frac{1}{2(1-\rho^2)}\left(\frac{(z+u-\mu_1)^2}{\sigma_1^2}+\frac{(u-\mu_2)^2}{\sigma_2^2}+\frac{2\rho(z+u-\mu_1)(u-\mu_2)}{\sigma_1\sigma_2}\right)
\end{aligned}
$$

である．e の指数部の一部を

$$
\frac{(z+u-\mu_1)^2}{\sigma_1^2}+\frac{(u-\mu_2)^2}{\sigma_2^2}+\frac{2\rho(z+u-\mu_1)(u-\mu_2)}{\sigma_1\sigma_2} = au^2+bu+c
$$

とおいて u に関して平方完成する． a, b, c は明示的には，それぞれ

$$
\begin{aligned}
a &= \frac{\sigma_1^2-2\rho\sigma_1\sigma_2+\sigma_2^2}{\sigma_1^2\sigma_2^2} \\
b &= \frac{2\{(\mu_1+\mu_2-z)\rho\sigma_1\sigma_2+(z-\mu_1)\sigma_2^2-\mu_2\sigma_1^2\}}{\sigma_1^2\sigma_2^2} \\
c &= \frac{\mu_2\sigma_1^2+2(z-\mu_1)\rho\mu_2\sigma_1\sigma_2+(z-\mu_1)^2\sigma_2^2}{\sigma_1^2\sigma_2^2}
\end{aligned}
$$

である．定数部分を $K = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}$ とおけば，

$$\begin{aligned}
g(z) &= K\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(a\left(u+\frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c\right)\right\} du \\
&= Ke^D \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)a^{-1}}\left(u+\frac{b}{2a}\right)^2\right\} du \\
&= Ke^D \int_{-\infty}^{\infty} \exp\left\{-\frac{(u+A)^2}{2C}\right\} du \\
&= Ke^D\sqrt{C}\sqrt{2\pi}
\end{aligned}$$

ただし，

$$A = \frac{b}{2a}, C = (1-\rho^2)a^{-1}, D = -\frac{1}{2(1-\rho^2)}\left(-\frac{b^2}{4a}+c\right)$$

である．C, D, K を代入して

$$\begin{aligned}
g(z) &= Ke^D\sqrt{C}\sqrt{2\pi} \\
&= \frac{1}{\sqrt{2\pi}(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2)} \exp\left\{-\frac{(z-(\mu_1-\mu_2))^2}{2(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2)}\right\}
\end{aligned}$$

を得る．$g(z)$ は確率変数 Z の確率密度関数であり，確かに正規分布の確率密度関数である．よって

$$Z \sim \mathrm{Normal}(\mu_1 - \mu_2, \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2)$$

が示せた．$W = X + Y$ と仮定した場合の分布も同様の計算により特定できる．

□

命題 3 (独立な正規分布の和と差の分布)**.** 確率変数 X, Y が独立に正規分布

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

に従うと仮定する．$Z = X - Y$ および $W = X + Y$ で定義すると，それぞれの分布は

$$Z \sim \mathrm{Normal}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$
$$W \sim \mathrm{Normal}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

である．

証明. 命題 1 で $\rho = 1$ とおけばよい． □

以上が正規分布同士を合成する場合に用いる初等的な計算です．

1　R による計算例

```
#　確率変数を足すことの意味を確認するコード
m1 <- 0
m2 <- 3
curve(dnorm(x,m1,1),-5,7)
par(new=TRUE)
curve(dnorm(x,m2,1),-5,7,ylab="")
```

平均が 0 の正規分布 X_1 と平均が 3 の正規分布 X_2 の確率密度関数をプロットします.

