# 数理統計学

大木

2018 年 6 月 4 日

## 1　事象と確率

この章は条件付き確率が登場する 1.3 のみ紹介する。

### 1.3　事象の独立性と従属性

**定義.** 2 つの事象 $A, B$ が確率的に独立であるとは、

$$P(A \cap B) = P(A)P(B) \tag{1}$$

が成り立つことをいう。独立でない場合は従属という。

**定義.** 2 つの事象 $A, B$ があって、$P(A) > 0$ のとき、

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{2}$$

を $A$ が与えられた時の $B$ の条件付き確率という。

言葉で書くなら
($A, B$ が同時に起こる確率 ) = ($A$ が起こる確率)×($A$ が起きたうえでの $B$ が起こる確率)
である。$A, B$ が独立の時は $P(B|A) = P(B), P(A|B) = P(A)$ が成立する。

**定理.** (全確率の公式) 母集団空間 $\Omega$ が互いに素な事象 $\{A_i\}_1^n$ によって直和分割されているとすると、

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i) \tag{3}$$

と書ける。

**定理.** (ベイズの定理) 事後確率 $P(A_j|B)$、つまり $B$ が起こった後での $A_j$ が起こる条件付き確率は、事前確率 $P(A_j)$ を用いて

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{P(B)} \tag{4}$$

と書ける。

例題 1.2 を確認して、この章は終えます。

# 2　確率変数と確率分布

## 2.3　分布の特性値：平均と分散

(連続的な) 確率変数 $X$ の分布関数を $F(x)$ とすると、平均 $E[X]$ と分散 $V[X]$ は次のように定義される。

$$E[X] = \int_{-\infty}^{\infty} x dF(x) \tag{5}$$

$$V[X] = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 dF(x) \tag{6}$$

力学に例えるなら、確率密度関数 (pdf) を質量密度だと思えば、平均は重心で、分散は重心周りの慣性モーメントである。確率変数が離散的な場合は、積分を和に書き換えればよい。
続いて分散公式を紹介する。分散を実際計算するときには、

$$V[X] = E[X^2] - (E[X])^2 \tag{7}$$

という分散公式を用いることが多い。

### 標準化

数学のテストの点数が平均 60 点の中で 80 点、国語の点数が平均 40 点の中で 60 点だった時、どちらが全体の中で上位にいるかの目安は偏差値でわかる。偏差値は平均や分散がバラバラな標本を平均 50、分散 100 にそろえることで比較しやすくするための尺度になっている。同様に、確率変数を平均 0 分散 1 の分布に従うように変換することを、標準化という。具体的には、平均 $\mu$、分散 $\sigma$ の確率変数 $X$ を

$$Z = \frac{X-\mu}{\sigma} \tag{8}$$

と $Z$ に変数変換することをいう。

## 2.4　分布関数の変換

### 積率母関数と特性関数

積率母関数についても取り上げておく。確率変数 $X$ を用いて

$$M(t) = E[\mathrm{e}^{tX}] \int_{-\infty}^{\infty} \mathrm{e}^{tX} dF(X) \tag{9}$$

という関数を、積率母関数という。積率母関数は任意回微分可能でその階数が積率 (モーメント) の次数に対応している。つまり、

$$\begin{gathered}
\left.\frac{dM(t)}{dt}\right|_{t=0} = E[X] \\
\left.\frac{d^2M(t)}{dt^2}\right|_{t=0} = E[X^2] \\
\vdots \\
\left.\frac{d^kM(t)}{dt^k}\right|_{t=0} = E[X^k]
\end{gathered} \tag{10}$$

となる。また、

$$\phi(t) = E[\mathrm{e}^{itX}] \int_{-\infty}^{\infty} \mathrm{e}^{itX} dF(X) \tag{11}$$

を特性関数という。特性関数は常に存在するが、積率母関数は必ずしも存在するとは限らないことに注意。特性関数や積率母関数が求まれば分布関数はその逆変換によって求まるので、分布関数の持つ情報は、特性関数や積率母関数を調べることでわかる。
確率変数が非負の離散値をとるときは、積率母関数と同等のものとして確率母関数がある。

#### キュミュラント母関数

積率母関数の対数をとったものをキュミュラント母関数という。キュミュラントとは、キュミュラント母関数を級数展開した

$$\psi(t) = \log M(t) = \sum_{n=1}^{\infty} \frac{c_n}{n!} t^n \tag{12}$$

の係数 $c_n$ をいう。

### 演習問題 2.14

表の出る確率が $p$, 裏が出る確率が $1-p$ の銅貨を独立に $n$ 回投げるとき、表が偶数回出る確率を $a_n$ とする。この $a_n$ を求める。

