確率統計 復習

Kosuke Toda

目次

第 1 章

事象と確率

1.1　集合と事象

- 試行 (trials)
  - 実験や観測，調査の総称.
- 全事象 (total event)
  - 試行を行ったときに起こりうる全ての結果の集まり
- 母集団 (population)
  - 統計学では，全事象のことを試行の対象となる集団とみなす.
- 事象 (event)
  - ある性質を満たす結果の集まり

Ω をある集合 (全集合) とし，$\emptyset$ を空集合 (空事象) とする．可測集合の定義を以下で与える．

定義 1.1　Ω の部分集合を要素とする集合族 $\mathscr{A}$ が次の 3 条件を満たすとき，σ-加法族という．

1. $\Omega \in \mathscr{A}$, $\emptyset \in \mathscr{A}$.
2. $A \in \mathscr{A}$ ならば $A^{\mathrm{C}} \in \mathscr{A}$.
3. $A_n \in \mathscr{A}$ $(n = 1, 2, \ldots)$ ならば，$\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$.

また，$(\Omega, \mathscr{A})$ を可測空間，$\mathscr{A}$ の要素を可測集合という[*1].

[*1] テキストでは，$\mathscr{A}$ (σ-加法族) を事象族，$\mathscr{A}$ の要素を事象と呼んでいる．

1.2　確率と確率空間

定義 1.2　事象族 $\mathscr{A}$ 上で定義された関数 P が次の 3 条件を満たすとき，確率 (または確率測度 (probability measure)) という．

(P1)　任意の事象 $A \in \mathscr{A}$ に対して，$0 \leq P(A) \leq 1$.

(P2)　$P(\emptyset) = 0,\ P(\Omega) = 1$.

(P3)　$A_n \in \mathscr{A}\,(n = 1, 2, \ldots)$ が互いに素 $(A_m \cap A_n = \emptyset\,(m \leq n))$ のとき[a]，

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

[a] $\sigma -$ 加法性 という．

このとき，$(\Omega, \mathscr{A}, P)$ を確率空間 (probability space) と呼ぶ．(P1)-(P3) から確率測度に関する基本的性質が導かれる:

1. $P(A^{\mathrm{C}}) = 1 - P(A)$
2. $A \subset B \Longrightarrow P(A) \leq P(B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $A_n \in \mathscr{A},\ A_n \supset A_{n+1}\,(n = 1, 2, \ldots) \Longrightarrow P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$
5. $A_n \in \mathscr{A},\ A_n \subset A_{n+1}\,(n = 1, 2, \ldots) \Longrightarrow P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$

1.3　事象の独立性と従属性

定義 1.3（独立性）　2 つの事象 A, B が独立 (independent) であるとは，

$$P(A \cap B) = P(A)P(B)$$

が成り立つことをいう．独立でないとき，従属 (dependent) であるという．

定義 1.4（条件付き確率）　2 つの事象 A, B があって，$P(A) > 0$ のとき，

$$P(B|A) \coloneqq \frac{P(A \cap B)}{P(A)}$$

を，A が与えられたときの B の条件付き確率という．

条件付き確率について，以下の性質が成り立つ．

1. $P(A \cap B) = P(A)P(B|A)$
2. A, B が独立のとき，$P(B|A) = P(B)$, $P(A|B) = P(A)$

定理 1.1（全確率の法則） $A_1, A_2 \ldots, A_n$ を互いに素な事象で，$\Omega = \bigcup_{i=1}^n A_i$ とする．$P(A_i) > 0\,(i = 1, 2, \ldots, n)$ のとき，任意の事象 B に対して，

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i) \tag{1.1}$$

が成立する．

証明 1（定理 1.1 の証明）

$$B = B \cap \Omega = B \cap \left(\bigcup_{i=1}^{n} A_i \right) = \bigcup_{i=1}^{n} (B \cap A_i)$$

および

$$(B \cap A_i) \cap (B \cap A_j) = \emptyset \; (i \neq j)$$

より，定義 1.2 から

$$P(B) = P\left(\bigcup_{i=1}^{n} (B \cap A_i) \right) = \sum_{i=1}^{n} P(B \cap A_i)$$

が成立し，定義 1.4 より，(1.1) 式が成立する． □

事象 B が起こる前と後のことをそれぞれ事前と事後と呼ぶ．A_i に対して $P(A_i)$ は B が起こる前に与えられている確率であるから事前確率 (prior probability) といい，$P(A_i|B)$ は B が起こった後で与えられる確率であるから事後確率 (posterior probability) という．次のベイズの定理が成立する[*2]．

