本読みゼミ第二回

多次元の確率分布・大数の法則と中心極限定理

高橋 優輝

森田研究室 B4

2022.4.21

目次

目次

同時確率分布

2 変数で考える．一般に 2 つの離散型確率変数 X, Y があるとし，これらを二次元のベクトル (X, Y) として表す．このとき，$X = x$ であり，同時に $Y = y$ である確率

$$P(X = x, Y = y) = f(x, y) \tag{7.1}$$

を 2 次元確率変数 (X, Y) の**同時確率分布** (joint probability distribution) という．
$f(x, y)$ は 1 次元のときと同じく，

$$f(x, y) \geq 0 \qquad \text{かつ} \qquad \sum_x \sum_y f(x, y) = 1 \tag{7.2}$$

を満たさなければならない．

同時確率分布

2 次元の確率変数の場合，**事象**も 2 次元空間の中にある．事象とは，一般に (x, y) の集まったある部分集合である．それを A とする．その確率 $P(A)$ は，以下の式で求められる．

$$P((X, Y) \in A) = \sum_{(x,y)\in A} f(x, y) \tag{7.3}$$

同時確率分布

X, Y が連続型の確率変数のときには，$f(x,y)$ は 2 次元の確率密度関数で，**同時確率密度関数** (joint probability density function) と呼ばれ，

$$f(x,y) \geq 0 \text{ かつ } \int\int_S f(x,y)dxdy = 1 \tag{7.4}$$

を満たす．ここで，S は**標本空間** (sample space)，すなわち確率が定義される基礎となる集合のことで，2 次元ユークリッド空間 (平面) の全範囲のことである．この $f(x,y)$ によって，事象 $A(S$ の部分集合 $)$ の確率は，積分で，

$$P((X,Y) \in A) = \int\int_A f(x,y)dxdy \tag{7.5}$$

と定義される．とくに，A が区間ならば，式 (7.5) は次のようになる．

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)dxdy \tag{7.6}$$

である．

周辺確率分布

X, Y 単独の確率分布は，離散型の場合，

$$g(x) = \sum_{y} f(x, y), \quad h(y) = \sum_{x} f(x, y) \tag{7.7}$$

で求められる．それぞれ X, Y の**周辺確率分布** (marginal probability distribution) と呼ばれる．
連続型の場合は，

$$g(x) = \int_{-\infty}^{\infty} f(x, y)dy, \quad h(y) = \int_{-\infty}^{\infty} f(x, y)dx \tag{7.8}$$

で与えられる．これらを**周辺確率密度関数** (marginal probability density function) という．この場合も，周辺確率密度関数が与える確率分布を周辺確率分布という．

周辺確率分布

離散型の場合は，同時確率分布 $f(x,y)$ から，X,Y の周辺確率分布 $g(x),h(y)$ を，連続型の場合は同時確率密度関数 $f(x,y)$ から，X,Y の周辺確率密度関数 $g(x),h(y)$ を求めることは可能であるが，$g(x),h(y)$ から，$f(x,y)$ を求めることは不可能である．(一意に定まらない．)
e.g.)
実現象に置き換えて考えてみても，父の身長と子の身長の関係を知るために父の身長の分布，子の身長の分布をそれぞれ知っても，何の役にも立たない．必要なのは，(父の身長，子の身長) の 2 次元の分布である．

共分散と相関係数

2 変数 X, Y の間に関連があれば，一方の変化は他方に及ぶと考えられるので，分散には単純な加法は成立しないと考えられる．すなわち，

$$V(X+Y) \neq V(X)+V(Y) \tag{7.9}$$

が成立すると考えられる．

共分散と相関係数

実際，$\mu_X = E(X), \mu_Y = E(Y), \mu_{X+Y} = E(X+Y) = \mu_X + \mu_Y$ として，定義に従って計算すると，

$$\begin{aligned}
V(X+Y) &= E[\{(X+Y) - \mu_{X+Y}\}^2] \\
&= E[\{(X-\mu_X) + (Y-\mu_Y)\}^2] \\
&= E[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)] \\
&= E[(X-\mu_X)^2] + E[(Y-\mu_Y)^2] + 2E[(X-\mu_X)(Y-\mu_Y)] \\
&= V(X) + V(Y) + 2Cov(X,Y) \qquad (7.10)
\end{aligned}$$

が成立する．
ただし，

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \qquad (7.11)$$

である．

共分散と相関係数

式 (7.11) で与えられる $Cov(X, Y)$ は X, Y の**共分散** (covariance) と呼ばれ，X と Y が，それぞれの平均 μ_X, μ_Y から互いに関連しながら，ばらつく程度を表す．
$X - \mu_X, Y - \mu_Y$ の正負の符号の全体 (平均) 的傾向から，$Cov(X, Y) > 0$ なら，X, Y は大小が同傾向，< 0 なら反対傾向の関係となる．

共分散と相関係数

共分散 $Cov(X, Y)$ は，X, Y の関係の方向を表すが，その強さの程度を判断する基準がない．そこで，この値を標準偏差で割って調整し，確率変数 X, Y の**相関係数** (correlation coefficient) を

$$\rho = \frac{Cov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \tag{7.12}$$

と定義する．ρ は必ず $-1 \leq \rho \leq 1$ の範囲に入るから，関連の絶対的な程度が -1 と 1 の間の数で表される．

共分散と相関係数

Proof.

$$
\begin{aligned}
Q(t) &= V(tX+Y) \\
&= E\left[\{(tX+Y)-E(tX+Y)\}^2\right] \\
&= E\left[\{t(X-\mu_X)+(Y-\mu_Y)\}^2\right] \\
&= t^2E[(X-\mu_X)^2]+2tE[(X-\mu_X)(Y-\mu_Y)]+E[(Y-\mu_Y)^2] \\
&= t^2V(X)+2tCov(X,Y)+V(Y)
\end{aligned}
$$

は t の値に依らず非負の値をとるので，この t についての 2 次式の判別式 D について $D \leq 0$ が必要より，

$$
\frac{D}{4} = Cov(X,Y)^2 - V(X)V(Y) \leq 0
$$

$$
\iff 1 \geq \frac{Cov(X,Y)^2}{V(X)V(Y)} \iff 1 \geq \rho^2 \iff -1 \leq \rho \leq 1
$$

共分散と相関係数

ρ の定義から，$\rho > 0$ なら，X, Y は同じ大小の向きに変化する傾向があり，$\rho < 0$ なら逆である．ここでいう傾向は平均的，確率的傾向であるが，$|\rho|$ が大きくなると確定的な関係に近づく．
最も極端な場合は，$\rho = \pm 1$ であり，このときは，X, Y の間には厳密に次の 1 次式の関係が成り立つ．

$$Y = aX + b \quad (\text{ただし，}\rho = 1 \text{ なら } a > 0, \rho = -1 \text{ なら } a < 0) \tag{7.13}$$

一方，$\rho = 0(\iff Cov(X, Y) = 0)$ のとき，X, Y は**無相関** (uncorrelated) であるという．

共分散と相関係数

$a \neq 0$ のとき，$|\rho| = 1 \iff Y = aX + b$ を示す．

Proof: $|\rho| = 1 \implies Y = aX + b.$

$\rho = 1$ のとき，

$$V\left(\frac{X}{\sqrt{V(X)}} - \frac{Y}{\sqrt{V(Y)}}\right) = \frac{V(X)}{V(X)} + \frac{V(Y)}{V(Y)} - 2\frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{(V(Y))}}$$
$$= 2 - 2\rho$$
$$= 0$$

ここで，分散が 0 より，$\frac{X}{\sqrt{V(X)}} - \frac{Y}{\sqrt{V(Y)}} = c \quad (c \in \mathbb{R})$ と表せる．

さらに，$a = \frac{\sqrt{V(Y)}}{\sqrt{V(X)}}, b = -c\sqrt{V(Y)}$ とすると，X と Y の関係式 $Y = aX + b$ を得る．

□

共分散と相関係数

$a \neq 0$ のとき，$|\rho| = 1 \iff Y = aX + b$ を示す．

Proof:$|\rho| = 1 \implies Y = aX + b$(続き).