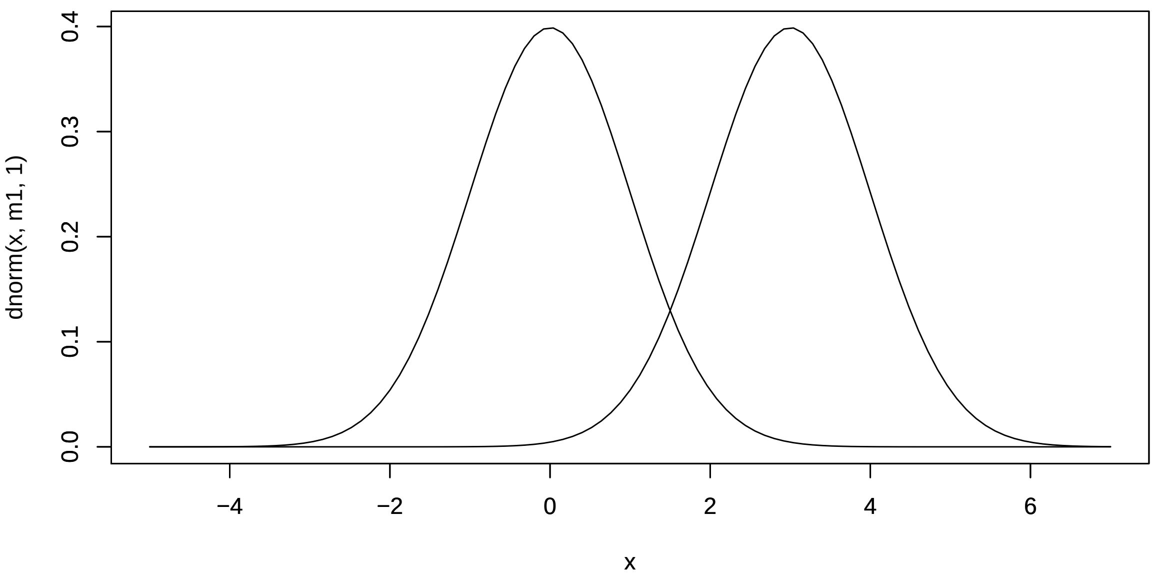

図 1: 2 つの正規分布

$Y = X_1 + X_2$ の分布を作ります.

```
# Y=X1+X2 の分布とは？ --------------------------
x1 <- rnorm(1000,m1,1)
x2 <- rnorm(1000,m2,1)
y <- x1+x2 # 発生させた正規乱数同士を足す.結果は1000個のデータ

hist(y,breaks=30)
```

合成した結果の Y のヒストグラムをプロットします.

図 2: X_1 と X_2 を足した分布

結果は正規分布です.

合成したあとの分布 Y の平均と分散を確認してみましょう.

```
> mean(y)
[1] 3.01811
> var(y)
[1] 1.979479
```

平均と分散はたしかに証明で示した結果と一致しています.

次に X_1 と X_2 の合併を考えます.

```
# x1,x2 の合併は二山の分布になる
z <- c(x1,x2)
hist(z,breaks=30)
```

2 つを合併したデータのヒストグラムをプロットします.

図 3: X_1 と X_2 を合併したときの分布

結果は二峰型の分布です.

最後に $W = X_1 - X_2$ の分布を作ります.

```
# W=X1-X2 の分布とは？ --------------------------
w <- x1-x2 # 発生させた正規乱数同士の差をとる.結果は 1000個のデータ
hist(w,breaks=30)
```

差で合成した W のヒストグラムをプロットします.

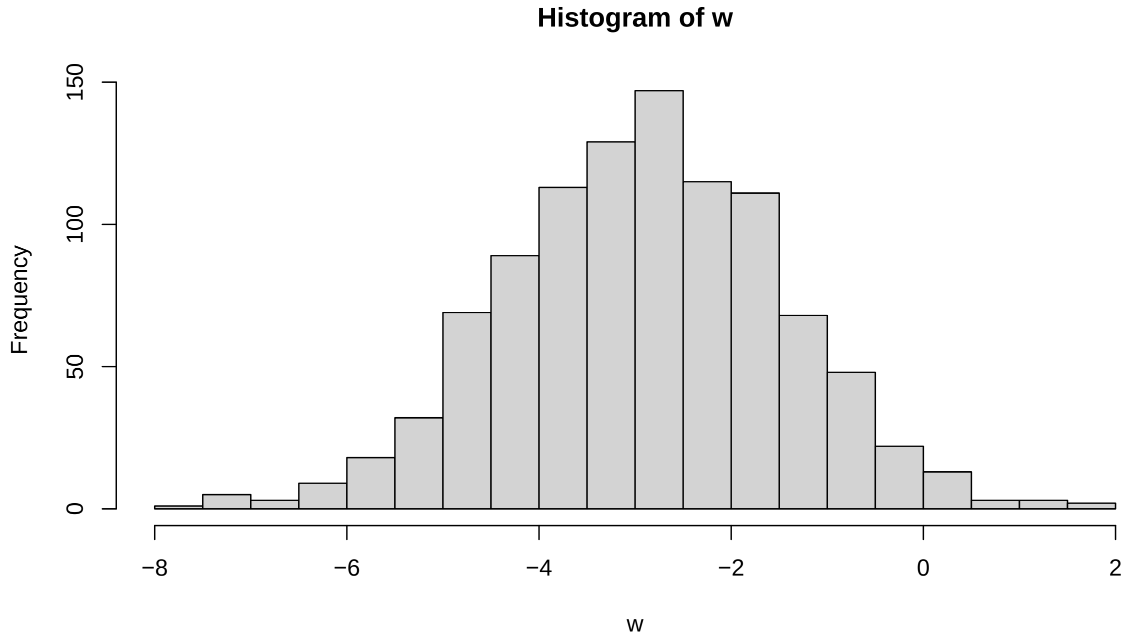

図 4: X_1 と X_2 の差をとった $W = X_1 - X_2$ の分布．平均の位置に注目．

```
> mean(w)
[1] -3.004647
> var(w)
[1] 1.979403
```

差によって合成場合の平均と分散は，証明の結果と確かに一致しています．分散は X_1 と X_2 の分散の和になっている点に注意します．

文献一覧

小針晛宏，1973，『確率・統計入門』岩波書店．