2008年度後期　情報統計学　第１２回
多次元確率分布と中心極限定理

ここまでの講義で，何度か「中心極限定理」が出てきました．中心極限定理は，「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．このことは，講義の他の場面でも出てきましたが，説明を省略してきました．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

この講義の前半では，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が高い値は標本に現れやすく，確率密度が低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

第６回の講義で説明したように，連続型確率変数 X が $a \le X \le b$ の範囲に入る確率 $P(a \le X \le b)$ は，X がしたがう確率密度関数 $f(x)$（要するにヒストグラム）のグラフの下の部分の面積，すなわち「$f(x)$ の a から b までの積分」で表され（図2），

$$P(a \le X \le b) = \int_a^b f(x)dx \tag{1}$$

と書きます．

この式の意味は，次のように考えられます．図3のように，X の非常に微小な幅 dx を考え，x と $x+dx$ の間では $f(x)$ の値は変化しないと考えます．そうすると，連続型確率変数 X が $x \le X \le x+dx$ の範囲に入る確率 $P(x \le X \le x+dx)$ は，図１のように本来「確率＝ヒストグラムの柱の面積」でしたから，この場合も幅 dx の柱と考えて

$$P(x \le X \le x+dx) = f(x)dx \tag{2}$$

となります．連続型確率変数 X が $a \le X \le b$ の範囲に入る確率 $P(a \le X \le b)$ は，この柱の面積を $a \le X \le b$ の範囲で合計したもので，図２右および図３のグレーの範囲となり，これを (1) 式のように書いて積分と呼んでいるのです．