同様に，$\rho = -1$ のとき，

$$
\begin{aligned}
V\left(\frac{X}{\sqrt{V(X)}} + \frac{Y}{\sqrt{V(Y)}}\right) &= \frac{V(X)}{V(X)} + \frac{V(Y)}{V(Y)} + 2\frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{(V(Y))}} \\
&= 2 + 2\rho \\
&= 0
\end{aligned}
$$

ここで，分散が 0 より，$\frac{X}{\sqrt{V(X)}} + \frac{Y}{\sqrt{V(Y)}} = c \ \ (c \in \mathbb{R})$ と表せる．

さらに，$a = -\frac{\sqrt{V(Y)}}{\sqrt{V(X)}}, b = c\sqrt{V(Y)}$ とすると，X と Y の関係式

$Y = aX + b$ を得る． □

共分散と相関係数

Proof:$|\rho| = 1 \impliedby Y = aX + b$.

$Y = aX + b$ のとき，

$$\begin{aligned} E(Y) &= aE(X) + b, \\ V(Y) &= a^2 V(X), \\ Cov(X, Y) &= E(XY) - E(X)E(Y) \\ &= E(aX^2 + bX) - E(X)E(aX + b) \\ &= a(E(X^2) - E(X)^2) = aV(X) \end{aligned}$$

が成り立つ．
よって，

$$\rho^2 = \frac{Cov^2(X, Y)}{V(X)V(Y)} = \frac{a^2 V^2(X)}{V(X) \cdot a^2 V(X)} = 1 \iff |\rho| = 1$$

$\square$

共分散と相関係数

共分散 $Cov(X,Y)$ は，次の式で計算されることが多い．

$$
\begin{aligned}
Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y \\
&= E[XY] - \mu_X \mu_Y \\
&= E[XY] - E[X]E[Y] \qquad (7.14)
\end{aligned}
$$

共分散と相関係数

なお，共分散 $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$ は同時確率分布で，

$$\begin{aligned} Cov(X,Y) &= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)f(x,y) (\text{離散型}) \\ Cov(X,Y) &= \iint_S (x-\mu_X)(y-\mu_Y)f(x,y)dxdy (\text{連続型}) \end{aligned} \tag{7.15}$$

と表されるが，式 (7.14) を用いて，

$$\begin{aligned} E[XY] &= \sum_x \sum_y xyf(x,y) (\text{離散型}) \\ E[XY] &= \iint_S xyf(x,y)dxdy (\text{連続型}) \end{aligned} \tag{7.16}$$

から計算することもできる．

条件付確率

条件付確率の定義によると，

$$
\begin{aligned}
P(X=x|Y=y) &= \frac{P(X=x, Y=y)}{P(Y=y)} \\
P(Y=y|X=x) &= \frac{P(X=x, Y=y)}{P(X=x)}
\end{aligned}
\tag{7.17}
$$

である．Y が y と与えられたときの X の**条件付確率密度関数** (conditional probability density function), X が x と与えられたときの Y の条件付確率密度関数を，式 (7.17) と同様に，それぞれ，

$$
\begin{aligned}
g(x|y) &= \frac{f(x,y)}{h(y)} \\
h(y|x) &= \frac{f(x,y)}{g(x)}
\end{aligned}
\tag{7.18}
$$

と定義する．($h(y) \neq 0, g(x) \neq 0$ を仮定する．)
ここで，$|$ の後が条件であり，前が変数である．

条件付確率

ここで，$g(x|y)$ の変数，すなわち x で和をとると，離散変数の場合，

$$\sum_x g(x|y) = \sum_x \frac{f(x,y)}{h(y)} = \frac{h(y)}{h(y)} = 1 \tag{7.19}$$

となり，$g(x|y)$ は確率分布の条件を満たしている．同様に，$h(y|x)$ や連続変数の場合もこれらは確率分布の条件を満たしている．

条件付確率

このように，条件付確率分布もひとつの確率分布であるから，その**条件付期待値** (conditional expectation), **条件付分散** (conditional variance) を離散変数のとき，次のように定義できる．

$$\begin{aligned} E(X|y) &= \sum_x x g(x|y) = \mu_{X|y} \\ E(Y|x) &= \sum_y y h(y|x) = \mu_{X|y} \end{aligned} \tag{7.20}$$

$$\begin{aligned} V(X|y) &= \sum_x (x - \mu_{X|y})^2 g(x|y) \\ V(Y|x) &= \sum_y (y - \mu_{Y|x})^2 h(y|x) \end{aligned} \tag{7.21}$$

連続変数のときは，$\sum$ を $\int$ に替えれば良い．

独立な確率変数

同時確率分布において，あらゆる x, y について，

$$f(x,y) = g(x)h(y) \tag{7.22}$$

が成り立つとき，確率変数 X, Y は互いに**独立** (independent) であるという．このとき，X, Y の同時確率分布 $f(x,y)$ は X, Y それぞれの確率分布 (周辺確率分布)$g(x), h(y)$ から求められる．
X, Y が独立であるとき，

$$g(x|y) = \frac{f(x,y)}{h(y)} = \frac{g(x)h(y)}{h(y)} = g(x), \quad h(y|x) = h(y) \tag{7.23}$$

となり，X の出方は y の値に依らず，Y の出方は x の値に依らないことが分かる．これが「独立」の分かりやすい意味である．また，式 (7.23) を式 (7.22) を代入して，

$$f(x,y) = g(x)h(y|x) = g(x|y)h(y) \tag{7.24}$$

が成立する．

独立な確率変数

n 個の確率変数 $X_1, X_2, \cdots, X_n$ に対しても，その同時確率分布 f が, 周辺確率分布 $f_1, f_2, \cdots, f_n$ の積に

$$f(x_1, x_2, \cdots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n) \tag{7.25}$$

のように分解されるなら，$X_1, X_2, \cdots, X_n$ は**独立**であるという．
独立とは関連がないことであるが，同じ「関連がない」といっても，**無相関**よりはずっと強い意味を持つ．なぜならば，無相関は平均的な性質であって確率分布から，式 (7.12) により決まる量 ρ によるが，独立性は基礎の確率分布そのものに関する仮定だからである．

積の期待値

X, Y が独立ならば，

$$E(XY) = E(X)E(Y) \tag{7.26}$$

が成立する．

Proof.

(離散型の場合)

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyf(x,y) = \sum_x \sum_y xyg(x)h(y) \\ &= \left(\sum_x xg(x)\right)\left(\sum_y yh(y)\right) = \mu_X \mu_Y \\ &\left(\because \sum_x xg(x) = \sum_x x \sum_y f(x,y) = \sum_x \sum_y xf(x,y) = E(X)\right) \end{aligned}$$

連続型の場合は $\sum$ と $\int$ を入れ替えれば良い． □

独立と無相関

X, Y が独立ならば，無相関となる．

Proof.

式 (7.14), 式 (7.26) より，

$$Cov(X,Y) = E(XY) - E(X)(Y) = 0 \tag{7.27}$$

よって，式 (7.12) から，$\rho = 0$ となる． □

ここで，独立ならば必ず無相関であるが，その逆は成立しない．(無相関であっても独立とは限らない．)
e.g)$f(x,y) = \frac{1}{4}, (x,y) \in \{(-1,0),(0,1),(0,-1),(1,0)\}$ のとき，
$f(x=1,y=0) = \frac{1}{4}$ である．一方，$g(x=1) \times h(y=0) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$ となり，$f(1,0) \neq g(1)h(0)$. すなわち，式 (7.22) を満たさない．

独立な確率変数の和とモーメント母関数

X, Y が独立ならば，

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ \phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \end{aligned} \tag{7.28}$$

となる．

Proof.

