2013 年度秋学期　解析応用　第１３回

多次元確率分布と中心極限定理

中心極限定理は，大雑把にいうと「たくさんの独立な確率変数の和は，もとの各々の確率変数がしたがう確率分布が何であっても，概ね正規分布にしたがう」というものです。この定理を証明するには，「独立な確率変数の和のしたがう確率分布」がどうなるかを考える必要があります。そこで，今回は複数の確率変数の関係を扱うための「多次元確率分布」について考えます。

変量の組における，データと確率分布

第１０回の講義で，母集団の相対度数分布と標本の確率分布が同じものであり，各々の標本の値はこの確率分布にしたがって現れるということを説明しました。つまり，この確率分布で確率密度の大きい値は標本に現れやすく，確率密度の小さい値は標本に現れにくくなっています。

これと同じ状況を，複数の項目（**変量**）の組からなるデータの場合に考えます。この場合，標本として得られるのは両変量の値の組（図１の例では数学の点数と英語の点数の組）で，図１の右下のように，各変量を軸とする図（**散布図**）上の点で表されます。このような標本が，母集団の相対度数分布と同じ確率分布にしたがって現れると考えます。

この図の場合の確率分布は，２つの量の組に対して確率分布が対応しているわけですから，図１の左下のように，２つの変量に対応する２つの確率変数の組がしたがう２次元の確率分布になります。この図では，確率密度を「雲」のようなもので表しており，色の濃いところは確率密度が大きいことを表しています。

同時確率分布と周辺確率分布

ここまでに何度か出てきたように，確率変数 X が $a \leq X \leq b$ の範囲に入る確率 $P(a \leq X \leq b)$ は，X の分布関数 $F(x)$ を使って

$$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b dF(x) \tag{1}$$

と表されます。

さて，２つの確率変数 X, Y があるときの確率分布を考えてみましょう。X, Y の値には相関があるかもしれませんから，それぞれ別々の確率分布ではなく，(X, Y) の組がしたがう確率分布を考えなければなりません。そこで，「X, Y それぞれについて，$a \leq X \leq b$ が成り立ち，同時に $c \leq X \leq d$ が成り立つ」確率 $P(a \leq X \leq b, c \leq Y \leq d)$ は，２変量の分布関数 $F_{XY}(x, y)$ を導入して，

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_{y=c}^d \int_{x=a}^b dF_{XY}(x, y) \tag{2}$$

と表されます。このように，「X にある条件が成り立ち，かつ，Y にある条件が成り立つ」確率によってあらわされる確率分布を，X, Y の**同時確率分布**といい，$F_{XY}(x, y)$ を**同時分布関数**といいます。

同時分布関数に対して密度関数 $f_{XY}(x, y)$ が存在する場合は，

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_{y=c}^d \int_{x=a}^b f_{XY}(x, y) dxdy \tag{3}$$

