Introduction to mathematical statistics ゼミ (第 2 回)

担当: 伊藤真道

7/24

2 Chap.2. Conditional Probability and Stochastic Independence

2.3 The correlation coefficient / 相関係数

確率変数 X, Y, Z が，密度関数 $f(x, y, z)$ を持つとする．もし，$u(x, y, z)$ が x, y, z の関数であるならば，$E[u(X, Y, Z)]$ が存在する (と仮定する．存在しないこともあるが，本節では，E[u(X,Y,Z)] の存在を仮定する．)．すると，X, Y, Z の平均 μ_1, μ_2, μ_3，分散 $\sigma_1^2, \sigma_2^2, \sigma_3^2$ はそれぞれ (煩雑なので，X の場合のみ記す)，

$$\begin{aligned}
\mu_1 &= E[X] = \int_{\mathscr{A}} x f(x, y, z) dydzdx \\
\sigma_1^2 &= V[X] = E[(X - \mu_1)^2] \\
&= \int_{\mathscr{A}} (x - \mu_1)^2 f(x, y, z) dydzdx
\end{aligned}$$

となる．ただし，ここで $\mathscr{A}$ は X, Y, Z の空間を表す．二つの確率変数の関係を表す指標として**共分散** (Covariance) がある．X と Y の共分散は以下の期待値で定義される．

$$\begin{aligned}
Cov(X, Y) &= E[(X - \mu 1)(Y - \mu_2)] \\
&= E[XY - \mu_2 X - \mu_1 Y + \mu_1 \mu_2] \\
&= E[XY] - \mu_2 E[X] - \mu_1 E[Y] + \mu_1 \mu_2 \\
&= E[XY] - \mu_2 \mu_1 - \mu_1 \mu_2 + \mu_1 \mu_2 \\
&= E[EY] - \mu_1 \mu_2
\end{aligned}$$

同様にして，X と Z の共分散は $E[(X - \mu_1)(Z - \mu_3)]$，$Y$ と Z の共分散は $E[(Y - \mu_2)(Z - \mu_3)]$ で定義される．もし，$\sigma_1, \sigma_2 > 0$ の時，X と Y の相関係数 ρ_{12} は

$$\rho_{12} = \frac{Cov[(X, Y)}{\sigma_1 \sigma_2}$$

で定義される．二つの確率変数の積の期待値は，それぞれの期待値の積と共分散の和である，つまり，

$$E[XY] = \mu_1 \mu_2 + Cov(X, Y)$$

であるということは重要な事実である．

¡note¿

次に，条件付き分布について考える．$f(x, y)$ を二つの確率変数 X, Y の同時確率密度関数 (joint probability density function, jpdf)，$f_1(x)$ を X の周辺分布とする．この時，$X = x$ を与えたもとでの Y の条件付き分布は，$f_1(x) > 0$ なる点に対して，

$$f(y|x) = \frac{f(x, y)}{f_1(x)}$$

である．また，$X = x$ のもとでの Y の条件付き期待値は

$$E[Y|x] = \int_{-\infty}^{\infty} y f(x, y) dy = \frac{\int_{-\infty}^{\infty} dy}{f_1(x)}$$

で与えられる (連続型確率変数の場合)．この $X = x$ のもとでの Y の条件付き期待値は，$\phi(x)$(x の関数) とみなすことができる．$Y = y$ のもとでの X の条件付き分布も同様に定義され．ここでは，$\psi(y)$ で表すこととする．

$\phi(x)$ が $\phi(x) = a + bx$ のような線形関数の場合，Y の条件付き期待値は，”x に関して線形である”，もしくは，”Y は線形の条件付き期待値を持つ”という．

$\phi(x) = a + bx$ が簡単な定数である場合を考える．以下，σ_1^2, σ_2^2 のいずれも 0 でないと仮定する．

$$E[Y|x] = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f_1(x)} = a + bx$$

から，

$$\int_{-\infty}^{\infty} y f(x, y) dy = (a + bx) f_1(x) \tag{1}$$

を得る．(1) 式の両辺を x について積分すると，

$$
\begin{aligned}
&\int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x,y)dxdy = \int_{-\infty}^{\infty} (a+bx) f_1(x)dx \\
\Leftrightarrow &\int_{-\infty}^{\infty} y f_2(y)dy = a \int_{-\infty}^{\infty} f_1(x)dx + b \int_{-\infty}^{\infty} x f_1(x)dx \\
\Leftrightarrow & E[Y] = a + bE[X] \\
(\Leftrightarrow & \mu_2 = a + b\mu_1)
\end{aligned}
$$