X, Y が独立ならば，e^X, e^Y も独立であるので，

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$$

特性関数についても同様． □

逆は不成立であることに注意したい．

多次元正規分布

複雑で多様な現象を統計学で扱うときは，多くの変数を考えなければならない．
今までに考えたように独立ではなく，むしろ関連しあう多数の確率変数を最初から仮定して用いられる確率分布が，**多 (n) 次元正規分布**である．
まずは，2 次元正規分布を導く．

2 次元正規分布の構成

標準正規分布 $N(0,1)$ に従う，独立な二つの確率変数 X_1, X_2 があるとき，a, b, c, d を定数 (ただし，$ad - bc \neq 0$) として，確率変数の変換

$$Y_1 = aX_1 + bX_2,\ \ Y_2 = cX_1 + dX_2 \tag{7.29}$$

を行ったとき，

(a) Y_1, Y_2 の期待値 μ_1, μ_2

(b) Y_1, Y_2 の分散 σ_1^2, σ_2^2，共分散 σ_{12}, 相関係数 ρ

(c) Y_1, Y_2 の同時確率密度関数 $g(y_1, y_2)$

を求めてみる．

2 次元正規分布の構成
(a)Y_1, Y_2 の期待値 μ_1, μ_2

$X_1, X_2 \sim N(0,1)$ より，$E(X_1) = E(X_2) = 0.$
期待値の線形性から，

$$\mu_1 = E(Y_1) = E(aX_1 + bX_2) = aE(X_1) + bE(X_2) = 0.$$

同様にして，

$$\mu_2 = E(Y_2) = 0.$$

2 次元正規分布の構成

(b)Y_1, Y_2 の分散 σ_1^2, σ_2^2，共分散 σ_{12}, 相関係数 ρ

$X_1, X_2 \sim N(0,1)$ より，$V(X_1) = V(X_2) = 1$.
X_1 と X_2 は独立より，分散の加法性から，

$$\sigma_1^2 = V(aX_1 + bX_2) = a^2V(X_1) + b^2V(X_2) = a^2 + b^2.$$

同様にして，

$$\sigma_2^2 = c^2 + d^2.$$

さらに，独立性から，$E(X_1X_2) = E(X_1)E(X_2)$ なので，

$$\begin{aligned}\sigma_{12} &= E[(aX_1 + bX_2)(cX_1 + dX_2)] = E[acX_1^2 + bdX_2^2] \\ &= acE[X_1^2] + bdE[X_2^2] = ac + bd\end{aligned}$$

よって，

$$\rho = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \frac{ac + bd}{\sqrt{a^2 + b^2}\sqrt{c^2 + d^2}} \tag{7.30}$$

2 次元正規分布の構成
(c)Y_1, Y_2 の同時確率密度関数 $g(y_1, y_2)$

X_1, X_2 は独立なので，それらの同時確率分布は密度関数の積となって，$X_1, X_2 \sim N(0, 1)$ より，

$$f(x_1, x_2) = \frac{1}{2\pi} \exp\left\{-\frac{x_1^2 + x_2^2}{2}\right\} \tag{7.31}$$

である．

2 次元正規分布の構成

(c)Y_1, Y_2 の同時確率密度関数 $g(y_1, y_2)$

ここで，(X_1, X_2) と (Y_1, Y_2) が $1:1$ に対応しているとき，(X_1, X_2) から (Y_1, Y_2) の変換

$$Y_1 = \phi_1(X_1, X_2),\ \ Y_2 = \phi_2(X_1, X_2)$$

は逆に解け，それを，

$$X_1 = \psi_1(Y_1, Y_2),\ \ X_2 = \psi_2(Y_1, Y_2)$$

と置く．このとき，$\psi_1(y_1, y_2)$ と $\psi_2(y_1, y_2)$ が微分できれば，次の定理が成立する．

$(\psi_1(y_1, y_2),\psi_2(y_1, y_2))$ のヤコビアンを行列式

$$J = \begin{vmatrix} \frac{\partial}{\partial y_1}\psi_1(y_1, y_2) & \frac{\partial}{\partial y_2}\psi_1(y_1, y_2) \\ \frac{\partial}{\partial y_1}\psi_2(y_1, y_2) & \frac{\partial}{\partial y_2}\psi_2(y_1, y_2) \end{vmatrix}$$

と置くと，(Y_1, Y_2) の同時確率密度関数は，次式で与えられる．

$$g(y_1, y_2) = f(\psi_1(y_1, y_2), \psi_2(y_1, y_2))|J| \tag{7.45}$$

2 次元正規分布の構成

(c)Y_1, Y_2 の同時確率密度関数 $g(y_1, y_2)$

この定理を利用すると，Y_1, Y_2 の同時確率密度関数 $g(y_1, y_2)$ は，

$$
\begin{aligned}
g(y_1, y_2) &= f\left(\frac{dy_1 - by_2}{ad - bc}, \frac{ay_2 - cy_1}{ad - bc}\right)\begin{vmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{vmatrix} \\
&= f\left(\frac{dy_1 - by_2}{ad - bc}, \frac{ay_2 - cy_1}{ad - bc}\right)\frac{1}{ad - bc}
\end{aligned}
$$

となる．ここで，(b) の結果より，

$$
\begin{aligned}
&(ad - bc)^2 = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 = \sigma_1^2\sigma_2^2(1 - \rho^2) \\
&(dy_1 - by_2)^2 + (ay_2 - cy_1)^2 = \sigma_2^2 y_1^2 - 2\sigma_{12} y_1 y_2 + \sigma_1^2 y2^2
\end{aligned}
$$

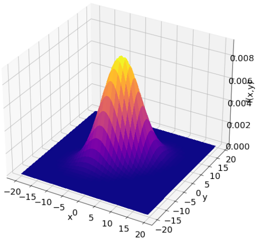
なので，式 (7.31) から，(Y_1, Y_2) の同時確率密度関数は，次式で与えられる．

$$
g(y_1, y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}}\exp\left\{-\frac{1}{2(1 - \rho^2)}\left(\frac{y_1^2}{\sigma_1^2} - \frac{2\rho y_1 y_2}{\sigma_1\sigma_2} + \frac{y_2^2}{\sigma_2^2}\right)\right\} \tag{7.34}
$$

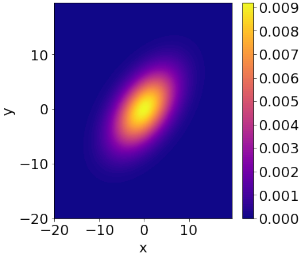
2 次元正規分布の構成
(c)Y_1, Y_2 の同時確率密度関数 $g(y_1, y_2)$

式 (7.34) を平均 (期待値) をそれぞれ 0,0, 分散をそれぞれ σ_1^2, σ_2^2，共分散を σ_{12} とする **2 次元 (2 変量) 正規分布** (bivariate normal distribution) といい，$N((0,0),(\sigma_1^2,\sigma_2^2,\sigma_{12}))$ と表す．なお，一般に平均 (期待値) が μ_1, μ_2 であるときは，y_1 が $y_1 - \mu_1$ に，y_2 が $y_2 - \mu_2$ になる．これを平均 (期待値)μ_1, μ_2, 分散 σ_1^2, σ_2^2，共分散 σ_{12} の 2 次元正規分布 (bivariate normal distribution) といい，$N((\mu_1,\mu_2),(\sigma_1^2,\sigma_2^2,\sigma_{12}))$ で表す．なお，$g(y_1, y_2)$ の指数部分は楕円の式であることに注意したい．

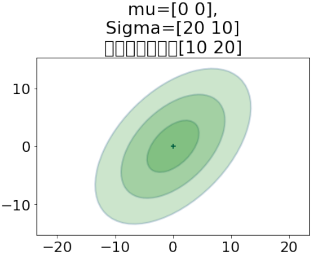
2 次元正規分布の構成
(c)Y_1, Y_2 の同時確率密度関数 $g(y_1, y_2)$



(a) 2 次元正規分布



(b) 上から俯瞰した図



(c) 等高線図

この楕円は，同時確率密度関数の等高線を表しており，$g(y_1, y_2) = k$ となる図形を表している．この楕円は，相関係数 ρ が大きくなるにつれて，直線

$$y_2 = \mathrm{sgn}(\rho)\frac{\sigma_2}{\sigma_1}(y_1 - \mu_1) + \mu_2$$

へと押しつぶされる．(Appendix 1.)

