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．

この再生性はけして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．

命題 1（**相関のある正規分布の和と差の分布**）．確率変数 X,Y が 2 次元正規分布

$$N\left((\mu_x,\mu_y),\begin{pmatrix}\sigma_x^2 & \sigma_{xy}\\ \sigma_{xy} & \sigma_y^2\end{pmatrix}\right)$$

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

$$W\sim N(\mu_x-\mu_y,\ \sigma_x^2-2\rho\sigma_x\sigma_y+\sigma_y^2),\ W\sim N(\mu_x+\mu_y,\ \sigma_x^2+2\rho\sigma_x\sigma_y+\sigma_y^2)$$

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．

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

すると $x-y=w$ である．確率変数 $W=X-Y$ の密度は変数変換定理より

$$\varphi(w,z)=f(x,y)\begin{vmatrix}\dfrac{\partial x}{\partial w} & \dfrac{\partial x}{\partial z}\\ \dfrac{\partial y}{\partial w} & \dfrac{\partial y}{\partial z}\end{vmatrix}=f(x,y)\begin{vmatrix}1 & 1\\ 0 & 1\end{vmatrix}=f(x,y)\cdot 1=f(w+z,z)$$

だから，合成積の定理[1]を使い

$$g(w)=\int_{-\infty}^{\infty}\varphi(w,z)\ dz=\int_{-\infty}^{\infty}f(w+z,z)\ dz$$

と表すことができる．明示的に書くと (X,Y) の同時確率密度関数の定義から

[1] 合成積の定理と変数変換定理については小針『確率・統計入門』（1973）と高木『定本解析概論』（2010）を参照

$$g(w)=\int_{-\infty}^{\infty}\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(w+z-\mu_x)^2}{\sigma_x^2}+\frac{(z-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(w+z-\mu_x)(z-\mu_y)}{\sigma_x\sigma_y}\right)\right\}dz$$

である．指数部を

$$\frac{(w+z-\mu_x)^2}{\sigma_x^2}+\frac{(z-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(w+z-\mu_x)(z-\mu_y)}{\sigma_x\sigma_y}=az^2+bz+c$$

とおいて z に関して平方完成する． a,b,c は明示的には，それぞれ

$$a=\frac{\sigma_x^2-2\rho\sigma_x\sigma_y+\sigma_y^2}{\sigma_x^2\sigma_y^2}\ ,b=\frac{2\{(\mu_x+\mu_y-w)\rho\sigma_x\sigma_y+(w-\mu_x)\sigma_y^2-\mu_y\sigma_x^2\}}{\sigma_x^2\sigma_y^2},$$
$$c=\frac{\mu_y^2\sigma_x^2+2(w-\mu_x)\mu_y\rho\sigma_x\sigma_y+(w-\mu_x)^2\sigma_y^2}{\sigma_x^2\sigma_y^2}$$

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

$$g(w)=K\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(a\left(z+\frac{b}{2a}\right)^2-\frac{b^2}{4a}+c\right)\right\}dz=K\exp\{D\}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-\rho^2)a^{-1}}\left(z+\frac{b}{2a}\right)^2\right\}dz$$
$$=K\exp\{D\}\int_{-\infty}^{\infty}\exp\left\{-\frac{(z+A)^2}{2C}\right\}dz=K\exp\{D\}\sqrt{C}\sqrt{2\pi}$$

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

$$g(w)=K\exp\{D\}\sqrt{C}\sqrt{2\pi}=\frac{1}{\sqrt{2\pi}(\sigma_x^2-2\rho\sigma_x\sigma_y+\sigma_y^2)}\exp\left\{-\frac{(w-(\mu_x-\mu_y))^2}{2(\sigma_x^2-2\rho\sigma_x\sigma_y+\sigma_y^2)}\right\}$$

を得る． $g(w)$ は確率変数 $W=X-Y$ の確率密度関数である．よって

$$W\sim N(\mu_x-\mu_y,\ \sigma_x^2-2\rho\sigma_x\sigma_y+\sigma_y^2)$$

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 2（**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

$$X \sim N(\mu_x, \sigma_x^2), \quad Y \sim N(\mu_y, \sigma_y^2)$$

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

$$W \sim N(\mu_x - \mu_y,\ \sigma_x^2 + \sigma_y^2),\ W \sim N(\mu_x + \mu_y,\ \sigma_x^2 + \sigma_y^2)$$