## 3　確率分布の代表的なモデル

この章で紹介する分布は最低限にとどめる。

### 3.1　離散分布

#### ベルヌーイ分布・二項分布

表の出る確率が $p$、裏が出る確率が $1-p$ のコインを投げることを考える。表を 1、裏を 0 という離散確率変数に対応付けると、その pmf はベルヌーイ分布

$$f(\epsilon) = p^{\epsilon}(1-p)^{1-\epsilon} \tag{13}$$

に従う。ベルヌーイ分布の期待値は $p$、分散は $p(1-p)$ である。
$n$ 回の独立なベルヌーイ試行 $\{\epsilon\}_1^n$ を考える。このとき $\epsilon = 1$ となった試行の合計を $X(0 \le X \le n)$ とすると、離散確率変数 $X$ は二項分布に従う。pmf は

$$f(x) = \ {}_nC_x p^x (1-p)^{n-x} \tag{14}$$

で与えられる。確率母関数は二項定理を用いればすぐに求まり、

$$P(t) = (pt + 1 - p)^n \tag{15}$$

となる。また、二項分布の期待値は $np$、分散は $np(1-p)$ となる。

**ポアソン分布**

まれな現象の大量観測 (何度も確認できるほどの規模・時間を設けたうえでの観測) でその現象を確認した回数はポアソン分布に従う。例えば、放射性崩壊によって所定期間内で観測される放射線の数はポアソン分布に従うことはラザフォードの実験で知られている。ポアソン分布の pmf は

$$f(x) = \frac{\lambda^x}{x!}\mathrm{e}^{-\lambda} \tag{16}$$

で与えられる。母数 $\lambda$ を強度という。確率母関数はすぐに確認でき、$\mathrm{e}^{\lambda(t-1)}$ となる。また、ここから平均と分散はどちらも $\lambda$ となることが確認できる。一般に、ポアソン分布の $n$ 次のキュミュラントは強度 $\lambda$ に等しい。

## 3.2　連続分布

**正規分布**

正規分布 (ガウス分布) はよく知っていると思うので、積率母関数を求めるだけにとどめる。(p.39〜)
正規分布 $\mathcal{N}(\mu, \sigma^2)$ の積率母関数は

$$\begin{aligned} M(t) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x-\mu)^2 - 2\sigma^2 tx}{2\sigma^2}\right\}dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x-(\mu+\sigma^2 t))^2 - 2\mu\sigma^2 t - (\sigma^2 t)^2}{2\sigma^2}\right\}dx \\ &= \exp(\mu t + \frac{1}{2}\sigma^2 t^2) \end{aligned} \tag{17}$$

となる。ここから、ガウス分布の 3 次以上のキュミュラントは 0 であることがわかる。

# 4　2 次元確率ベクトルの分布

## 4.1　2 つの確率変数の同時分布

この章では 2 つの確率変数を同時に考え、その分布を通してそれらの関係について調べる。具体的には、2 つの確率変数を $X, Y$ として、同時分布 $P(X, Y)$ を考える。また、確率変数 $X, Y$ それぞれについて $P(X), P(Y)$ を周辺分布という。

**条件付き期待値・条件付き分散**

$Y = y$ が与えられたときの確率変数 $X$ についての条件付き期待値・条件付き分散をそれぞれ

$$E[X|Y=y] = \int xf(x|Y=y)dx \tag{18}$$

$$V[X|Y=y] = \int (x - E[X|Y=y])^2 f(x|Y=y)dx \tag{19}$$

と書く。

### 4.2　共分散と相関係数

$(X,Y)$ の関数 $h(X,Y)$ の平均は

$$E[h(X,Y)] = \int\int h(x,y)f(x,y)dxdy \tag{20}$$

ここで $h(X,Y) = \mathrm{e}^{sX+tY}$ とすると、同時分布の同時積率母関数は

$$M(s,t) = \int\int \mathrm{e}^{sx+ty}f(x,y)dxdy \tag{21}$$

と書ける。確率変数 $X,Y$ が独立の時、$E[h_1(X)h_2(Y)] = E[h_1(X)]E[h_2(Y)]$ となることから

$$M(s,t) = E[\mathrm{e}^{sX+tY}] = E[\mathrm{e}^{sX}]E[\mathrm{e}^{tY}] = M_1(s)M_2(t) \tag{22}$$