2 次元正規分布の構成

(c)Y_1, Y_2 の同時確率密度関数 $g(y_1, y_2)$

一般の**多次元 (多変量) 正規分布** (multivariate normal distribution) について，$\boldsymbol{X} = (X_1, X_2, \cdots, X_n)^\top$ が多変量正規分布に従うとすると，その密度関数は，

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

で与えられる．ここで，$\boldsymbol{\mu}$ は期待値ベクトル (expectation vector)，対称行列 Σ は分散共分散行列 (variance covariance matrix) と呼ばれ，

$$\boldsymbol{\mu} = (E(X_1), E(X_2), \cdots, E(X_n))^\top,$$

$$\Sigma = \begin{pmatrix} V(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & V(X_2) & \cdots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \cdots & V(X_n) \end{pmatrix}$$

と定義される．

分散の加法性

2 個の確率変数 X, Y に対して，期待値については常に加法性

$$E(X+Y) = E(X) + E(Y) \tag{7.35}$$

が成り立つ．さらに X, Y が独立であるときには，**分散の加法性**

$$V(X \pm Y) = V(X) + V(Y) \tag{7.36}$$

が成立する．独立でないときは，一般に，

$$\begin{aligned} V(X+Y) &= V(X) + V(Y) + 2Cov(X,Y) \\ &= V(X) + V(Y) + 2\rho_{XY} D(X) D(Y) \end{aligned} \tag{7.37}$$

となる．したがって，分散の加法性には無相関の条件があれば十分である．

分散の加法性

一方で，期待値の加法性には独立性は不要である．

Proof.

$$
\begin{aligned}
E(X+Y) &= \sum_x \sum_y (x+y) f(x,y) \\
&= \sum_x \left\{ x \sum_y f(x,y) \right\} + \sum_y \left\{ y \sum_x f(x,y) \right\} \\
&= \sum_x x g(x) + \sum_y y h(y) \\
&= E(X) + E(Y)
\end{aligned}
$$

□

n 個の場合

n 個の確率変数 $X_1, X_2, \cdots, X_n$ に対しても同じく独立性に関わらず，

$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) \tag{7.38}$$

であるが，$X_1, X_2, \cdots, X_n$ が独立のときには，

$$V(X_1 + X_2 + \cdots + X_n) = V(X_1) + V(X_2) + \cdots + V(X_n) \tag{7.39}$$

が成り立つ．

同一分布

$X_1, X_2, \cdots, X_n$ が同一の確率分布に従うとし，これらの期待値，分散を μ, σ^2 とすれば，$E(X_1) = E(X_2) = \cdots = E(X_n) = \mu,\ \ V(X_1) = V(X_2) = \cdots = V(X_n) = \sigma^2$ であるから，

$$E(X_1 + X_2 + \cdots + X_n) = n\mu,\ \ V(X_1 + X_2 + \cdots + X_n) = n\sigma^2 \quad (7.40)$$

が成立する．したがって，標準偏差は，$D(X_1 + X_2 + \cdots + X_n) = \sqrt{n}\sigma$ となる．

標準偏差は $\boldsymbol{\sqrt{n}}$ **に比例**することは重要である．

相加平均

$X_1 + X_2 + \cdots + X_n$ を n で割った相加平均を

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{7.41}$$

と置くと，式 (7.40) より，

$$E(\bar{X}) = \mu, V(\bar{X}) = \sigma^2/n \tag{7.42}$$

を得る．
すなわち，相加平均 $\bar{X}$ について，期待値は n に無関係につねに μ に一致するが，分散は n に反比例して減少し 0 に収束する．この安定化の傾向を定理の形に発展させたのが後に述べる**大数の法則**である．

和の確率分布

離散型の確率変数 X, Y が独立であるとし，その確率分布を $g(x), h(y)$ としよう．和の $X+Y$ の確率分布 $k(z)$ は確率 $P(X+Y=z)$ を考えれば得られる．$X+Y=z$ となるのは，$X=x, Y=z-x$ の形である $x, z-x$ の和が z になる全ての組み合わせであるから，それを確率の積で表して，

$$k(z) = \sum_x g(x)h(z-x) \tag{7.43}$$

となる．関数 g, h から k を作る数学操作を g, h の**たたみこみ** (convolution) といい，$k = g * h$ と書く．g, h が密度関数のときも同様で，たたみこみは積分

$$k(z) = \int_{-\infty}^{\infty} g(x)h(z-x)dx \tag{7.44}$$

となる．

和の確率分布

このたたみこみを用いて，X, Y が独立のとき，いくつかのよく知られた確率分布に対しては，次の結果が得られる．(再生性)

(a) 二項分布
$X \sim Bi(n,p), Y \sim Bi(m,p) \implies X+Y \sim Bi(n+m,p)$

(b) ポアソン分布
$X \sim Po(\lambda), Y \sim Po(\mu) \implies X+Y \sim Po(\lambda+\mu)$

(c) 正規分布
$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2) \implies X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

(d) ガンマ分布 (+指数分布)
$X \sim \Gamma(\alpha_1, \lambda), Y \sim \Gamma(\alpha_2, \lambda) \implies X+Y \sim \Gamma(\alpha_1+\alpha_2, \lambda)$

(e) 負の二項分布 (+幾何分布)
$X \sim NB(k_1, p), Y \sim NB(k_2, p) \implies X+Y \sim NB(k_1+k_2, p)$

(f) カイ二乗分布
$X \sim \chi^2_{k_1}, Y \sim \chi^2_{k_2} \implies X+Y \sim \chi^2_{k_1+k_2}$

これらの証明は，特性関数を用いることが容易である．

和の確率分布

Proof.

(a) 二項分布

$$\begin{aligned}\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \\ &= (pe^{it}+q)^n(pe^{it}+q)^m \\ &= (pe^{it}+q)^{n+m}\end{aligned}$$

(b) ポアソン分布

$$\begin{aligned}\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \\ &= \exp\{\lambda(e^{it}-1)\}\exp\{\mu(e^{it}-1)\} \\ &= \exp\{(\lambda+\mu)(e^{it}-1)\}\end{aligned}$$

□

和の確率分布

Proof.

(c) 正規分布

$$\begin{aligned}\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \\ &= \exp\left(i\mu_1 t - \frac{\sigma_1^2 t^2}{2}\right)\exp\left(i\mu_2 t - \frac{\sigma_2^2 t^2}{2}\right) \\ &= \exp\left(i(\mu_1+\mu_2)t - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}\right)\end{aligned}$$

(d) ガンマ分布 (+指数分布)

$$\begin{aligned}\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \\ &= (1-i\beta t)^{-\alpha_1}(1-i\beta t)^{-\alpha_2} \\ &= (1-i\beta t)^{-(\alpha_1+\alpha_2)}\end{aligned}$$

□

和の確率分布

Proof.