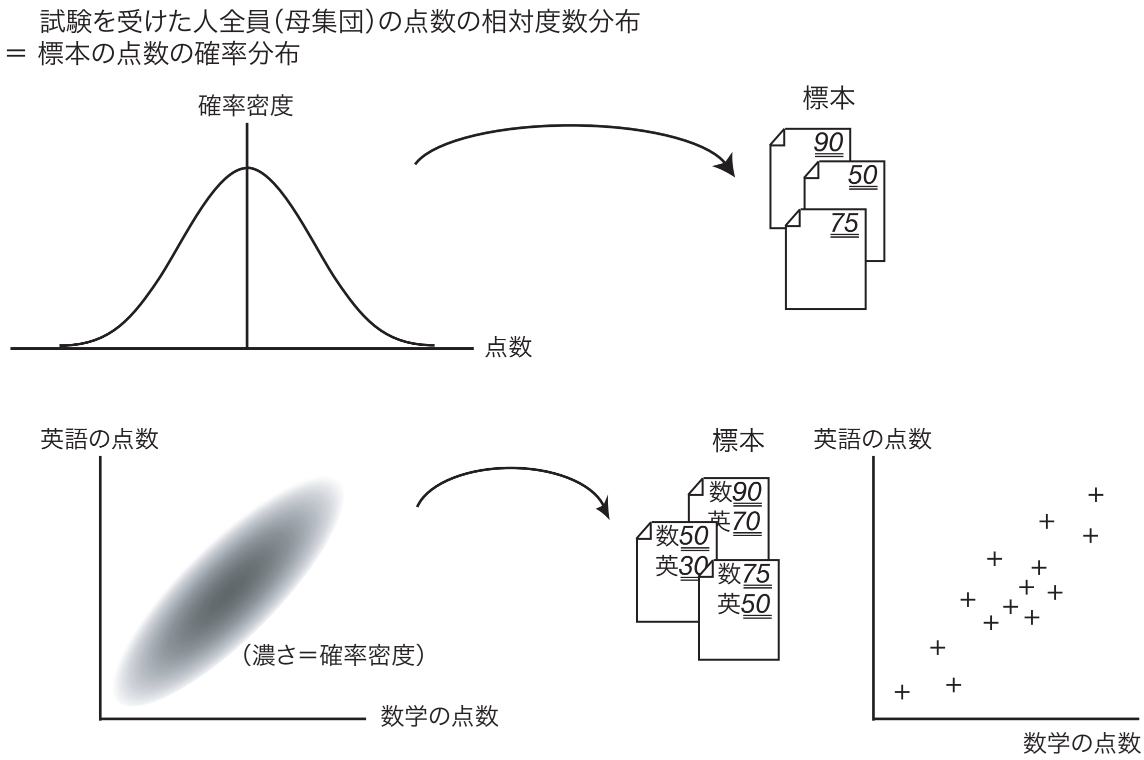

図 1: 多変量におけるデータと確率分布［試験の点数を例にとって］

と表されます。この密度関数を**同時確率密度関数**といいます。図 2 では，同時確率密度関数（立体のヒストグラム）がハット型の立体で表現されています。

以上のように同時確率分布が与えられているとき，X や Y 単独の確率分布はどうなるでしょうか。X 単独の確率分布，すなわち「$x \leq X \leq x + dx$ である」確率とは，「$x \leq X \leq x + dx$ が成り立ち，かつ，Y は何でもよい」確率のことです。「Y は何でもよい」とは $-\infty < Y < \infty$ であることを意味します。$-\infty < Y < \infty$ である確率を求めることは，同時確率密度関数を Y について $-\infty$ から ∞ まで積分することに相当しますから，

$$P(x \leq X \leq x + dx) = \{\int_{y=-\infty}^{\infty} f_{XY}(x, y) dy\} dx \tag{4}$$

と表します。ここで，

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \tag{5}$$

と表すと，(4) 式は $P(x \leq X \leq x + dx) = f_X(x)dx$ となります。このように，同時確率分布から X だけについての確率分布を抜き出したものを X の周辺確率分布といい，$f_X(x)$ を X の**周辺確率密度関数**といいます。同様に，Y の周辺確率密度関数は

$$P(y \leq Y \leq y + dy) = \{\int_{x=-\infty}^{\infty} f_{XY}(x, y) dx\} dx \tag{6}$$

となります。

同時確率密度関数と周辺確率密度関数との関係を，図 3 でみてみましょう。同時確率密度関数 $f_{XY}(x, y)$ は図 2 と同様にハット型の立体で表現されています。これを X および Y で積分したものが，Y および