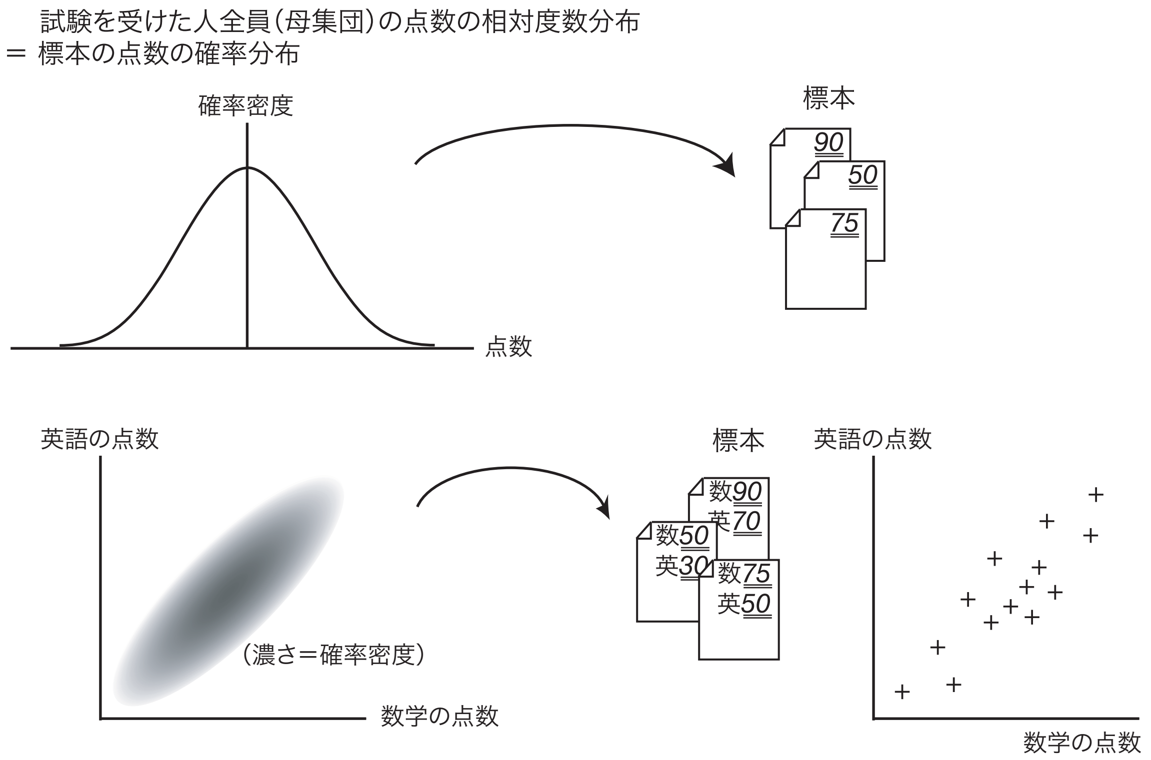

図 1: 多変量におけるデータと確率分布［試験の点数を例にとって］

さて，２つの連続型確率変数 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，２変量の確率密度関数 $f_{XY}(x, y)$ を導入して，微少な dx, dy にたいして「$x \leq X \leq x + dx$ が成り立ち，同時に $y \leq Y \leq y + dy$ が成り立つ」確率を考えます．図 3 と同様に考えて，図 4 のように２変量による立体のヒストグラムの柱を考えたとき，底が $x \leq X \leq x + dx, y \leq Y \leq y + dy$ の範囲（底面積 $dxdy$）にある柱では確率密度関数 $f_{XY}(x, y)$ の値は変化しないとします．このとき，求める確率は「確率＝ヒストグラムの柱の体積」となって

$$P(x \leq X \leq x + dx, y \leq Y \leq y + dy) = f(x, y)dxdy \tag{3}$$

と表されます．したがって，「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 $P(a \leq X \leq b, c \leq X \leq d)$ は，この柱の体積を $a \leq X \leq b, c \leq X \leq d$ の範囲で合計したもので

