Publication: lifemater.com, Author: peekAboo, Creation: 2021/12/30

平均・分散・共分散（期待値で記述）

1. 基本的な定義と関係式

n 個のデータ $\{(x_i, y_i)\}$ について、i 番目のデータが得られる確率 p_i を、

$$p_i = \frac{1}{n}, \tag{1-1}$$

とすれば、$\{(x_i, y_i)\}$ の平均値、分散、共分散を、確率変数に関する表現を用いて表すことができる。

ここで、以下のように 2 つの式を定義する。

$$E[x] = \sum_i x_i p_i\,, \tag{1-2}$$

$$\mathrm{Cov}[x, y] = E\Big[\big(x - E[x]\big)\big(y - E[y]\big)\Big], \tag{1-3}$$

ここで、a, b を定数とすると、

$$E[ax+b] = \sum_i (ax_i + b)\,p_i = \sum_i (ax_i p_i + bp_i) = a\sum_i x_i p_i + b\sum_i p_i\,,$$

となる。p_i は確率なので、

$$\sum_i p_i = 1, \tag{1-4}$$

が成立する。よって、Eqs. (1-2, 1-4)より、

$$E[ax+b] = a\sum_i x_i p_i + b\sum_i p_i = aE[x] + b,$$

となる。つまり、

$$E[ax+b]=aE[x]+b\,, \tag{1-5}$$

が成立する。

2. 平均

Eqs. (1-1, 1-2)より、

$$E[x]=\sum_i x_i p_i=\frac{1}{n}\sum_i x_i\,,$$

であり、$\{x_i\}$の平均値μ_{x}は、以下の式で与えられる。

$$\mu_{\mathrm{x}}=E[x]. \tag{2-1}$$

ここで、a, bを定数として、

$$X_i=a(x_i-b), \tag{2-2}$$

とすると、Eq. (1-5)より、

$$E[X]=E[a(x-b)]=E[ax-ab]=aE[x]-ab=a(E[x]-b),$$

となる。つまり、

$$E[X]=a(E[x]-b), \tag{2-3}$$

となる。

3. 分散

Eqs. (1-1 ~ 1-3, 2-1)より、

$$\mathrm{Cov}[x,x]=E\left[\left(x-E[x]\right)\left(x-E[x]\right)\right]=E\left[\left(x-E[x]\right)^2\right]=\frac{1}{n}\sum_i\left(x_i-\mu_{\mathrm{x}}\right)^2,$$

となるので、$\{x_i\}$ の分散 s_{xx}^2 は以下の式で与えられる。

$$s_{\mathrm{xx}}^2=\mathrm{Cov}[x,x]. \tag{3-1}$$

この式を展開すると、Eq. (1-5)より、

$$\begin{aligned}\mathrm{Cov}[x,x]&=E\left[\left(x-E[x]\right)^2\right]=E\left[x^2-2xE[x]+\left(E[x]\right)^2\right]\\&=E\left[x^2\right]+E\left[-2xE[x]\right]+E\left[\left(E[x]\right)^2\right]=E\left[x^2\right]-2E[x]E[x]+\left(E[x]\right)^2E[1],\\&=E\left[x^2\right]-\left(E[x]\right)^2\end{aligned}$$

となるので、

$$\mathrm{Cov}[x,x]=E\left[x^2\right]-\left(E[x]\right)^2, \tag{3-2}$$

が成立する。

ここで、*a, b* を定数として、Eq. (2-2)のように X_i を定義すると、Eqs. (1-5, 2-3, 3-2)より、

$$\begin{aligned}\mathrm{Cov}[X,X]&=E\left[X^2\right]-\left(E[X]\right)^2=E\left[\left(a(x-b)\right)^2\right]-\left(E\left[a(x-b)\right]\right)^2\\&=E\left[a^2\left(x^2-2bx+b^2\right)\right]-\left(a\left(E[x]-b\right)\right)^2=a^2E\left[\left(x^2-2bx+b^2\right)\right]-a^2\left(E[x]-b\right)^2\\&=a^2\left\{E\left[x^2\right]+E[-2bx]+E\left[b^2\right]-\left(\left(E[x]\right)^2-2bE[x]+b^2\right)\right\}\\&=a^2\left\{E\left[x^2\right]-2bE[x]+b^2-\left(\left(E[x]\right)^2-2bE[x]+b^2\right)\right\}\\&=a^2\left\{E\left[x^2\right]-\left(E[x]\right)^2\right\}=a^2\mathrm{Cov}[x,x]\end{aligned},$$