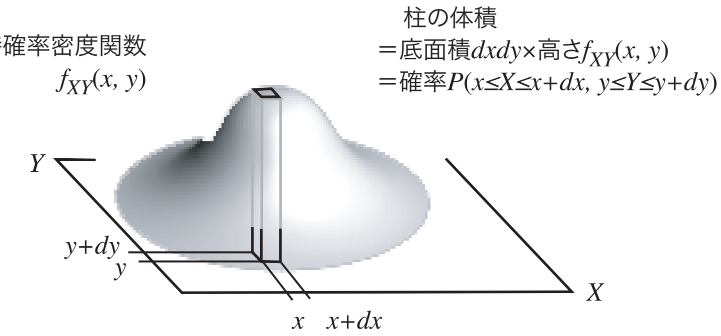

図 2: ２変量の確率密度関数と確率

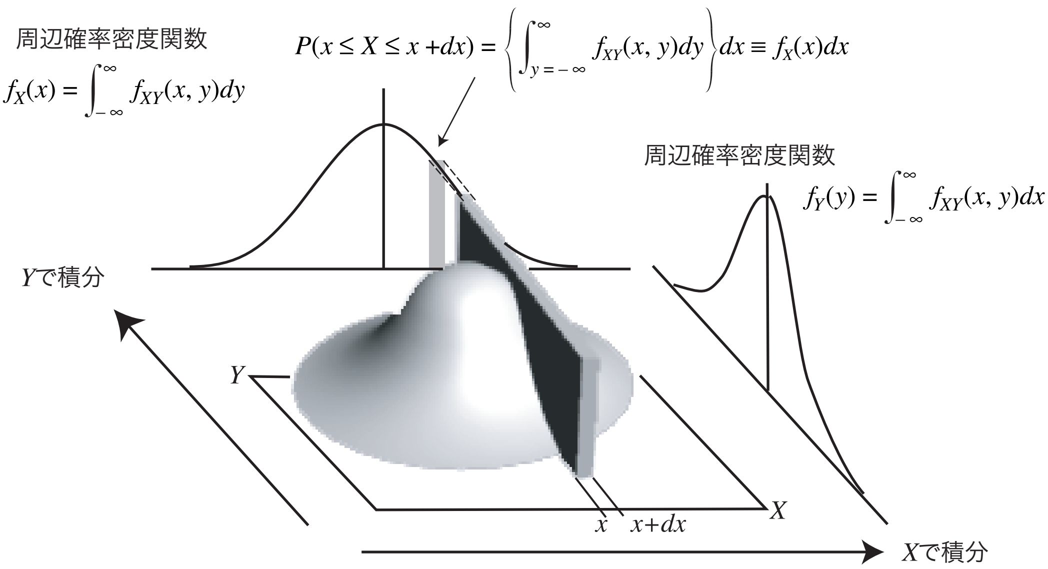

図 3: 同時確率密度関数と周辺確率密度関数

X の周辺確率密度関数 $f_Y(y)$ および $f_X(x)$ です。周辺確率密度関数が，同時確率密度関数を「周辺」に「投影」したものであることがわかると思います。また，(4) 式，(6) 式の関係は，ハット型の立体に挟まった「壁」とその周辺確率密度関数への「影」に相当します。

独立な確率変数

ここまでの知識を使って，「独立な確率変数」について考えます。確率変数 X と Y の同時確率密度関数が X, Y それぞれの周辺確率密度関数になっている場合，すなわち

$$f_X(x)f_Y(x) = f_{XY}(x, y) \tag{7}$$

となるとき，確率変数 X と Y は独立である，といいます。X と Y が互いに独立であるということは，「$X = x$ である確率は Y の値には関係ない」「X と Y の同時確率分布は，X，Y それぞれの周辺確率分布だけで決まる」ことを意味します。

母集団から無作為抽出された，複数個のデータからなる標本は，互いに独立な確率変数のセットと考

えられるので，統計的推測の理論の展開には，独立な確率変数についての性質が多く用いられます。ここでは，確率変数 X と Y が独立の時になりたつ性質のいくつかをあげてみましょう。以下，$E()$ は期待値，$V()$ は分散を意味します。