定理 1.2（ベイズの定理） $A_1, A_2, \ldots, A_n$ を互いに素な事象で，$\Omega = \bigcup_{i=1}^n A_i$ とする．$P(A_i) > 0$ $(i = 1, 2, \ldots, n)$ のとき，任意の事象 B に対して

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)} \quad (i = 1, 2, \ldots, n) \tag{1.2}$$

が成立する．

証明 2（定理 1.2 の証明） 定義 1.4 より，

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}.$$

条件付き確率の性質より，

$$P(A_i \cap B) = P(A_i)P(B|A_i).$$

[*2] ベイズの定理は，事前確率 $P(A_i)$, $P(B|A_i)$ $(i = 1, 2, \ldots, n)$ が求まれば，事後確率 $P(A_i|B)$ が導出できることを述べている．

定理 1.1 より，

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i) \tag{1.3}$$

であるから，(1.2) 式が成立する． □

第 2 章

確率変数と確率分布

2.1　母集団と標本

既知の無作為標本によって母集団の未知の性質を知ろうとすることを統計的推測 (statistical inference) という (図 2.1).

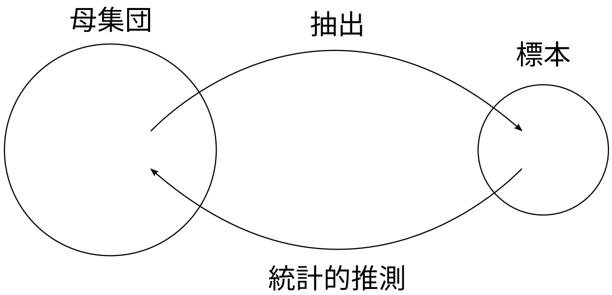


図 2.1: 統計的推測のイメージ

2.2　確率変数と確率分布

2.2.1　確率変数

まず，離散型の確率変数とその分布を考える．X を 2 個のサイコロを投げたときに出た目の和を表す変数とする (表 2.1).

表 2.1: 確率変数の例

X の値	2	3	4	$\cdots$	12	計
確率	1/36	2/36	3/36	$\cdots$	1/36	1

この X のように，取りうる値に対して，確率が対応する変数を確率変数 (random variable) と呼ぶ．確率変数の取りうる値 (実現値) とその確率をペアにしたものを確率分布 (probability distribution) と呼ぶ.

定義 2.1（分布関数）　確率変数 X の分布関数 (distribution function) を

$$F(x) = P(X \leq x)$$

によって定義する．

分布関数 $F(\cdot)$ は以下の性質を持つ．

(DF1)　任意の $x \in \mathbb{R}^1$ に対して，$0 \leq F(x) \leq 1$ であり，かつ，

$$F(-\infty) \coloneqq \lim_{x \to -\infty} F(x) = 0,\ \ F(\infty) \coloneqq \lim_{x \to \infty} F(x) = 1.$$

(DF2)　$F(x)$ は単調非減少である：$x < y \Longrightarrow F(x) \leq F(y)$.

(DF3)　$F(x)$ は右側連続である：$\lim_{y \to x+0} F(y) = F(x)$.

逆に，これらの性質を持つものを分布関数として定義してもよい．

確率変数 X の分布関数が $F(x)$ である場合に，「確率変数 X は分布 $F(\cdot)$ に従う」といい，“$X \sim F$” と書くことがある．

定義 2.2（離散分布）　確率変数 X の実現値が有限個または可算無限個の値の離散値 $x_1, x_2, \ldots, x_n, \ldots$ のみをとるとき，離散型確率変数といい，その分布を離散分布 (discrete distribution) という．各値 x_i に対する確率を p_i とすると，離散分布は

$$f_k = f(x_k) = P(X = x_k) = p_k\ \ (k = 1, 2, \ldots) \tag{2.1}$$

である．つまり，離散分布は各値に対する確率によって決定される．ただし，$\sum_{k=1}^{\infty} p_k = 1$, $p_k \geq 0$ $(k = 1, 2, \ldots)$ である．$f(\cdot)$ を確率関数 (probability function) という．

ただし，以下の点に注意する．

1. $\mathscr{X} = \{x_1, x_2, \ldots\}$ を標本空間 (sample space) という．
2. $P(X = x_k)$ は，確率変数 X が x_k をとる確率を意味する．
3. 確率変数は大文字 (ex. X)，実現値は小文字 (ex. x) で表す．
4. 確率変数 X の実現値が離散的なとき，離散型確率変数と呼ぶ．