と同時積率母関数は周辺積率母関数の積で表せる。$X,Y$ が独立であることと、同時積率母関数が周辺積率母関数の積で書けることは同値である。

$X,Y$ の平均・分散

$$E[X] = \mu_1, V[X] = \sigma_1^2$$
$$E[Y] = \mu_2, V[Y] = \sigma_2^2$$

に加えて、平均周りの相互モーメントを共分散といい

$$Cov[X,Y] = E[(X-\mu_1)(Y-\mu_2)] = \sigma_{12} \tag{23}$$

と表す。このとき、共分散公式

$$Cov[X,Y] = E[XY] - E[X]E[Y] \tag{24}$$

が成り立つ。

$X,Y$ の標準化 $Z_1 = \frac{X-\mu_1}{\sigma_1}, Z_2 = \frac{Y-\mu_2}{\sigma_2}$ の積の平均を相関係数といい、

$$Corr[X,Y] = \frac{Cov[X,Y]}{V[X]V[Y]} = \frac{\sigma_{12}}{\sigma_1\sigma_2} = \rho \tag{25}$$

で表す。$\rho > 0$ の時は正の相関、$\rho = 0$ の時は無相関、$\rho < 0$ の時は負の相関という。相関係数には以下のような性質がある。

(1) 相関係数の絶対値は 1 以下である。

(2) $\rho = \pm 1$ の時、$X,Y$ に線形関係が成り立つ。

$$Y = \mu_2 \pm \frac{\sigma_2}{\sigma_1}(X-\mu_1), \quad \text{(複号同順)} \tag{26}$$

(3) $X,Y$ が独立の時、相関係数は 0 である。

**証明**

(1) 任意の定数 $t$ に対して、$W = Z_2 - tZ_1$ とすると、

$$0 \leq E[W^2] = E[(Z_2)^2 tZ_1Z_2 + t^2(Z_1)^2] = 1 - 2\rho t + t^2 = (t-\rho)^2 + 1 - \rho^2$$

よって $|\rho| \leq 1$ となる。

(2) $\rho = \pm 1$ の時、$0 \leq E[W^2] = (t \mp 1)^2$ であるから、$t = \pm 1$ として $Z_2 = \pm Z_1$ が成り立つ (複号

同順)。
(3) $X, Y$ が独立なら、

$$Cov[X, Y] = E[(X-\mu_1)(Y-\mu_2)] = E[(X-\mu_1)]E[Y-\mu_2] = 0$$

となるから、$\rho = 0$。
※誤解しがちだが、独立 $\neq$ 無相関である。独立であっても無相関であることがあり得る (例題 4.1)。相関係数は 2 つの確率変数の線形関係に近いかどうかを確かめる指標であって、連関性をとらえる指標ではないことに注意。

## 4.4 独立な確率変数の和の分布

**定理.** 確率変数 $X, Y$ は独立でそれぞれ分布関数 $F(x), G(y)$ を持つとき、和 $Z = X + Y$ の分布関数を $H(z)$ とすれば

$$H(z) = \int_{-\infty}^{\infty} F(z-y)dG(y) = \int_{-\infty}^{\infty} G(z-x)dF(x) \tag{27}$$

が成り立つ。これを $H = F * G$ と書き、$F$ と $G$ のたたみこみという。それぞれの密度関数を小文字で書けば $Z$ の密度関数 $h(z)$ は

$$h(z) = \int_{-\infty}^{\infty} f(z-y)g(y)dy = \int_{-\infty}^{\infty} g(z-x)f(x)dx \tag{28}$$

となる。これを $h = f * g$ と書き、$f$ と $g$ のたたみこみという。$X, Y$ の特性関数をそれぞれ $\phi, \psi$ とすると $Z$ の特性関数 $\eta(t)$ はそれらの積で書ける: $\eta(t) = \phi(t)\psi(t)$。

**証明.** $X, Y$ の同時分布関数は $F(x)G(y)$ であるから

$$\begin{aligned} H(z) = P(X+Y \leq z) &= \int\int_{\{x+y\leq z\}} dF(x)dG(y) \\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-y} dF(x)\right)dG(y) = \int_{-\infty}^{\infty} F(z-y)dG(y) \end{aligned} \tag{29}$$