となる．(1) 式が，x をかけた後に積分されたとすると，

$$
\begin{aligned}
\int_{-\infty}^{\infty} xyf(x,y)dxdy = & \int_{-\infty}^{\infty} x(a+bx)f_1(x)dx \\
& \Leftrightarrow E[XY] = aE[X] + bE[X^2] \\
& (\Leftrightarrow \rho\sigma_1\sigma_2 = a\mu_1 + b(\sigma_1^2 + \mu_1^2))
\end{aligned}
$$

となる．これらの結果から，(a,b) について連立して解くと，

$$a = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1$$

,

$$b = \rho\frac{\sigma_2}{\sigma_1}$$

が得られ，以上から，Y の条件付き期待値が x と線形の関係を持つとき，$X = x$ の時の Y の条件付き期待値は，

$$\phi(x) = E[Y|x] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

となる．逆の場合は，

$$\psi(y) = E[X|y] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$$

となる．

¡note¿

次に，条件付き期待値が線形であるという仮定のもとでの，条件付き分布の分散について考えていく．Y の条件付き分散は，

$$
\begin{aligned}
E[(Y-E[Y|x])^2|x] &= \int_{-\infty}^{\infty}[y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)]^2 f(y|x)dy \\
&= \frac{1}{f_1(x)}\int_{-\infty}^{\infty}[(y-\mu_2)-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)]^2 f(x,y)dy
\end{aligned}
$$

で与えられる (連続型確率変数の場合)．この分散は，非負で，x の関数である．そのため，もし，$f_1(x)$ をかけて，x について積分されたなら，得られる結果は非負になるであろう．この結果は，

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[(y-\mu_2)-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)]^2 f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[(y-\mu_2)^2-2\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)(y-\mu_2)+\rho^2\frac{\sigma_2^2}{\sigma_1^2}(x-\mu_1)^2]f(x,y)dxdy \\
&= E[(y-\mu_2)^2]-2\rho\frac{\sigma_2}{\sigma_1}E[(x-\mu_1)(y-\mu_2)]+\rho^2\frac{\sigma_2^2}{\sigma_1^2}E[(x-\mu_1)^2] \\
&= \sigma_2^2-2\rho\frac{\sigma_2}{\sigma_1}\rho\sigma_1\sigma_2+\rho^2\frac{\sigma_2^2}{\sigma_1^2}\sigma_1^2 \\
&= \sigma_2^2+2\rho^2\sigma_2^2+\rho^2\sigma_2^2 \\
&= \sigma_2^2(1-\rho^2) \geq 0
\end{aligned}
$$

のように確かめることができる．つまり，もし，分散が $k(x)=E[(Y-E[Y|x])^2|x]$ のように表される時，$E[k(X)]=\sigma_2^2(1-\rho^2)\geq 0$ である．結果として，$\rho^2\leq 1$ もしくは，$-1\geq\rho\geq 1$ となる．条件付き期待値が，線形であるかどうかに関わらず，$-1\geq\rho\geq 1$ となることは，練習問題とする．

$E[(Y-E[Y|x])^2|x]$ が，正であり，x **の関数でない**とする．つまり，$E[(Y-E[Y|x])^2|x]$ が正定数 $k>0$ とする．今，もし k が $f_1(x)$ をかけてから，x で積分されたとすると，その結果は，$k=\sigma_2^2(1-\rho^2)$ である．そのため，この場合， $X=x$ での Y の条件付き分散は，$\sigma_2^2(1-\rho^2)$ である．もし，$\rho=0$ ならば，それぞれの Y の条件付き分布の分散は，Y の周辺分布の分散 σ_2^2 と一致する．他方，もし ρ が 1 に近いならば，$X=x$ の時の Y の条件付き分布の分散は，比較的小さくなり，分布の形状は平均 $E[Y|x]=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$ の付近に集中した形になる．これらの議論は，離散型の確率分布に対しても成り立つ．