確率関数 $f(\cdot)$ は次の性質を満たす．

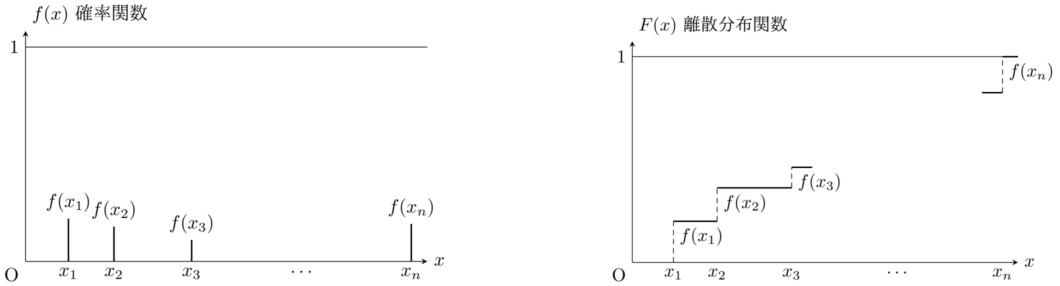
(PF1)　$f(x_k) \geq 0$.

(PF2)　$\sum_{k=1}^{\infty} f(x_i) = 1$.

(PF3)　分布関数は $F(x) = \sum_{\{i : x_i \leq x\}} f(x_i)$.

- $P(X = x_k) = F(x_k) - F(x_{k-1})$.

逆に，これらの性質をもつものを確率関数と定義してもよい．図 2.2 に確率変数が離散の場合の確率関数および分布関数を示す．



(a) 確率関数　　**(b)** 離散分布関数

図 2.2: 確率関数および離散分布関数

次に，連続型の確率変数と分布について定義する．

定義 2.3（連続分布）　確率変数 X が連続的な値をとるとき，連続型確率変数といい，その分布を連続分布 (continuous distribution) という．連続分布の分布関数が C^1 級である時，その導関数

$$f(x) = \frac{dF(x)}{dx}$$

を X の確率密度関数 (probability density function) という．

確率密度関数 $f(\cdot)$ は次の性質を満たす．

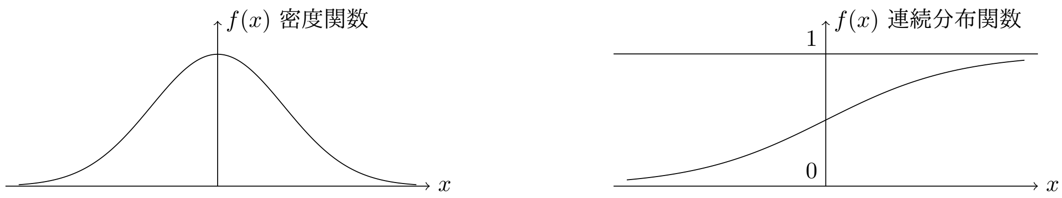
(Dn1)　$f(x) \geq 0$.

(Dn2)　$\int_{-\infty}^{\infty} f(x)dx = 1$.

(Dn3)　分布関数は $F(x) = \int_{-\infty}^{x} f(y)dy$.

- $P(a \leq X \leq b) = \int_a^b f(x)dx$

逆にこれらの性質をもつものを確率密度関数と定義してもよい．連続型確率変数の 1 点における確率はゼロである．図 2.3 に確率変数が連続の場合の確率密度関数および分布関数を示す．



(a) 確率密度関数　　**(b)** 連続分布関数

図 2.3: 確率密度関数および連続分布関数

以下，確率変数を r.v.，確率密度関数を p.d.f. と略記する．

2.3　分布の特性値：平均値と分散

定義 2.4（平均値）　1.　X が離散型 r.v. のとき $(P(X = x_i) = p_i\ (i = 1, 2, \ldots))$
X の平均値 $E[X]$ は，

$$E[X] \coloneqq \sum_{i=1}^{\infty} x_i p_i$$

とする．

2.　X が連続型 r.v. のとき $(P(a \leq X \leq b) = \int_a^b f(x)dx)$
X の平均値 $E[X]$ は，

$$E[X] \coloneqq \int_{-\infty}^{\infty} xf(x)dx \left(= \int_{-\infty}^{\infty} xdF(x) \right)$$