1. $E(XY) = E(X)E(Y)$ が成り立ちます（証明は演習とします）。
2. $(X + Y)$ のモーメント母関数と X, Y それぞれのモーメント母関数との間には，$M_{X+Y}(t) = M_X(t)M_Y(t)$ がなりたちます。証明は，$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}]$ であることから，1. の証明の X と Y をそれぞれ e^{tX} と e^{tY} に置き換えればできます。
3. $V(X + Y) = V(X) + V(Y)$ が成り立ちます。この関係は，第１０回の「大数の法則」の説明で用いました（証明は付録をみてください）。
4. 確率変数 $X + Y$ の確率密度関数はどうなるでしょうか？ これを求めるには，$X + Y = z$ のときの確率密度を求めればよいわけです。「$X + Y = z$」とは，「$X = x, Y = z - x$ で，x はどんな値でもよい」ということと同じです。(7) 式で，y を $z - x$ と書き直すと，

$$f_{XY}(x, z - x) = f_X(x)f_Y(z - x) \tag{8}$$

となります。この式は「$X = x, Y = z - x$ のときの確率密度」を表していますが，x はどんな値でもよいので，求める確率密度 $f_{X+Y}(z)$ は (8) 式をすべての x について合計したもの，すなわち

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{XY}(x, z - x)dx = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)dx \tag{9}$$

となります。この計算は，つまり $f_X(x)$ と $f_Y(y)$ のコンヴォリューション（たたみこみ積分）$[f_X * f_Y](z)$ です。

中心極限定理

中心極限定理は，

> $X_1, X_2, \ldots, X_n$ を，互いに独立で期待値 μ，分散 σ^2 の同じ確率分布にしたがう確率変数とし，また $\bar{X}_n = (X_1 + X_2 + \cdots + X_n)/n$ とする。このとき，$\dfrac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}}$ がしたがう確率分布は，$n \to \infty$ のとき標準正規分布 $N(0, 1)$ に収束する[1]。したがって，n が十分大きいとき，$\bar{X}_n$ は正規分布 $N(\mu, \sigma^2/n)$ で近似できる。

というものです。とくに，$X_1, X_2, \ldots, X_n$ の各々がしたがう確率分布は何であってもよい[2]，ということが重要です。正規分布が統計的現象のあちこちに現れる理由は，まさにここにあります。

今日の講義では，中心極限定理の証明を説明します。ただ，中心極限定理を完全に証明するには高度な確率論の知識が必要ですので，ここでは，これまでに説明した知識だけで説明できる，証明の概略を示すことにします。以下の証明では，

[1]ここでの「収束」は，大数の法則の説明に出てきた「確率収束」とは違い，「分布の形が，ある形（ここでは標準正規分布）収束する」という「法則収束」です。

[2]細かいことをいえば，各々の確率変数のしたがう確率分布が違っていてもよかったり，逆に確率分布に制約がついたりもしますが，本文の内容以上の詳細は省略します。

1. $X_1, X_2, \ldots, X_n$ がしたがう確率分布のモーメント母関数が存在する

2. モーメント母関数に対して確率分布が一意に定まる

3. モーメント母関数の列があるモーメント母関数に収束するとき，それらに対応する確率分布も，収束したモーメント母関数に対応する確率分布に収束する

の３点を認めてください。

まず，$Y_1 = (X_1 - \mu)/\sigma, Y_2 = (X_2 - \mu)/\sigma, \ldots, Y_n = (X_n - \mu)/\sigma$ とおきます。こうすると，$E(Y_1) = E(Y_2) = \cdots = E(Y_n) = 0$，$V(Y_1) = V(Y_2) = \cdots = V(Y_n) = 1$ となります。ここで，Y_1 のモーメント母関数を $M_{Y1}(t)$ とすると，その１階微分・２階微分について

$$\begin{aligned} M'_{Y_1}(t)\big|_{t=0} &= E(Y_1) = 0 \\ M''_{Y_1}(t)\big|_{t=0} &= E(Y_1^2) = V(Y_1) + (E(Y_1))^2 = 1 \end{aligned} \tag{10}$$