¡note¿

$f(x,y)$ を二つの確率変数の同時確率密度関数とする．もし，$-h_1 < t_1 < h_1, -h_2 < t_2 < h_2$ の範囲に $E[e^{t_1X+t_2Y}]$ が存在するならば，これは，同時密度関数の積率母関数と呼ばれ，$M(t_1,t_2)$ で表される．1 変数の場合と同様に，$M(t_1,t_2)$ は X,Y の同時密度関数の形状を完全に決定し，それゆえに，X,Y の周辺分布も決定する．実際，

$$M(t_1,0) = E[e^{t_1X}] = M(t_1)$$

,

$$M(0,t_2) = E[e^{t_2Y}] = M(t_2)$$

となることが容易にわかる．さらに，高次モーメントも 1 変数の場合と同様に，

$$\frac{\partial^{k+m}M(t_1,t_2)}{\partial t_1^k \partial t_2^m}|_{t_1=t_2=0} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^m f(x,y)dxdy = E[X^kY^m]$$

と計算できる．以上の議論は，離散型の確率変数の場合にも成り立つ．
¡note¿

2.4 Stochastic Independence / 確率的独立性

定義 2.1 (確率論での独立性 / Stochastically independence)**.** 確率変数 X_1, X_2 が同時確率密度関数 $f(x_1, x_2)$ と周辺分布 $f_1(x_1), f_2(x_2)$ を持つとする．確率変数 X_1, X_2 が確率的に独立 *(stochastically independent)* であるというのは，$f(x_1, x_2) = f_1(x_1)f_2(x_2)$ が成立する時，つまり，同時確率密度関数がそれぞれの確率変数の周辺密度関数の積となることを言う．確率変数が，確率的に独立でないことを，確率的に従属する *(stochastically dependent)* という．

¡note¿

定理 2.1. 確率変数 X_1, X_2 が同時確率密度関数 $f(x_1, x_2)$ を持つとする．X_1 と X_2 は $f(x_1, x_2)$ が x_1 のみの非負値関数と x_2 のみの非負値関数の積でかける時，独立であるという．つまり，

$$f(x_1, x_2) \equiv g(x_1)h(x_2)$$

のようにかけることを言う．ただし，ここで，$g(x_1) > 0, x_1 \in \mathscr{A}, 0\ elsewhere$，$h(x_2) > 0, x_2 \in \mathscr{A}, 0\ elsewhere$ である．

証明 2.1. もし，X_1, X_2 が独立ならば，定義から，同時密度関数 $f(x_1, x_2)$ は周辺密度関数 $f_1(x_1), f_2(x_2)$ を用いて，$f(x_1, x_2) \equiv f_1(x_1)f_2(x_2)$ である．それゆえに，定理の $f(x_1, x_2) \equiv g(x_1)h(x_2)$ が満たされる．

逆に，もし，$f(x_1, x_2) \equiv g(x_1)h(x_2)$ ならば，連続型確率変数の場合，

$$f_1(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_2 = g(x_1)\int_{-\infty}^{\infty} h(x_2)dx_2 = c_1 g(x_1)$$

，

$$f_2(x_2) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1 = h(x_2)\int_{-\infty}^{\infty} g(x_1)dx_1 = c_2 h(x_2)$$

とできる．ただし，ここで c_1, c_2 は定数である．さらに，

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)dx_1 dx_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1 dx_2 \\ &= [\int_{-\infty}^{\infty} g(x_1)dx_1][\int_{-\infty}^{\infty} h(x_2)dx_2] = c_1 c_2 \end{aligned}$$

から，$c_1 c_2 = 1$ であることもわかる．これらの結果は，

$$f(x_1, x_2) \equiv g(x_1)h(x_2) \equiv c_1 g(x_1) c_2 h(x_2) \equiv f_1(x_1) f_2(x_2)$$

を意味し，このことから，逆も成立する． □

¡note¿

定理 2.2. もし X_1, X_2 が独立ならば，

$$Pr(a < X_1 < b, c < X_2 < d) = Pr(a < X_1 < b)Pr(c < X_2 < d)$$

である．

証明 2.2. X_1, X_2 の独立性から，同時密度関数は $f_1(x_1)f_2(x_2)$ である．よって，(連続型確率変数の場合，)

$$\begin{aligned} Pr(a < X_1 < b, c < X_2 < d) &= \int_a^b \int_c^d f_1(x_1)f_2(x_2)dx_1dx_2 \\ &= [\int_a^b f_1 f(x_1)dx_1][\int_c^d f_1 f_2(x_2)dx_2] \\ &= Pr(a < X_1 < b)Pr(c < X_2 < d) \quad \square \end{aligned}$$