定義 2.5　$h(X)$ を連続関数として，$Y = h(X)$ となる r.v. を考える．

1. X が離散型 r.v. のとき

$$E[Y] = E[h(X)] \coloneqq \sum_{i=1}^{\infty} h(x_i) p_i$$

2. X が連続型 r.v. のとき

$$E[Y] = E[h(X)] \coloneqq \int_{-\infty}^{\infty} h(x) f(x) dx$$

期待値は，以下の性質を持つ．

(E1)　c を定数とすると，$E[c] = c$

(E2)　X, Y を r.v.，$h(X), k(X)$ を連続関数，a, b を定数とする．

$$E[ah(X) + bk(Y)] = aE[h(X)] + bE[k(Y)]$$

(E3)　$h(X) \geq 0 \Longrightarrow E[h(X)] \geq 0$

(E4)　$|E[h(X)]| \leq E[|h(X)|]$

定義 2.6（分散）　X を r.v. とし，$E[X] = \mu$, $E[X^2] < \infty$ とする．このとき，X の分散 $V[X]$ を

$$V[X] \coloneqq E[(X - E[X])^2] = E[(X - \mu)^2]$$

と定義する．また，標準偏差を $D[X] = \sqrt{V[X]}$ と定義する．

分散は，以下の性質を持つ．X を r.v.，$E[X] = \mu$, $E[X^2] < \infty$ とし，a, b を定数とすると，

(V1)　$V[X] = E[X^2] - \mu^2$

(V2)　$V[aX + b] = a^2 V[X]$

(V1) について，

$$\begin{aligned} V[X] &= E[(X - E[X])^2] = E[(X-\mu)^2] \\ &= E[X^2 - 2E[X]\mu + \mu^2] = E[X^2] - \mu^2 \end{aligned}$$

である．

(V2) について，

$$\begin{aligned} V[aX+b] &= E[(aX+b-E[aX+b])^2] \\ &= E[(aX+b-aE[X]-b)^2] = E[a^2(X-E[X])^2] \\ &= a^2E[(X-E[X])^2] = a^2V[X] \end{aligned}$$

である．

2.4 分布関数の変換

確率分布は分布関数

$$F(x) = \begin{cases} \sum_{x_k \le x} P(X = x_k) & (\text{離散型}) \\ \int_{-\infty}^{x} f(t)dt & (\text{連続型}) \end{cases}$$

として与えられる．関数のままでは分布の特徴をうまく表現できない．そこで，確率分布を扱いやすい関数に変換する．

1. 確率母関数 (probability generating function, p.g.f.)

   X: 離散型 r.v. $(x = 0, 1, 2, \dots)$

   $$P(t) \coloneqq E[t^X] = \sum_{x=0}^{\infty} t^x P(X = x),\ |t| \le 1$$

   を X の確率母関数という．

2. 積率母関数 (moment generating function, m.g.f.)

   X: r.v.

   $$M(t) \coloneqq E[e^{tX}] = \begin{cases} \sum_{x=0}^{\infty} e^{tx} P(X = x) & (\text{離散型}) \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & (\text{連続型}) \end{cases}$$

   を X の積率母関数という．

3. 特性関数 (characteristic function, c.f.)

   X: r.v.

   $$\phi(t) \coloneqq E[e^{itX}] = \begin{cases} \sum_{x=0}^{\infty} e^{itx} P(X = x) & (\text{離散型}) \\ \int_{-\infty}^{\infty} e^{itx} f(x)dx & (\text{連続型}) \end{cases}$$

   を X の特性関数という．

ただし，以下の性質が成り立つ．

- X が離散型 r.v. のとき，$M(t) = P(e^t)$, $\phi(t) = P(e^{it})$

- $M(t)$ は一般に存在するとは限らないが，$\phi(t)$ は常に存在する．
- $\phi(t) = M(it)$

モーメントの計算について，それぞれの関数は以下の性質が成り立つ．

p.g.f.　X: 離散型 r.v., $E[X] < \infty, V[X] < \infty$

$$\begin{aligned}\frac{dP(t)}{dt} &= \frac{d}{dt}\left(\sum_{x=0}^{\infty} t^x P(X=x)\right) = \frac{d}{dt}\left(P(X=0) + \sum_{x=1}^{\infty} t^x P(X=x)\right) \\ &= \sum_{x=1}^{\infty} \frac{d}{dt}(t^x P(X=x)) = \sum_{x=1}^{\infty} x t^{x-1} P(X=x)\end{aligned}$$