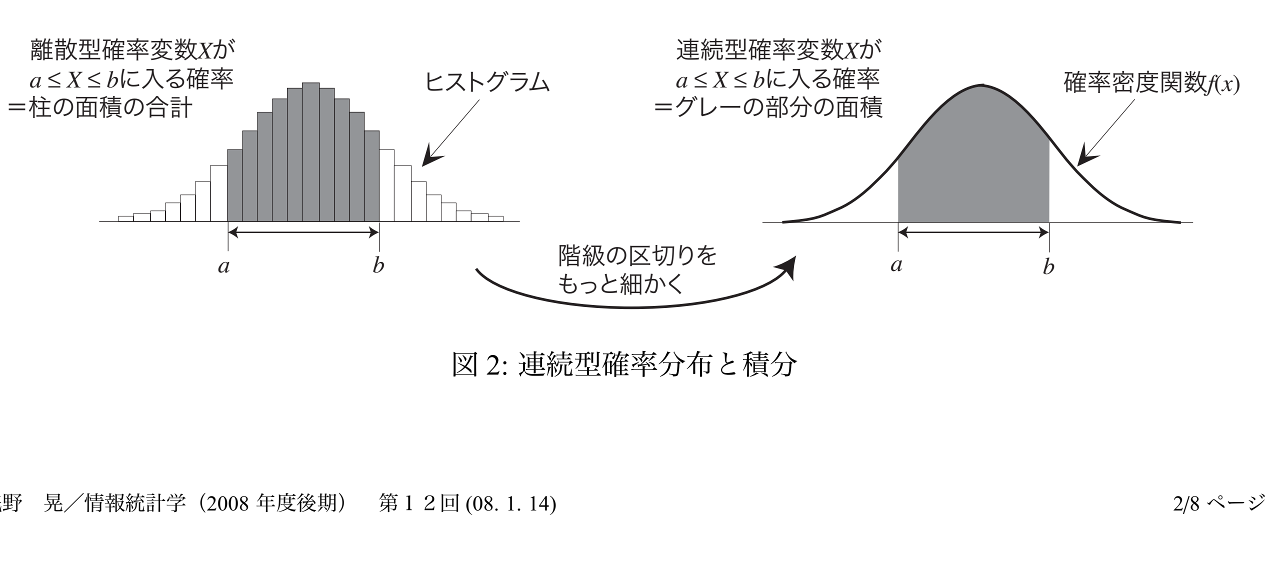

図 2: 連続型確率分布と積分

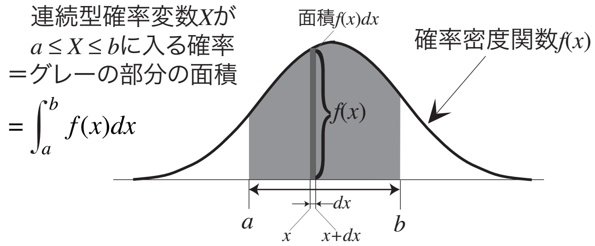

図 3: 積分の意味

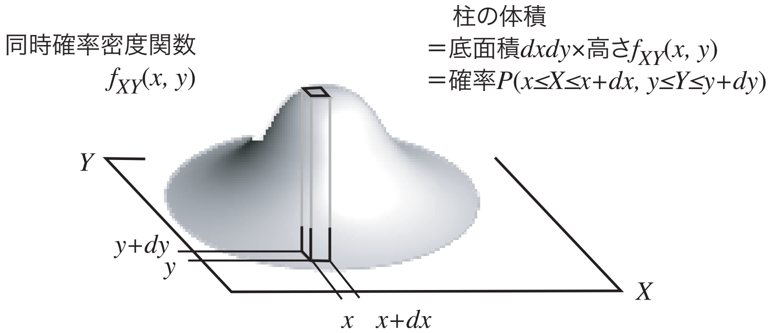

図 4: ２変量の確率密度関数と確率

す．この計算は数学では重積分とよばれ，

$$P(a \le X \le b, c \le X \le d) = \int_{y=c}^{d} \int_{x=a}^{b} f_{XY}(x, y) dxdy \tag{4}$$

と書きます．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，この $f_{XY}(x, y)$ を同時確率密度関数といいます．図 4 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「$x \le X \le x + dx$ である」確率とは,「$x \le X \le x + dx$ 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，

$$P(x \le X \le x + dx) = \{\int_{y=-\infty}^{\infty} f_{XY}(x, y) dy\} dx \tag{5}$$

と表します．ここで，

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \tag{6}$$

と表すと，(5) 式は $P(x \le X \le x + dx) = f_X(x)dx$ 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

$$P(y \le Y \le y + dy) = \{\int_{x=-\infty}^{\infty} f_{XY}(x, y) dx\} dx \tag{7}$$

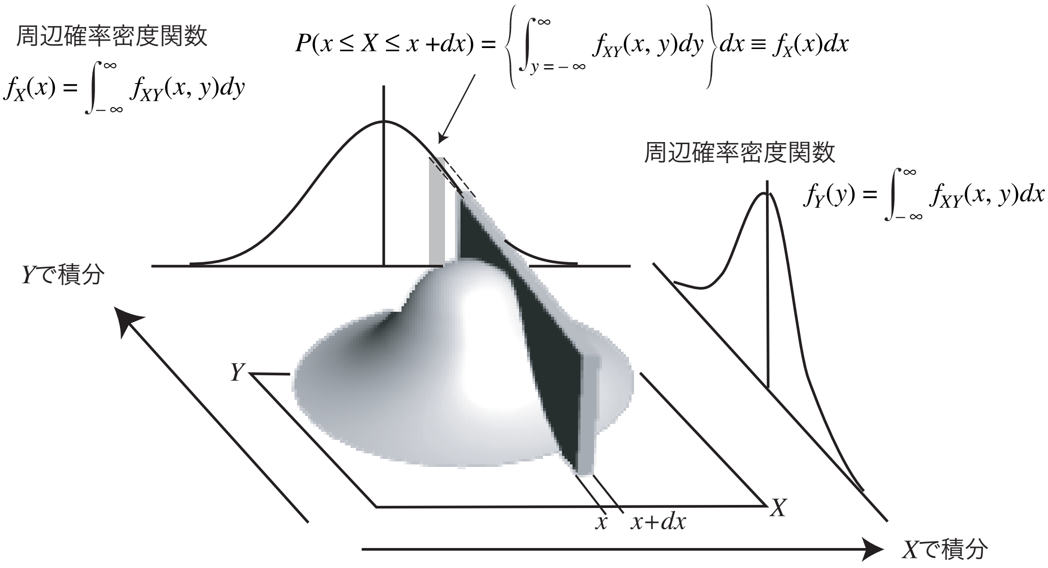

図 5: 同時確率密度関数と周辺確率密度関数

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 5 でみてみましょう．同時確率密度関数 $f_{XY}(x, y)$ は図 4 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 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，(5) 式，(7) 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