(e) 負の二項分布 (+幾何分布)

$$\begin{aligned}\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \\ &= \left(\frac{1-p}{1-pe^{it}}\right)^{k_1}\left(\frac{1-p}{1-pe^{it}}\right)^{k_2} \\ &= \left(\frac{1-p}{1-pe^{it}}\right)^{k_1+k_2}\end{aligned}$$

(f) カイ二乗分布

$$\begin{aligned}\phi_{X+Y}(t) &= \phi_X(t)\phi_Y(t) \\ &= \left(\frac{1}{1-2it}\right)^{\frac{k_1}{2}}\left(\frac{1}{1-2it}\right)^{\frac{k_2}{2}} = \left(\frac{1}{1-2it}\right)^{\frac{k_1+k_2}{2}}\end{aligned}$$

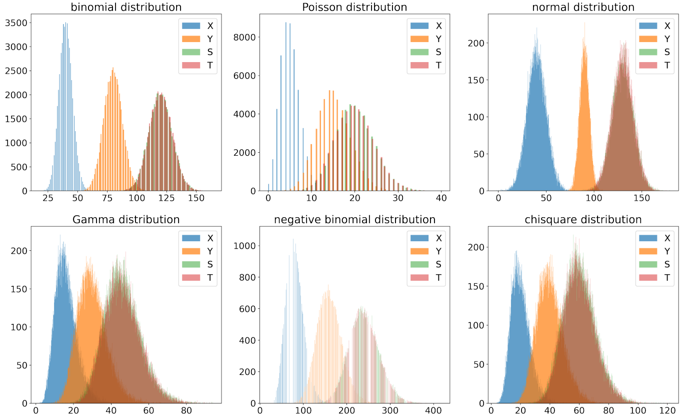
□

和の確率分布

このように，たたみこみの結果として，全く別の分布でなくふたたび同一種類の確率分布 (確率分布族) が得られるならば，取り扱いが便利である．このとき，この確率分布族は**再生的** (reproductive) であるという．

和の確率分布

再生性の確認．X(青) と Y(オレンジ) は各確率分布に従う乱数．S(緑) はそれらの和．T(赤) は再生性が成り立つときに X+Y が従うはずの確率分布に従う乱数である ($E(T) = E(X) + E(Y)$ かつ X, Y と同じ確率分布に従う乱数)．(Appendix 2.)



和の確率分布

指数分布 $Ex(\lambda)$ について，この特性関数は，

$$\phi_X(t) = \frac{\lambda}{\lambda - it}$$

より，$X \sim Ex(\lambda_1), Y \sim Ex(\lambda_2)$ とすると，

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \frac{\lambda_1}{\lambda_1 - it}\frac{\lambda_2}{\lambda_2 - it} = \frac{1}{1 - i(\lambda_1 + \lambda_2)t - \lambda_1\lambda_2 t^2}$$

となり，指数分布の特性関数にならないので，指数分布の再生性は成り立たない．

和の確率分布

また，幾何分布 $Geo(p)$ について，この特性関数は，

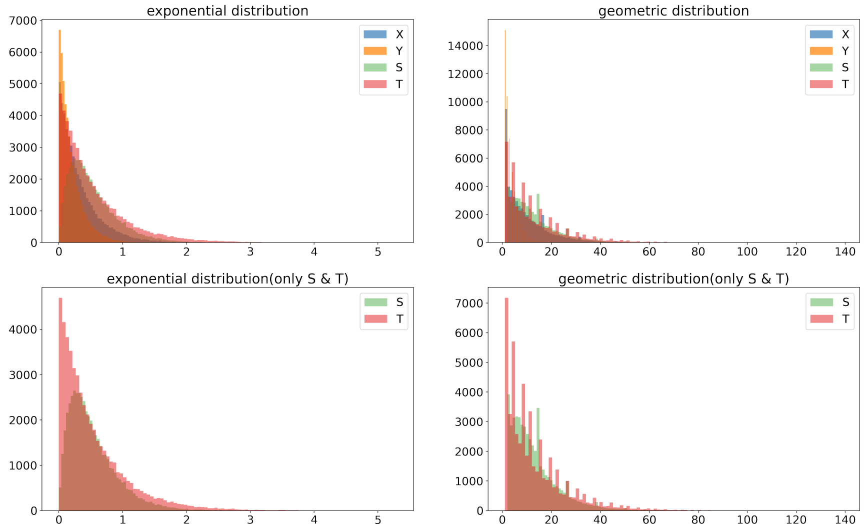
$$\phi_X(t) = \frac{1-p}{1-pe^{it}}$$

より，$X \sim Geo(p_1), Y \sim Geo(p_2)$ とすると，

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t) = \frac{1-p_1}{1-p_1e^{it}}\frac{1-p_2}{1-p_2e^{it}} = \frac{1-(p_1+p_2)+p_1p_2}{1-(p_1+p_2)e^{it}+p_1p_2e^{2it}}$$

となって，幾何分布の特性関数にならないので，幾何分布の再生性は成り立たない．

和の確率分布


exponential distribution
geometric distribution
exponential distribution(only S & T)
geometric distribution(only S & T)
X
Y
S
T

正規分布の再生性

正規分布については，再生性により，以下が成立する．

(a) $X_1, X_2, \cdots, X_n$ が独立で，それぞれ正規分布 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2), \cdots, N(\mu_n, \sigma_n^2)$ に従っているならば，
  - i) $X_1 + X_2 + \cdots + X_n$ は，$N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$ に従い，
  - ii) $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ は，$N(c_1\mu_1 + c_2\mu_2 + \cdots + c_n\mu_n, c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_n^2\sigma_n^2)$ に従う．

(b) とくに $X_1, X_2, \cdots, X_n$ の確率分布がすべて正規分布 $N(\mu, \sigma^2)$ なら，
  - i) $X_1 + X_2 + \cdots + X_n$ は，$N(n\mu, n\sigma^2)$ に従い，
  - ii) $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$ は，$N(\mu, \sigma^2/n)$ に従う．

目次

大数の法則

$X_1, X_2, \cdots, X_n$ が互いに独立で，同じ分布に従う確率変数とする．その確率分布の平均を $E(X) = \mu$ とすると，任意の $\epsilon > 0$ に対し，

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

が成立する．
このとき，確率論では $\bar{X}_n$ は μ に**確率収束** (converge in probability) するという．
大数の法則 (law of large numbers) は，一般的に，**大標本では，観察された標本平均を母集団の真の平均 (母集団) とみなしてよい**という常識を，数学的に厳密に証明したものに他ならない．

大数の法則

Proof.

チェビシェフの不等式

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \qquad (k>0, \mu=E(X), \sigma^2=V(X))$$

から，$\sigma_n^2 = V(\bar{X}_n)$ として，確率変数 X を $\bar{X}$ とすると，

$$P(|\bar{X}_n-\mu| > k\sigma_n) \leq \frac{1}{k^2}$$

が成立する．ここで，$\sigma^2 = V(X)$ とすると，$\sigma_n^2 = \sigma^2/n$ であるので，
$k\sigma_n = \epsilon$ とおくと，$k = \epsilon/\sigma_n = \sqrt{n}\epsilon/\sigma$ であるから，

$$P(|\bar{X}_n-\mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \qquad (n \to \infty)$$

が成立する． □

中心極限定理

平均 μ, 分散 σ^2 の分布からの無作為標本 $X_1, X_2, \cdots, X_n$ に対し，

$$
\begin{aligned}
&\lim_{n\to\infty} P\left(a \leq \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \leq b\right) \\
&= \lim_{n\to\infty} P\left(a \leq \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq b\right) \\
&= \lim_{n\to\infty} P\left(a \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq b\right) \\
&= \Phi(b) - \Phi(a) \qquad (8.1)
\end{aligned}
$$

が成り立つ．これを**中心極限定理** (central limit theorem) という．ここで，$\Phi(x)$ は標準正規分布の (累積) 分布関数であり，巻末の数値表で値が与えられている．

中心極限定理は，大数の法則よりくわしい定理であり，ごく大まかにいえば，母集団分布が何であっても，**和 $X_1 + \cdots + X_n$ の確率分布は,n が大なるときには，大略正規分布と考えてよい**ということである．

中心極限定理

Proof.