が成り立ちます（下段の式では，付録の (A3) 式で示されている $V(Y_1) = E(Y_1^2) - (E(Y_1))^2$ という性質を用いました)。

１階微分，２階微分が (10) 式の条件を満たす関数 $M_{Y_1}(t)$ は，$t \to 0$ のとき $g(t) \to 0$ となる任意の関数 $g(t)$ を使って

$$M_{Y_1}(t) = 1 + \frac{t^2}{2} + g(t) \cdot t^2 \tag{11}$$

と表せます。

さて，$T = Y_1 + Y_2 + \cdots + Y_n$ とおくと，$Y_1, Y_2, \ldots, Y_n$ は独立なので，前節の「確率変数 X と Y が独立の時になりたつ性質」の 2. で述べたように，T のモーメント母関数 $M_T(t)$ は $Y_1, Y_2, \ldots, Y_n$ の各々のモーメント母関数の積になり，

$$M_T(t) = M_{Y_1}(t) \cdot M_{Y_2}(t) \cdot \cdots \cdot M_{Y_n}(t) = \left\{1 + \frac{t^2}{2} + g(t) \cdot t^2\right\}^n \tag{12}$$

となります。ですから，$T/\sqrt{n}$ のモーメント母関数は (12) 式で t を $t/\sqrt{n}$ で置き換えることで得られ，

$$\begin{aligned} M_T(\frac{t}{\sqrt{n}}) &= \left\{1 + \frac{\left(\frac{t}{\sqrt{n}}\right)^2}{2} + g(\frac{t}{\sqrt{n}}) \cdot \left(\frac{t}{\sqrt{n}}\right)^2\right\}^n \\ &= \left\{1 + \frac{\left(\frac{t}{2}\right)}{n} + \frac{g(\frac{t}{\sqrt{n}}) \cdot t^2}{n}\right\}^n \end{aligned} \tag{13}$$

となります。ここで $n \to \infty$ とすると，

$$M_T(\frac{t}{\sqrt{n}}) \to \exp\left(\frac{t^2}{2}\right) \tag{14}$$

となります（理由は解析学の教科書を見てください)。付録に示すように，$\exp(t^2/2)$ は標準正規分布 $N(0,1)$ のモーメント母関数ですから，$T/\sqrt{n}$ の分布は $n \to \infty$ のとき $N(0,1)$ に収束することがわかり

ます。ここで

$$\begin{aligned}\frac{T}{\sqrt{n}} &= \frac{Y_1+Y_2+\cdots+Y_n}{\sqrt{n}} \\ &= \frac{\frac{X_1-\mu}{\sigma}+\frac{X_2-\mu}{\sigma}+\cdots\frac{X_n-\mu}{\sigma}}{\sqrt{n}} \\ &= \frac{(X_1+X_2+\cdots+X_n)-n\mu}{\sqrt{n\sigma^2}} \qquad (15)\end{aligned}$$

であり，これが $N(0,1)$ にしたがいます。ですから，n が大きいとき $(X_1+X_2+\cdots+X_n)/n$ の分布は $N(\mu,\sigma^2/n)$ で近似できます。

今日の演習

$E(X)$ を確率変数 X の期待値，$M_X(t)$ を確率変数 X がしたがう確率分布のモーメント母関数とします。連続型確率変数 X と Y が独立のとき，
1. $E(XY)=E(X)E(Y)$ を証明してください。
2. $M_{X+Y}(t)=M_X(t)M_Y(t)$ を証明してください。