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

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４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．(8) 式で，y を $z-x$ と書き直すと，

$$f_{XY}(x, z-x) = f_X(x) f_Y(z-x) \tag{9}$$

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 (9) 式をすべての x について合計したもの，すなわち

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{XY}(x, z-x)dx = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx \tag{10}$$

となります．この計算を $f_X(x)$ と $f_Y(y)$ の**コンヴォリューション（たたみこみ積分）**といい，$[f_X * f_Y](z)$ と書きます．

中心極限定理

中心極限定理は，

> $X_1, X_2, \ldots, X_n$ を，互いに独立で平均 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．第６回の講義で触れたように，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，第６回のときには説明しなかった，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. $X_1, X_2, \ldots, X_n$ 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

[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」という意味です．これを「法則収束」といいます．

[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．

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

$$
\begin{aligned}
M'_{Y_1}(t)\Big|_{t=0} &= E(Y_1) = 0 \\
M''_{Y_1}(t)\Big|_{t=0} &= E(Y_1^2) = V(Y_1) + (E(Y_1))^2 = 1
\end{aligned} \tag{11}
$$

が成り立ちます（下段の式では，付録の (A4) 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 (11) 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

$$
M_{Y_1}(t) = 1 + \frac{t^2}{2} + g(t) \cdot t^2 \tag{12}
$$

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，

$$
M_T(t) = M_{Y_1}(t) \cdot M_{Y_2}(t) \cdot \cdots \cdot M_{Y_n}(t) = \left\{1 + \frac{t^2}{2} + g(t) \cdot t^2\right\}^n \tag{13}
$$

となります．ですから，$T/\sqrt{n}$ のモーメント母関数は (13) 式で t を $t/\sqrt{n}$ で置き換えることで得られ，

$$
\begin{aligned}
M_T(\frac{t}{\sqrt{n}}) &= \left\{1 + \frac{\left(\frac{t}{\sqrt{n}}\right)^2}{2} + g(\frac{t}{\sqrt{n}}) \cdot \left(\frac{t}{\sqrt{n}}\right)^2\right\}^n \\
&= \left\{1 + \frac{\left(\frac{t}{2}\right)}{n} + \frac{g(\frac{t}{\sqrt{n}}) \cdot t^2}{n}\right\}^n
\end{aligned} \tag{14}
$$

となります．ここで $n \to \infty$ とすると，

$$
M_T(\frac{t}{\sqrt{n}}) \to \exp\left(\frac{t^2}{2}\right) \tag{15}
$$

となります（理由は解析学の教科書を見てください）．第６回の正規分布の説明（付録）で示したように，$\exp(t^2/2)$ は標準正規分布 $N(0,1)$ のモーメント母関数ですから，T/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

$$
\begin{aligned}
\frac{T}{\sqrt{n}} &= \frac{Y_1 + Y_2 + \cdots + Y_n}{\sqrt{n}} \\
&= \frac{\frac{X_1-\mu}{\sigma} + \frac{X_2-\mu}{\sigma} + \cdots \frac{X_n-\mu}{\sigma}}{\sqrt{n}} \\
&= \frac{(X_1 + X_2 + \cdots + X_n) - n\mu}{\sqrt{n\sigma^2}}
\end{aligned} \tag{16}
$$

であり，これが $N(0,1)$ にしたがいます．ですから，第６回の「正規分布の性質」で説明したように，n が大きいとき $X_1 + X_2 + \cdots + X_n$ の分布は $N(n\mu, n\sigma^2)$ で近似でき，$(X_1 + X_2 + \cdots + X_n)/n$ の分布は $N(\mu, \sigma^2/n)$ で近似できます．