標準化変数

$$\frac{X_1+X_2+\cdots+X_n-n\mu}{\sqrt{n}\sigma}=\frac{Y_1+Y_2+\cdots+Y_n}{\sqrt{n}}$$

の確率分布のモーメント母関数が，n が大きいとき正規分布のモーメント母関数に近づくことを示す．ここで，

$$Y_1=\frac{X_1-\mu}{\sigma},\ Y_2=\frac{X_2-\mu}{\sigma},\ \cdots,Y_n=\frac{X_n-\mu}{\sigma}$$

は，各々 $X_1,X_2,\cdots,X_n$ の標準化変数であって，

$$E(Y_1)=E(Y_2)=\cdots=E(Y_n)=0, V(Y_1)=V(Y_2)=\cdots=V(Y_n)=1$$

となっている． □

中心極限定理

Proof(続き).

$X_1, X_2, \cdots, X_n$ の確立分布はみな同一，よって $Y_1, Y_2, \cdots, Y_n$ についてもそうであるから，その一つを Y とおいてモーメント母関数を作る．1 次のモーメント (期待値) が $E(Y) = 0$,2 次のモーメントが $E(Y^2) = V(Y) = 1$ であるので，

$$M_Y(t) = 1 + \frac{t^2}{2} + \cdots$$

となり，$Y_1 + Y_2 + \cdots + Y_n \equiv T$ のモーメント母関数は $Y_1, Y_2, \cdots, Y_n$ のそれらの積であるから，

$$M_T(t) = \{M_Y(t)\}^n = \left(1 + \frac{t^2}{2} + \cdots\right)^n$$

となる． □

中心極限定理

Proof(続き).

最終的に $(Y_1 + Y_2 + \cdots + Y_n)/\sqrt{n}$ のモーメント母関数は，t を $t/\sqrt{n}$ でおきかえて，

$$M_T\left(\frac{t}{\sqrt{n}}\right) = \left\{M_Y\left(\frac{t}{\sqrt{n}}\right)\right\}^n = \left(1 + \frac{t^2}{2n} + \cdots\right)^n \to \exp\left(\frac{t^2}{2}\right)$$

ここで，$\exp(t^2/2)$ は標準正規分布のモーメント母関数であるため，中心極限定理は証明された． □

中心極限定理

また，母集団分布の平均，分散 (母平均，母分散) を μ, σ^2 とすると，母集団が何であっても，標本の大きさ n が大なるときは，大略

$$
\begin{aligned}
S_n = X_1 + X_2 + \cdots + X_n \qquad & \text{は} \qquad N(n\mu, n\sigma^2) \qquad \text{に，} \\
\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \qquad & \text{は} \qquad N(\mu, \frac{\sigma^2}{n}) \qquad \text{に}
\end{aligned}
\tag{8.6}
$$

従うと考えてよい．

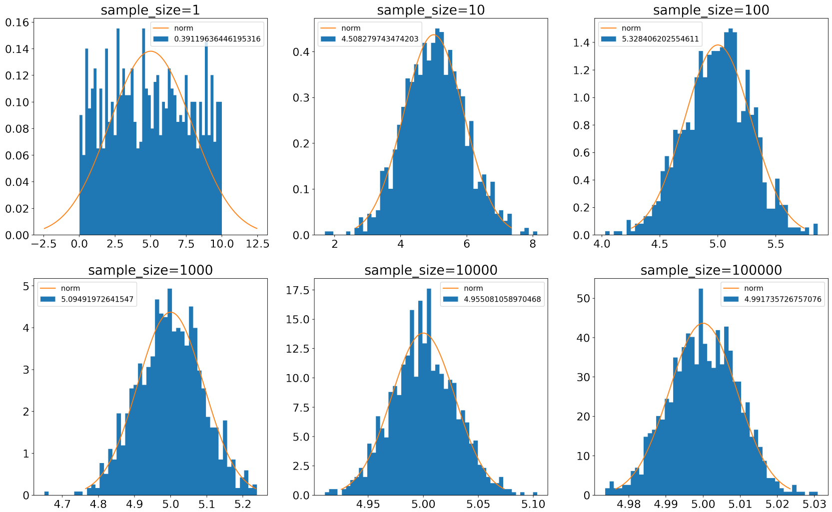
とくに，$\bar{X}$ については集中を保証する大数の法則よりくわしい．正規分布の形をとりながら集中する ($\sigma^2/n \to 0$) ことを示しているからである．(Appendix 3.,4.)

中心極限定理

平均 μ, 標準偏差 σ の母集団に対して，十分大きな n 個の標本の平均値 $\bar{X}$ は平均 μ, 標準偏差 $\sigma/\sqrt{n}$ の正規分布に近似できることを確認する．
方法としては，各分布に従う乱数を $n \in \{1, 10, 100, 1000, 10000, 100000\}$ 個生成して，各 $n(sample\_size)$ についてそれらの平均を求めることを 1000 回行い，その結果をヒストグラムとして表した．ここで，ヒストグラムの帯の面積の和は 1 となるように設定しており，オレンジの線は各分布の期待値，分散をパラメータとする理想的な正規分布である．

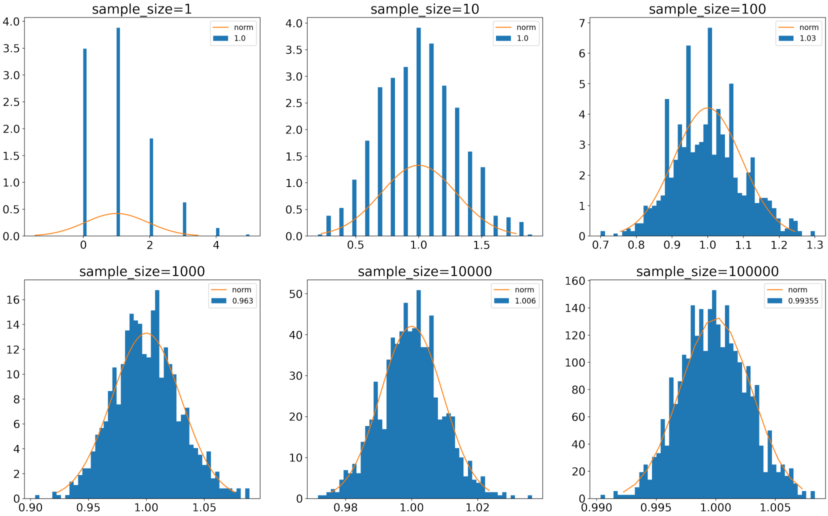
中心極限定理

一様分布


sample_size=1
norm
0.39119636446195316
sample_size=10
norm
4.508279743474203
sample_size=100
norm
5.328406202554611
sample_size=1000
norm
5.09491972641547
sample_size=10000
norm
4.955081058970468
sample_size=100000
norm
4.991735726757076

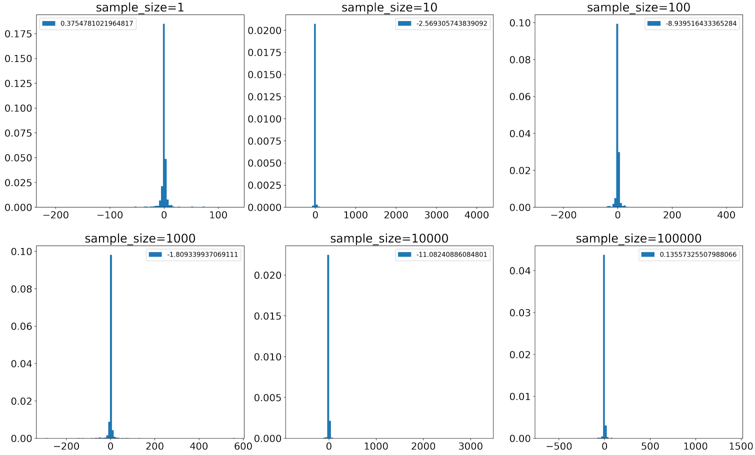
中心極限定理

二項分布


sample_size=1
norm
1.0
sample_size=10
norm
1.0
sample_size=100
norm
1.03
sample_size=1000
norm
0.963
sample_size=10000
norm
1.006
sample_size=100000
norm
0.99355