離散型の場合は実際にやってみましょう．

¡note¿

定理 2.3. 独立な確率変数 X_1, X_2 がそれぞれ，周辺密度関数 $f_1(x_1), f_2(x_2)$ を持つとする．X_1 の関数 $u(X_1)$ と X_2 の関数 $v(X_2)$ の積の期待値は，その存在性を仮定すると，$u(X_1)$ の期待値，$v(X_2)$ の期待値の積に一致する．つまり．

$$E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)]$$

が成立する．

証明 2.3. X_1, X_2 の独立性から，同時密度関数は $f_1(x_1)f_2(x_2)$ である．ゆえに，数学的期待 *(=俗に言う期待値だと思います)* の定義から，連続型の場合，

$$\begin{aligned} E[u(X_1)v(X_2)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_1)v(x_2)f_1(x_1)f_2(x_2)dx_1dx_2 \\ &= [\int_{-\infty}^{\infty} u(x_1)f_1(x_1)dx_1][\int_{-\infty}^{\infty} v(x_2 f_2(x_2)dx_2] \\ &= E[u(X_1)]E[v(X_2)] \quad \square \end{aligned}$$

離散型の場合は簡単なのでやってみましょう *(again)*.

¡note¿

定理 2.4. X_1, X_2 を同時密度関数 $f(x_1, x_2)$，周辺密度関数 $f_1(x_1), f_2(x_2)$ を持つ確率変数とする．さらに，$M(t_1, t_2)$ をその分布の積率母関数とする．この時，

$$M(t_1, t_2) = M(t_1, 0)M(0, t_2)$$

であるならば，X_1, X_2 は独立である

証明 2.4. もし，X_1, X_2 が独立であるならば，

$$\begin{aligned} M(t_1, t_2) &= E[e^{t_1 X_1 t_2 X_2}] \\ &= E[e^{t_1 X_1} e^{t_2 X_2}] \\ &= E[e^{t_1 X_1}]E[e^{t_2 X_2}] \\ &= M(t_1, 0)M(0, t_2) \end{aligned}$$

であるため，X_1, X_2 の独立性は，同時密度関数の積率母関数が，二つの周辺分布関数の積率母関数に因数分解されるということを表す．

次に，同時密度関数の積率母関数が，$M(t_1, t_2) = M(t_1, 0)M(0, t_2)$ で与えられるとする．いま，X_1 が

$$M(t_1, 0) = \int_{-\infty}^{\infty} e^{t_1 x_1} f_1(x_1) dx_1$$

で与えられる唯一の積率母関数をもち，X_2 が

$$M(0, t_2) = \int_{-\infty}^{\infty} e^{t_2 x_2} f_2(x_2) dx_2$$

で与えられる唯一の積率母関数を持つとする．すると，

$$\begin{aligned} M(t_1, 0)M(0, t_2) &= [\int_{-\infty}^{\infty} e^{t_1 x_1} f_1(x_1) dx_1][\int_{-\infty}^{\infty} e^{t_2 x_2} f_2(x_2) dx_2] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1 x_1 + t_2 x_2} f_1(x_1) f_2(x_2) dx_1 dx_2 \end{aligned}$$

である．また，

$$M(t_1,t_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{t_1x_1+t_2x_2}f_1(x_1)f_2(x_2)dx_1dx_2$$

ここで，$M(t_1,t_2)$ は X_1,X_2 の同時密度関数の積率母関数である．積率母関数の一意性から，

$$f(x_1,x_2)=f_1(x_1)f_2(x_2)$$

がわかる．つまり，もし，$M(t_1,t_2)=M(t_1,0)M(0,t_2)$ ならば，X_1,X_2 は独立である． $\square$

離散型の場合もやってみましょう．
¡note¿

これらの 2 つの確率変数間の独立性は，確率変数が n 個の場合にも拡張できる．しかし，ペアの間で独立性が成り立っていても，それが必ずしも複数の確率変数間の独立性を保証するわけではないということに注意されたい．
¡note¿