同様に $y$ について先に積分を実行すれば $H(z) = \int_{-\infty}^{\infty} G(z-x)dF(x)$ が得られる。
分布の集合を分布族という。分布族には確率関数や密度関数は共通で母数によって特徴づけられた分布の集合を考える。

$$\mathcal{F} = \{f(x|\theta) : \theta \in \Theta\} \tag{30}$$

$X, Y$ が独立でそれらの分布 $F, G$ が同じ分布族 $\mathcal{F}$ に属するとき $Z = X + Y$ の分布 $H$ も $\mathcal{H}$ に属するなら、$\mathcal{F}$ は再生性を持つという。再生性を持つ分布族は、確率変数の和の分布を求める場合、分布型が変わらないので母数を求めればよい。前の章で紹介したガウス分布、二項分布、ポアソン分布はすべて再生性を持つ。(定理 4.4) 再生性を持つかどうかの確認は、積率母関数を用いて確認できる。

演習問題 4.2

# 5 多次元確率ベクトルの分布

2 次元確率変数ベクトルの議論は、多次元確率変数ベクトルに対しても拡張できる。$n$ 次元確率変数ベクトル $X_1, X_2, ..., X_n$ に対して、分散共分散行列を

$$\Sigma_{ij} = Cov[X_i, X_j] = E[(X_i - \mu_i)(X_j - \mu_j)] \tag{31}$$

を行列要素にもつ実対称行列と定義する。また、$n$ 次元確率ベクトル $X$ と $m$ 次元確率ベクトル $Y$ に対して相互共分散行列を

$$(\Sigma_{\mathbf{xy}})_{ij} = Cov[X_i, Y_j] \tag{32}$$

を行列要素に持つ $n \times m$ 行列と定義する。

**定理.** $n$ 次元確率ベクトル $X$ と $m$ 次元確率ベクトル $Y$ を考える。定数列ベクトル $\mathbf{a}, \mathbf{b}$(それぞれ $n, m$ 次元) に対して、

$$E[\mathbf{a}^{\mathrm{T}} X] = \mathbf{a}^{\mathrm{T}} E[X] \tag{33}$$

$$V[\mathbf{a}^{\mathrm{T}} X] = \mathbf{a}^{\mathrm{T}} V[X] \mathbf{a} \tag{34}$$

$$Cov[\mathbf{a}^{\mathrm{T}} X, \mathbf{b}^{\mathrm{T}} Y] = \mathbf{a}^{\mathrm{T}} Cov[X, Y] \mathbf{b} \tag{35}$$

が成り立つ。証明は p.56 の定理 4.2 と同様。
これを拡張して、$p \times n$ 定数行列 $A$ と $q \times m$ 定数行列 $B$ による $X, Y$ の 1 次変換 $AX, BY$( それぞれ $p, q$ 次元ベクトル) に対して、

$$E[AX] = AE[X] \tag{36}$$

$$V[AX] = AV[X]A^{\mathrm{T}} \tag{37}$$

$$Cov[AX, BY] = ACov[X, Y]B^{\mathrm{T}} \tag{38}$$

が成り立つ。

**証明** 1 次変換を行ベクトルによる線形結合のベクトルとみなす。

$$AX = \begin{pmatrix} \mathbf{a}_1 X \\ \vdots \\ \mathbf{a}_p X \end{pmatrix}, \quad BY = \begin{pmatrix} \mathbf{b}_1 Y \\ \vdots \\ \mathbf{b}_q Y \end{pmatrix} \tag{39}$$

このベクトルの各成分についての平均、分散、共分散を考えると、上の定理より成り立つことがわかる。

**定理.** $n \times n$ 行列 $C$ と $n \times m$ 行列 $D$ に対して、二次形式

$$X^{\mathrm{T}} C X = \sum_{i,j}^{n} c_{ij} X_i X_j, \quad X^{\mathrm{T}} D Y = \sum_{i}^{n} \sum_{j}^{m} d_{ij} X_i Y_j$$

を新しい確率変数と考えれば、それらの平均は以下のように与えられる。

$$E[X^{\mathrm{T}} C X] = E[X]^{\mathrm{T}} C E[X] + trace[CV[X]] \tag{40}$$

$$E[X^{\mathrm{T}} D Y] = E[X]^{\mathrm{T}} D E[Y] + trace[DCov[X, Y]] \tag{41}$$