付録

独立な確率変数の和の分散

２つの確率変数の和の分散 $V(X+Y)$ を計算するには，まず２つの確率変数の和の期待値 $E(X+Y)$ を計算する必要があります。

２つの確率変数 X,Y については，「$x\le X\le x+dx$ が成り立ち，同時に $y\le Y\le y+dy$ が成り立つ」確率が，$f_{XY}(x,y)dxdy$ となります。したがって，「$X+Y$」という確率変数を考えるとき，［確率変数のとりうる値 $\times$ その値をとる確率］は $(x+y)f_{XY}(x,y)dxdy$ で表されます。ですから，確率変数「$X+Y$」の期待値 $E(X+Y)$ は，

$$\begin{aligned}E(X+Y) &= \iint_{x,y}(x+y)f_{XY}(x,y)dxdy \\ &= \iint_{x,y}xf_{XY}(x,y)dxdy+\iint_{x,y}yf_{XY}(x,y)dxdy \\ &= \int_x x\left\{\int_y f_{XY}(x,y)dy\right\}dx+\int_y y\left\{\int_x f_{XY}(x,y)dx\right\}dy \\ &= \int_x xf_X(x,y)dx+\int_y yf_Y(x,y)dy \\ &= E(X)+E(Y) \qquad (A1)\end{aligned}$$

となります。同様の計算で，$E(X+Y)^2=E(X^2)+E(Y^2)+2E(XY)$ も導かれます。これらをもちい

て $V(X+Y)$ を求めますが，ここで

$$\begin{aligned} E[(X-E(X))(Y-E(Y))] &= E[XY - E(X)Y - E(Y)X + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned} \tag{A2}$$

が成り立ち ($E(X)$ や $E(Y)$ は $E[]$ の計算においては定数であることに注意)，さらに (A2) 式で $X=Y$ とすると

$$E[(X-E(X))^2] = V(X) = E(X^2) - (E(X))^2 \tag{A3}$$

となるので，これを使って

$$\begin{aligned} V(X+Y) &= E\{(X+Y)^2\} - \{E(X+Y)\}^2 \\ &= \left[E(X^2) + E(Y^2) + 2E(XY)\right] - \left[\{E(X)\}^2 + 2E(X)E(Y) + \{E(Y)\}^2\right] \\ &= \left[E(X^2) - \{E(X)\}^2\right] + \left[E(Y^2) - \{E(Y)\}^2\right] + 2\left[E(XY) - E(X)E(Y)\right] \\ &= V(X) + V(Y) + 2\left[E(XY) - E(X)E(Y)\right] \end{aligned} \tag{A4}$$

となります。本文のとおり，X と Y が独立のときは $E(XY) = E(X)E(Y)$ ですから，(A4) 式から $V(X+Y) = V(X) + V(Y)$ となります。

さらに (A2) 式を使うと，

$$V(X+Y) = V(X) + V(Y) + 2E[(X-E(X))(Y-E(Y))] \tag{A5}$$

が得られます。(A5) 式の最後の項の $E[(X-E(X))(Y-E(Y))]$ を確率変数 X と Y の**共分散**とよび，$Cov(X,Y)$ と書きます。また，

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \tag{A6}$$

を確率変数 X と Y の**相関係数**とよびます。X と Y が独立でなくても，$Cov(X,Y) = 0$ であれば，$V(X+Y) = V(X) + V(Y)$ が成り立ちます。これを「X と Y は**無相関**である」といいます。

確率変数の共分散の意味を知るため，図 A1 を見てください。各図で雲状の絵は，図 1 と同様に，確率変数 X と Y の同時確率密度関数でグレーの濃いところが値が大きいと思ってください。(a) では，「X，Y ともそれぞれその平均 $E(X)$，$E(Y)$ よりも大きい」または「両者とも各々の平均よりも小さい」という値の確率密度が大きい，つまりそういう値がこの確率分布では起こりやすいことを意味しています。したがって，$(X-E(X))(Y-E(Y))$ の値が正になるデータ組が多くなり，共分散 $Cov(X,Y) = E[(X-E(X))(Y-E(Y))]$ は正で大きくなります。同様に考えて，(b) では $(X-E(X))(Y-E(Y))$ の値が負になるデータ組が多くなるので，共分散の値は負でその絶対値が大きくなります。(c) はどちらでもない場合で，このときは $Cov(X,Y) = 0$，すなわち無相関となります。