である．

証明．命題 1 で $\rho = 0$ とおけばよい．

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=””)
```

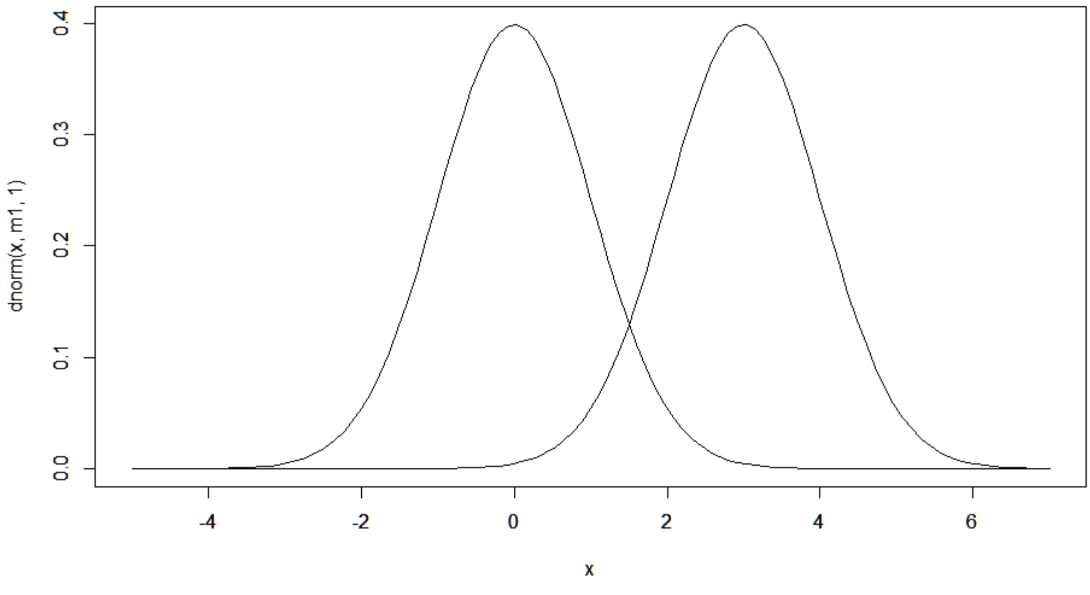

図:2 つの正規分布