中心極限定理

コーシー分布



コーシー分布は期待値・分散が存在しないため，中心極限定理は成立しない．

中心極限定理

また，平均値の標準偏差は $\sqrt{n}$ に反比例することを確認した．方法としては，各分布に従う乱数を $n \in \{1, 10, 100, 1000, 10000, 100000\}$ 個生成して，その平均を求めることを 1000 回行い，n とそれらの標準偏差と両対数グラフにプロットした．
ここで，両対数グラフの傾きが -0.5 程度であれば，n 個の平均の標準偏差を σ_n,1 個の平均 (すなわち，ある分布に従う乱数の標準偏差) を σ としたとき，

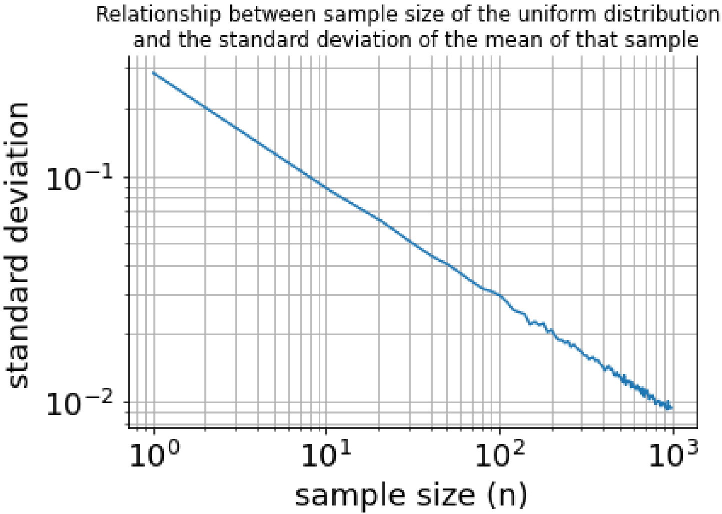
$$\log \sigma_n = -0.5 \log n + log\sigma$$

$$\iff \sigma_n = n^{-0.5}\sigma = \frac{\sigma}{\sqrt{n}}$$

となり，平均値の標準偏差は $\sqrt{n}$ に反比例するといえる．

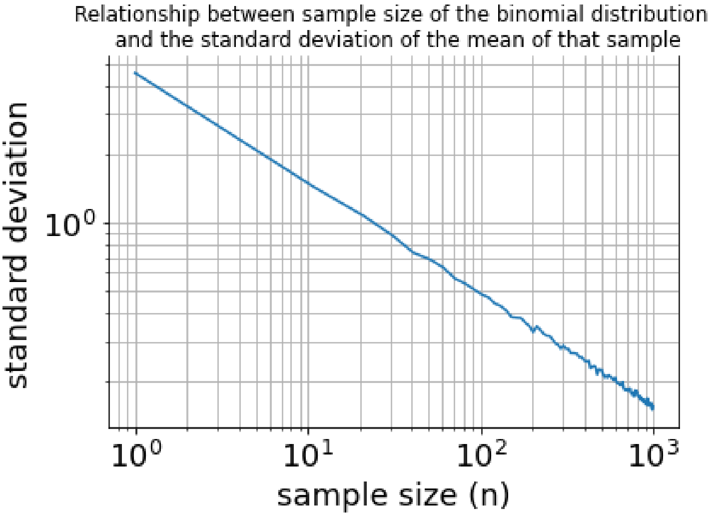
中心極限定理

一様分布


Relationship between sample size of the uniform distribution
and the standard deviation of the mean of that sample
standard deviation
10^{-1}
10^{-2}
10^0
10^1
10^2
10^3
sample size (n)

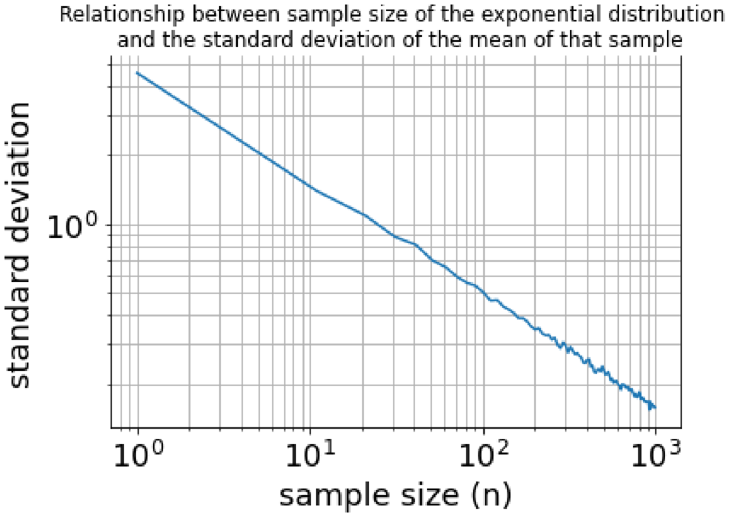
中心極限定理

二項分布


Relationship between sample size of the binomial distribution
and the standard deviation of the mean of that sample
standard deviation
10^0
10^0
10^1
10^2
10^3
sample size (n)

中心極限定理

指数分布


Relationship between sample size of the exponential distribution
and the standard deviation of the mean of that sample
standard deviation
10^0
10^0
10^1
10^2
10^3
sample size (n)

ある分布に従う乱数の発生

確率変数 X は密度関数 $f(x)$, 分布関数 $F(x)$ を持つとする．関数 $y = h(x)$ は連続な狭義単調関数であるとして，この逆関数を $x = h^{-1}(y) = t(y)$ とする．このとき，$Y = h(X)$ の密度関数は，次式で与えられる．c.f (7.45)

$$g(y) = f\left(t(y)\right)\left|t'(y)\right|$$

Proof.

($h(x)$ が狭義単調増加の場合)

$$G(y) = P(Y \leq y) = P(h(X) \leq y) = P(X \leq t(y)) = F(t(y))$$

この両辺を微分すれば，Y の密度関数 $g(y) = f\left(t(y)\right)t'(y)$ が得られる．
($h(x)$ が狭義単調減少の場合)

$$G(y) = P(Y \leq y) = P(h(X) \leq y) = P(X \geq t(y)) = 1 - F(t(y))$$

両辺を微分して，$g(y) = -f\left(t(y)\right)t'(y) = f\left(t(y)\right)\left|t'(y)\right|$ を得る． □

ある分布に従う乱数の発生

一般に，求める確率変数 (乱数) の累積分布関数を $G(x)$ とする．$G(x)$ は狭義単調増加，連続とする．
$y = G(x)$ の逆関数を $x = G^{-1}(y)$ とする．$G(x)$ は狭義単調増加かつ連続より，これは必ず存在する．
いま，$[0,1]$ 上の一様分布に従う確率変数を U とするとき，$X = G^{-1}(U)$ が求めるべき，累積分布関数 $G(x)$ を持つ確率変数である．

Proof.