正規分布 $N(\mu, \sigma^2)$ のモーメント母関数

モーメント母関数の定義から，

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{2\sigma^2 tx - (x-\mu)^2}{2\sigma^2}\right\} dx \end{aligned} \tag{A7}$$

となります。ここで，$\exp\{\}$ の中の分子の部分を完全平方にすると，

$$
\begin{aligned}
2\sigma^2 tx-(x-\mu)^2 &= -\left[x^2-2\mu x+\mu^2-2\sigma^2 tx\right] \\
&= -\left[x^2-2(\mu+\sigma^2 t)x+(\mu+\sigma^2 t)^2-(\mu+\sigma^2 t)^2+\mu^2\right] \\
&= -(x-(\mu+\sigma^2 t))^2+(2\mu\sigma^2 t+\sigma^4 t^2) \qquad \text{(A8)}
\end{aligned}
$$

となりますから，(A7) 式は

$$
\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x-(\mu+\sigma^2 t))^2+(2\mu\sigma^2 t+\sigma^4 t^2)}{2\sigma^2}\right\}dx \\
&= \exp\left\{\frac{2\mu\sigma^2 t+\sigma^4 t^2}{2\sigma^2}\right\}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x-(\mu+\sigma^2 t))^2}{2\sigma^2}\right\}dx \\
&= \exp\left\{\mu t+\frac{\sigma^2 t^2}{2}\right\}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{\frac{-(x-(\mu+\sigma^2 t))^2}{2\sigma^2}\right\}dx \qquad \text{(A9)}
\end{aligned}
$$

となります。ここで，式 (A9) の積分は正規分布 $N(\mu+\sigma^2 t,\sigma^2)$ の確率密度関数を全実数で積分したものですから，1 です[3]。よって，モーメント母関数は

$$
M_X(t)=\exp\left\{\mu t+\frac{\sigma^2 t^2}{2}\right\} \qquad \text{(A10)}
$$

となります。

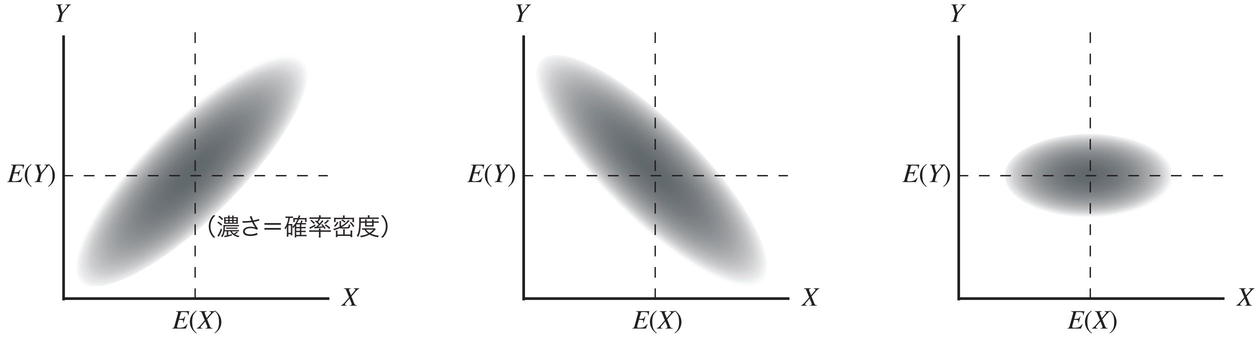

図 A1: 確率変数の共分散

[3]確かに 1 になることの証明は省略します。求める積分の 2 乗を重積分で表して，さらに変数を極座標に変換すると求められます。解析学の本を参考にしてください。