証明は教科書を確認。

## 無相関化・白色化

分散共分散行列は正定値行列[*1]であるから、直交行列を用いて対角化できる。

$$U^{\mathrm{T}}\Sigma U = \Lambda$$

ここで $U$ は直交行列で、$\Lambda$ は対角行列。このことから、確率変数ベクトル $X$ を $U^{\mathrm{T}}X$ と変換すれば、変換された新たな確率変数ベクトルに対する分散共分散行列は

$$V[U^{\mathrm{T}}X] = U^{\mathrm{T}}\Sigma U = \Lambda$$

と対角行列となり、したがって確率変数の相関が 0 となる。これを無相関化という。無相関化は主成分分析 (PCA) と関連が深い。
無相関化では一般に各確率変数の分散は異なるが、各確率変数を中心化してその分散を 1 に統一する操作を白色化と呼ぶ。すなわち、確率変数ベクトルを平均 0、分散共分散行列が単位行列になるように変数変換する。平均を 0 にするのは単純で、平均ベクトル $E[X] = \boldsymbol{\mu}$ を用いて $\tilde{X} = X - \boldsymbol{\mu}$ と変換すればよい。次に、中心化された確率変数ベクトルに対して、その分散共分散行列が単位行列になるような変換を考える。

$$\begin{aligned}
V[\tilde{X}] &= I \\
&= \Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}} \\
&= \Lambda^{-\frac{1}{2}}U^{\mathrm{T}}\Sigma U\Lambda^{-\frac{1}{2}} \\
&= \Lambda^{-\frac{1}{2}}V[U^{\mathrm{T}}\tilde{X}]\Lambda^{-\frac{1}{2}} \\
&= V[\Lambda^{-\frac{1}{2}}U^{\mathrm{T}}\tilde{X}]
\end{aligned}$$

よって、

$$X \to \Lambda^{-\frac{1}{2}}U^{\mathrm{T}}(X - \boldsymbol{\mu})$$

と変換すればよい。

独立な同一分布に従う確率変数 $X_1, ..., X_n$ のことを無作為標本という。その標本平均と標本分散は

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

で定義される。これまでの分布関数による平均のことを、特に、母平均 (数学的平均) と呼ぶことがある。$\bar{X}, S^2$ のような標本の関数を一般に統計量という。

[*1] 厳密には半正定値行列だが、簡単のため正定値行列とする。

**定理.** 平均 $\mu$、分散 $\sigma^2$ を持つ分布からの標本 $X_1, ..., X_n$ の標本平均 $\bar{X}$ の母平均・分散と、標本分散の母平均は次のようになる。

$$
\begin{aligned}
E[\bar{X}] &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu \\
V[\bar{X}] &= \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] = \frac{\sigma^2}{n} \\
E[S^2] &= \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \bar{X})^2] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[(X_i - \mu)^2] + \frac{1}{n}\sum_{i=1}^{n} E[-2\bar{X}X_i + 2\mu X_i + \bar{X}^2 - \mu^2] \\
&= \sigma^2 - E[-2\bar{X}^2 + 2\mu\bar{X} + \bar{X}^2 - \mu^2] \\
&= \sigma^2 - \underbrace{E[(\bar{X} - \mu)^2]}_{V[\bar{X}]} \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

## 5.3 多変量分布の代表的モデル一多変量正規分布

ここでは、多変量分布の例として、多変量正規分布について取り上げる。
$n$ 次元確率ベクトル $X = (X_1, ..., X_n)^{\mathrm{T}}$ に対して、$n$ 次元正規分布の同時確率密度関数は

$$
f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}))\right\} \tag{42}
$$

で与えられる。ここで $\boldsymbol{\mu}, \Sigma$ は $X$ の平均ベクトルと分散共分散行列である。
(1) $N_n(\mathbf{0}, \boldsymbol{I})$ の従う確率変数ベクトル $Z$ について考える。このとき、同時密度関数は

$$
\varphi_n(\boldsymbol{z}) = \exp\left(-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2}} \exp\left(-\frac{z_i^2}{2}\right) = \prod_{i=1}^{n} \varphi(z_i) \tag{43}
$$

と、標準正規分布の密度関数の積で表される。つまり、$Z$ は独立な $n$ 個の確率変数 $Z_1, ..., Z_n$ を成分に持つ確率変数ベクトルであることがわかる。この分布を $n$ 次元標準正規分布と呼ぶ。これに関連して、多変量正規分布の重要な性質の一つについて述べる。

**定理.** 確率変数ベクトル $X$ が多変量正規分布に従う時、$X$ が無相関ならば独立である。

$X$ が無相関の時、分散共分散行列は対角行列となるので、その同時密度関数は各成分の密度関数の積で表すことができ、この定理は成立する。

**定理.** 確率変数ベクトル $X$ が多変量正規分布 $\mathcal{N}(\boldsymbol{\mu}, \Gamma)$ に従う時、その一次変換 $AX$ を施した後の $X$ は $\mathcal{N}(A\boldsymbol{\mu}, A\Gamma A^{\mathrm{T}})$ に従う。

# 11 線形回帰