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．

第３回の「確率変数の期待値」の説明で，確率変数の期待値は「［確率変数のとりうる値×その値をとる確率］の和」であることを説明しました．１変量の連続型確率変数 X が $x \le X \le x+dx$ の範囲に入る確率 $P(x \le X \le x+dx)$ は，(2) 式の通り，確率密度関数 $f(x)$ を使って $f(x)dx$ と表されます．ですから，期待値 $E(X)$ は，「［確率変数のとりうる値（x）×その値をとる確率（$f(x)dx$）］の和」，すなわち

$$E(X) = \int_{x=-\infty}^{\infty} x f(x) dx \tag{A1}$$

となります．これは第６回の連続型確率分布のところで説明した通りです．

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，(3) 式の通り $f_{XY}(x,y)dxdy$ です．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×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，

$$\begin{aligned}
E(X+Y) &= \iint_{x,y} (x+y) f_{XY}(x,y)dxdy \\
&= \iint_{x,y} x f_{XY}(x,y)dxdy + \iint_{x,y} y f_{XY}(x,y)dxdy \\
&= \int_x x\left\{\int_y f_{XY}(x,y)dy\right\}dx + \int_y y\left\{\int_x f_{XY}(x,y)dx\right\}dy \\
&= \int_x x f_X(x,y)dx + \int_y y f_Y(x,y)dy \\
&= E(X) + E(Y)
\end{aligned} \tag{A2}$$

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

$$\begin{aligned}
E[(X-E(X))(Y-E(Y))] &= E[XY - E(X)Y - E(Y)X + E(X)E(Y)] \\
&= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned} \tag{A3}$$

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 (A3) 式で $X = Y$ とすると

$$E[(X-E(X))^2] = V(X) = E(X^2) - (E(X))^2 \tag{A4}$$

となるので，これを使って

$$\begin{aligned}
V(X+Y) &= E\{(X+Y)^2\} - \{E(X+Y)\}^2 \\
&= \left[E(X^2) + E(Y^2) + 2E(XY)\right] - \left[\{E(X)\}^2 + 2E(X)E(Y) + \{E(Y)\}^2\right] \\
&= \left[E(X^2) - \{E(X)\}^2\right] + \left[E(Y^2) - \{E(Y)\}^2\right] + 2\left[E(XY) - E(X)E(Y)\right] \\
&= V(X) + V(Y) + 2\left[E(XY) - E(X)E(Y)\right]
\end{aligned} \tag{A5}$$

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，(A5) 式から $V(X+Y) = V(X) + V(Y)$ となります．

さらに (A3) 式を使うと，

$$V(X+Y) = V(X) + V(Y) + 2E[(X - E(X))(Y - E(Y))] \tag{A6}$$

が得られます．(A6) 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，

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \tag{A7}$$

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．

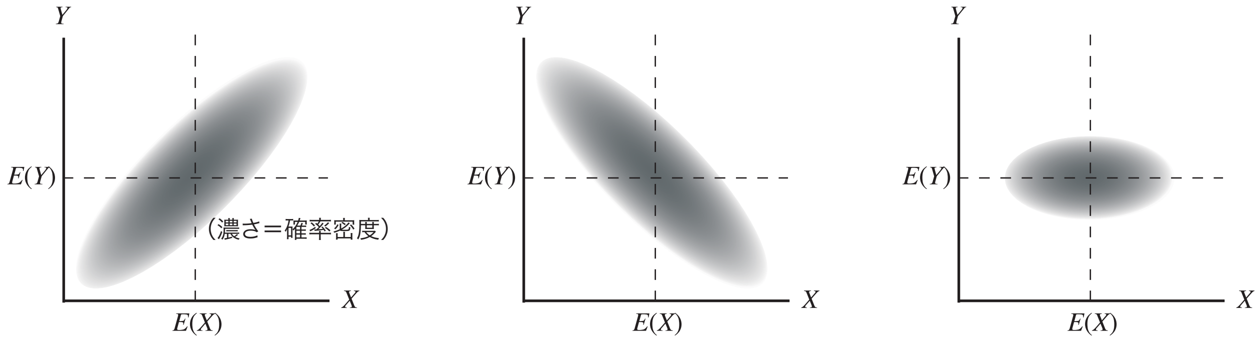

図 A1: 確率変数の共分散