となる。つまり、

$$\mathrm{Cov}[X,X]=a^2\mathrm{Cov}[x,x], \tag{3-3}$$

となる。Eq. (3-3)は b によらないため、分散は b に依存しないことが分かる。

4. 共分散

Eqs. (1-1 ~ 1-3, 2-1)より、

$$\mathrm{Cov}[x,y]=E\left[\left(x-E[x]\right)\left(y-E[y]\right)\right]=\frac{1}{n}\sum_i\left(x-E[x]\right)\left(y-E[y]\right)=\frac{1}{n}\sum_i\left(x-\mu_{\mathrm{x}}\right)\left(y-\mu_{\mathrm{y}}\right),$$

となるので、$\{(x_i,y_i)\}$ の共分散 s_{xy}^2 は以下の式で与えられる。

$$s_{\mathrm{xy}}^2=\mathrm{Cov}(x,y). \tag{4-1}$$

この式を展開すると、Eq. (1-5)より、

$$\begin{aligned}&\mathrm{Cov}[x,y]=E\left[\left(x-E[x]\right)\left(y-E[y]\right)\right]=E\left[xy-yE[x]-xE[y]+E[x]E[y]\right]\\&=E[xy]+E\left[-yE[x]\right]+E\left[-xE[y]\right]+E\left[E[x]E[y]\right]\\&=E[xy]-E[x]E[y]-E[y]E[x]+E[x]E[y]E[1]\\&=E[xy]-2E[x]E[y]+E[x]E[y]=E[xy]-E[x]E[y]\end{aligned},$$

となるので、

$$\mathrm{Cov}[x,y]=E[xy]-E[x]E[y], \tag{4-2}$$

が成立する。

ここで、a, b, c, d を定数として、

$$X_i=a\left(x_i-b\right), \tag{4-3}$$

$$Y_i=c\left(y_i-d\right), \tag{4-4}$$

とすると、Eqs. (1-5, 2-3, 4-2 ~ 4-4)より、

$$\begin{aligned}&\mathrm{Cov}[X,Y]=E[XY]-E[X]E[Y]=E\left[a(x-b)c(y-d)\right]-E\left[a(x-b)\right]E\left[c(y-d)\right]\\&=acE[xy-by-dx+bd]-a\left(E[x]-b\right)c\left(E[y]-d\right)\\&=ac\left\{E[xy]-bE[y]-dE[x]+E[bd]-\left(E[x]E[y]-bE[y]-dE[x]+bd\right)\right\}\\&=ac\left\{E[xy]+bd-\left(E[x]E[y]+bd\right)\right\}=ac\left\{E[xy]-E[x]E[y]\right\}=ac\mathrm{Cov}[x,y]\end{aligned},$$

となる。つまり、

$$\mathrm{Cov}[X, Y] = ac\mathrm{Cov}[x, y], \tag{4-5}$$

となる。Eq. (4-5)は *b, d* によらないため、共分散は *b, d* に依存しない。

5. 相関係数

Eqs. (3-1, 4-1)より、$\{(x_i, y_i)\}$の相関係数ρ_{xy}は以下の式で与えられる。

$$\rho_{\mathrm{xy}} = \frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{Cov}(x, x)}\sqrt{\mathrm{Cov}(y, y)}}. \tag{5-1}$$

Eqs. (4-3, 4-4)と同様にX_i, Y_iを定義すると、Eqs. (3-3, 4-5, 5-1)より、

$$\begin{aligned}\rho_{\mathrm{XY}} &= \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Cov}(X, X)}\sqrt{\mathrm{Cov}(Y, Y)}} = \frac{ac\mathrm{Cov}(x, y)}{\sqrt{a^2\mathrm{Cov}(x, x)}\sqrt{c^2\mathrm{Cov}(y, y)}} \\ &= \frac{ac}{\sqrt{a^2c^2}}\frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{Cov}(x, x)}\sqrt{\mathrm{Cov}(y, y)}} = \frac{ac}{|ac|}\rho_{\mathrm{xy}}\end{aligned},$$

となる。つまり、以下の関係式が成立する。

$$\rho_{\mathrm{XY}} = \begin{cases}\rho_{\mathrm{xy}} & (ac > 0) \\ -\rho_{\mathrm{xy}} & (ac < 0)\end{cases}. \tag{5-2}$$