$X = G^{-1}(U)$　としたとき，$P(X \leq x) = G(x)$ であることを示せばよい．
いま，不等式 $G^{-1}(U) \leq x$ と $U \leq G(x)$ が同値であるから，

$$P(X \leq x) = P(G^{-1}(U) \leq x) = P(U \leq G(x)) = G(x)$$

となって，$G^{-1}(U) = X$ は求める累積分布関数に従っている．
ここで，一様分布の累積分布関数について，$P(U \leq u) = u \ \ (0 \leq u \leq 1)$ であることを用いた．　□

ある分布に従う乱数の発生

母数 λ の指数分布 $Ex(\lambda)$ に対して，その累積分布関数は $G(x) = 1 - e^{-\lambda x}$ であるので，逆関数 $G^{-1}(y)$ を求めると，

$$G^{-1}(y) = -\frac{\log(1-y)}{\lambda}$$

となる．したがって，

$$G^{-1}(U) = -\frac{\log(1-U)}{\lambda}$$

が $Ex(\lambda)$ に従う．$1-U$ も $[0,1]$ 上の一様分布に従うから，$-\log U/\lambda$ でもよい．

二項分布の正規分布による近似

二項分布 $Bi(n, p)$ は，

$$f(x) = {}_n\mathrm{C}_x p^x (1-p)^{n-x}, \ x = 0, 1, \cdots, n \tag{8.9}$$

で表される．
したがってこの式を用いれば，いかなる n, p において $f(x)$ を求めることが原理的には可能である．しかし，実際には，二項係数 ${}_n\mathrm{C}_x$ の値が大きくなるためにそれほど簡単ではない．

二項分布の正規分布による近似

中心極限定理によると，n が大きいときに $Bi(n,p)$ は正規分布に近づくので，それを近似することができる．なぜなら，二項分布における成功の回数 S は，それぞれが二項分布 $Bi(1,p)$ に従う確率変数 $X_1, X_2, \cdots, X_n$ の和 $S = X_1 + X_2 + \cdots + X_n$ となるからである．
$E(S) = np, V(S) = np(1-p)$ より，中心極限定理 (8.6) により，$Bi(n,p)$ は $N\left(np, \sqrt{np(1-p)}\right)$ によって近似できるので，成功回数が k 回以上 k' 回以下である確率は，n が大きければ，

$$
\begin{aligned}
P(k \le S \le k') &= P\left(\frac{k-np}{\sqrt{np(1-p)}} \le Z \le \frac{k'-np}{\sqrt{np(1-p)}}\right) \\
&\fallingdotseq \Phi\left(\frac{k'-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right)
\end{aligned}
\tag{8.11}
$$

で求めることができる．

二項分布の正規分布による近似

さて，二項分布で n の値がどのくらい大きければ，正規分布による近似を用いてよいか．実用上充分な精度を得るために通常言われている必要条件は，$np > 5$ かつ $n(1-p) > 5$ である．また，p が小さく n が大きいとき $(np < 5)$ はポアソン分布 $Po(np)$ で近似できる．(Appendix 5.)

Proof.

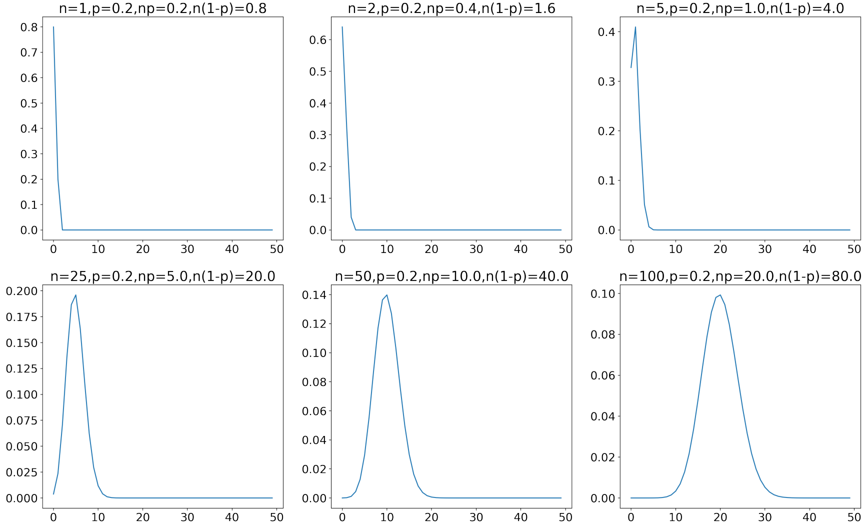
$np = \lambda(> 0)$ として，$n \to \infty, p \to 0$ となる極限では，二項分布 $B(n, p)$ に従う確率変数 X の特性関数を考えると，e の定義から，

$$\phi_X(t) = (pe^{it}+1-p)^n = \left(1 + \frac{\lambda(e^{it}-1)}{n}\right)^n \to \exp\left(\lambda(e^{it}-1)\right) \;\; as\; n \to \infty$$

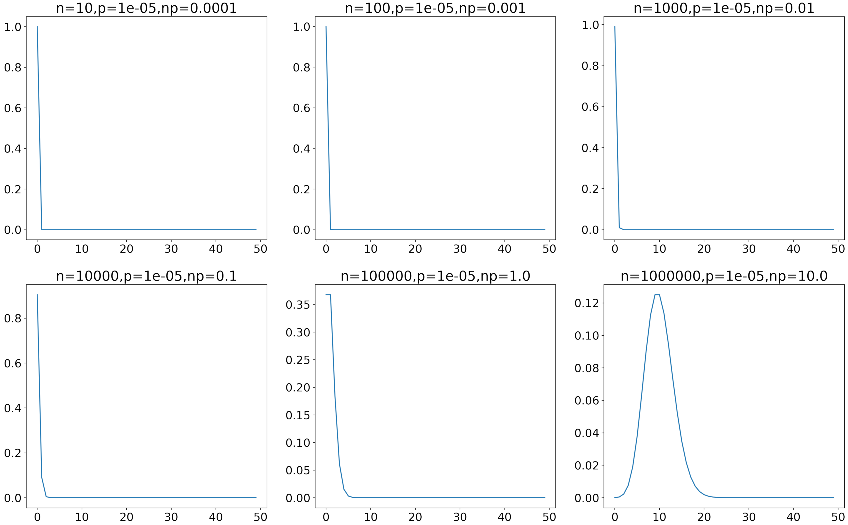
となるが，この特性関数は平均 $\lambda = np$ のポアソン分布の特性関数に他ならない．

(See.p114 ポアソンの小数の法則) □

二項分布の正規分布による近似



二項分布の正規分布による近似



正規乱数の発生

指定された平均，分散に従う乱数 (正規乱数) を区間 $(0,1)$ 上の一様乱数から作るにはどうすればよいだろうか．
正規分布の累積分布関数は単純な式ではないから逆関数を求めることができず，したがって指数乱数の場合のように逆変換法を用いることはできない．
しかし，中心極限定理によれば，区間 $(0,1)$ 上の一様乱数を n 個発生させ，n が十分に大きければその標本平均は一つの正規乱数とみなせる．
具体的な方法は次の通りである．

正規乱数の発生

まず $(0,1)$ 上の一様乱数を n 個発生させ，それらを $r_1, r_2, \cdots, r_n$ とする．区間 (a,b) 上の一様分布の期待値，分散はそれぞれ

$$E(r_i) = \frac{b+a}{2}, \quad V(r_i) = \frac{(b-a)^2}{12}$$

であるから，$U(0,1)$ の場合はそれぞれ $1/2, 1/12$ である．中心極限定理から，

$$z = \frac{\sum_{i=1}^{n} r_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \tag{8.13}$$

は n が大きいときほぼ標準正規分布 $N(0,1)$ に従う．予定の正規乱数 x の期待値を μ，分散を σ^2 とすると，x の標準化の逆関数 $x = \sigma z + \mu$ に代入して，

$$x = \sigma \sqrt{\frac{12}{n}} \left(\sum_{i=1}^{n} r_i - \frac{n}{2} \right) + \mu \tag{8.14}$$

となる．これがほぼ正規分布 $N(\mu, \sigma^2)$ に従う正規乱数である．