```
# 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)
```

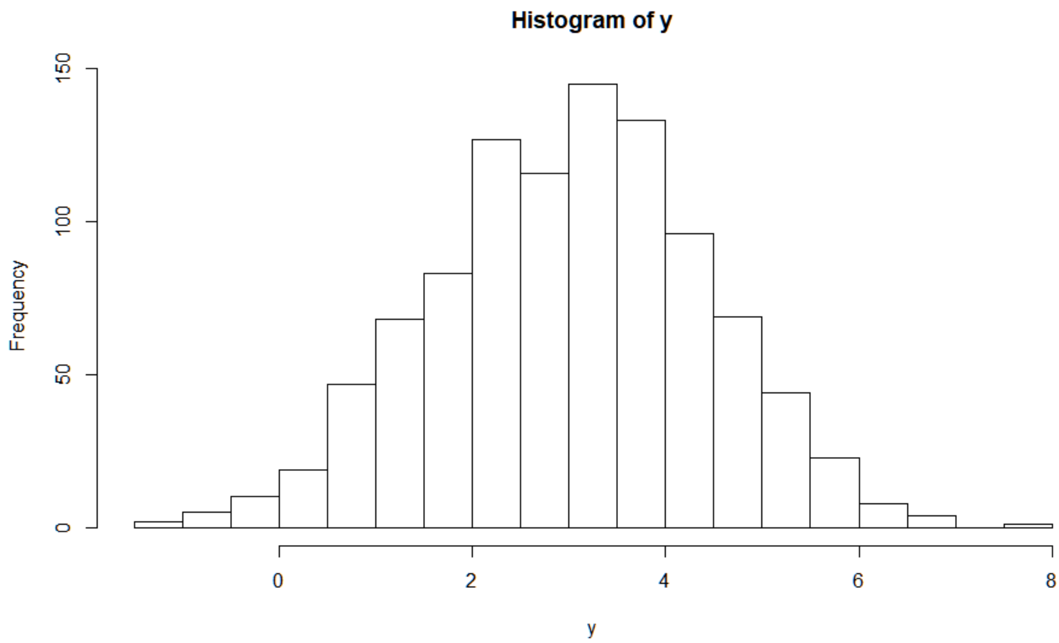

図:2 つを足した結果

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

平均と分散は証明の通り.

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

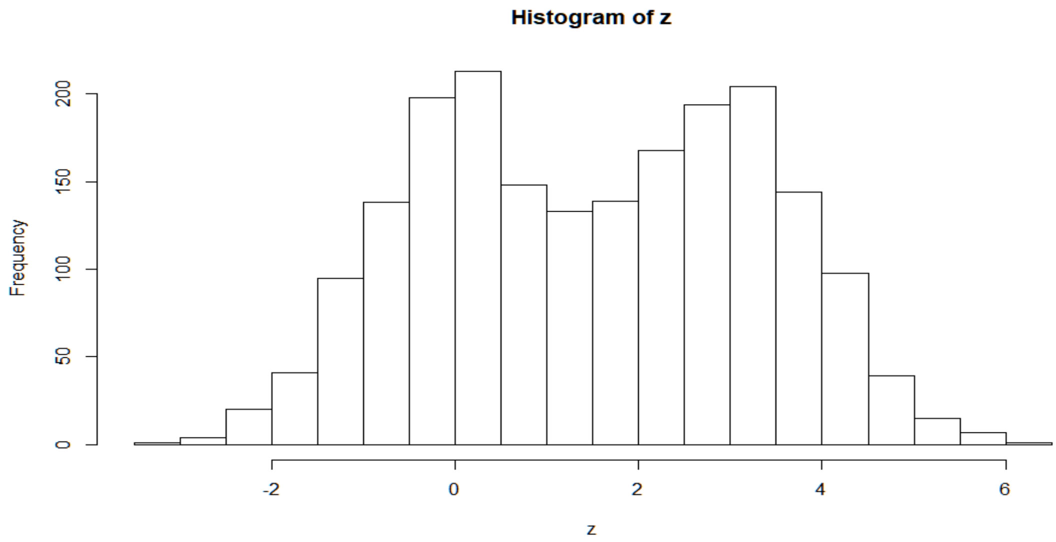

図:2 つを合併したときの分布

```
# Y=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)
```

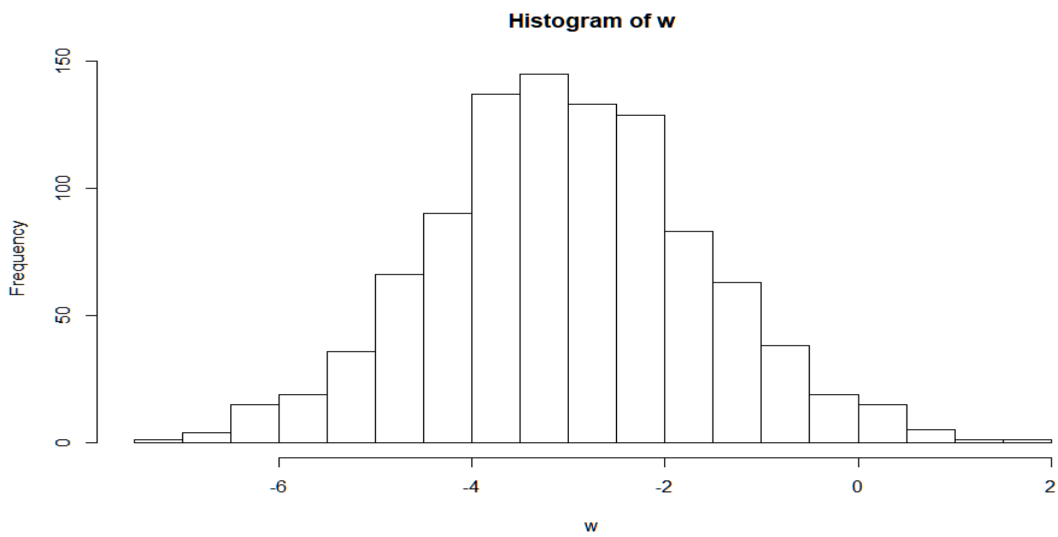

図:2 つの差をとった分布. 平均の位置に注目.

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

証明の結果と確かに一致する. 分散は和になっている点に注意.