より，

$$\left.\frac{dP(t)}{dt}\right|_{t=1} = \sum_{x=1}^{\infty} xP(X=x) = \sum_{x=0}^{\infty} xP(X=x) = E[X]$$

が成り立つ．同様にして，

$$\begin{aligned}\frac{d^2P(t)}{dt^2} &= \frac{d}{dt}\left(\sum_{x=1}^{\infty} x t^{x-1} P(X=x)\right) = \frac{d}{dt}\left(P(X=1) + \sum_{x=2}^{\infty} x t^{x-1} P(X=x)\right) \\ &= \sum_{x=2}^{\infty} x(x-1) t^{x-2} P(X=x)\end{aligned}$$

より，

$$\left.\frac{d^2P(t)}{dt}\right| = \sum_{x=2}^{\infty} x(x-1)P(X=x) = \sum_{x=0}^{\infty} x(x-1)P(X=x) = E[X(X-1)]$$

分散は，

$$\begin{aligned}V[X] &= E[X^2] - (E[X])^2 = E[X(X-1)] + E[X] - (E[X])^2 \\ &= \left.\frac{d^2P(t)}{dt^2}\right|_{t=1} + \left.\frac{dP(t)}{dt}\right|_{t=1} + \left(\left.\frac{dP(t)}{dt}\right|_{t=1}\right)^2\end{aligned}$$

で求められる．

m.g.f.　X: (主に連続型)r.v.

$$\begin{aligned}\frac{dM(t)}{dt} &= \frac{d}{dt}\left(\int_{-\infty}^{\infty} e^{tx} f(x)dx\right) = \int_{-\infty}^{\infty}\left(\frac{d}{dt}e^{tx}\right) f(x)dx \\ &= \int_{-\infty}^{\infty} x e^{tx} f(x)dx\end{aligned}$$

より，

$$\left.\frac{dM(t)}{dt}\right|_{t=0} = \int_{-\infty}^{\infty} xf(x)dx = E[X].$$

また，

$$\frac{d^2M(t)}{dt^2} = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)dx$$

より，

$$\left.\frac{d^2M(t)}{dt^2}\right|_{t=0} = \int_{-\infty}^{\infty} x^2 f(x)dx = E[x^2].$$

離散型 r.v. についても同様.

c.f.　$E[X^2] < \infty$ とする．m.g.f. と同様の議論により，

$$\left.\frac{d\phi(t)}{dt}\right|_{t=0} = E[iX] = iE[X] \quad \therefore E[X] = (-i) \left.\frac{d\phi(t)}{dt}\right|_{t=0}$$

および

$$\left.\frac{d^2\phi(t)}{dt^2}\right|_{t=0} = E[-X^2] \quad \therefore E[X^2] = -\left.\frac{d^2\phi(t)}{dt^2}\right|_{t=0}$$

が得られ，

$$V[X] = \left.\frac{d^2\phi(t)}{dt^2}\right|_{t=0} + \left(\left.\frac{d\phi(t)}{dt}\right|_{t=0}\right)^2$$

となる.

積率母関数の対数 $\psi(t) = \log M(t)$ をキュミュラント母関数 (cumulant generating function, c.g.f.) という．c.g.f. について，

$$\frac{d\psi(t)}{dt} = \frac{M'(t)}{M(t)}, \quad \left.\frac{d\psi(t)}{dt}\right|_{t=0} = \frac{M'(0)}{M(0)} = E[X]$$

および

$$\frac{d^2\psi(t)}{dt^2} = \frac{M(t)M''(t) - M'(t)^2}{M(t)^2}, \quad \left.\frac{d^2\psi(t)}{dt^2}\right|_{t=0} = \frac{E[X^2] - (E[X])^2}{1} = V[X]$$

が成り立つ.

以上はモーメントの計算について述べたが，確率分布の同値性および分布の終息について，以下の定理が成り立つ.

定理 2.1　1.　一致性

X_1, X_2　: r.v.

F_1, F_2　: X_1, X_2 の分布関数

ϕ_1, ϕ_2　: X_1, X_2 の特性関数

とするとき，次が成立.

$$F_1 = F_2 \Longleftrightarrow \phi_1 = \phi_2$$

2.　連続性

$X, X_n\ (n = 1, 2, \ldots)$　: r.v.

$F, F_n\ (n = 1, 2, \ldots)$　: X, X_n の分布関数

$\phi, \phi_n\ (n = 1, 2, \ldots)$　: X, X_n の特性関数

とするとき，次が成立.

$$\lim_{n\to\infty} F_n(x) = F(x) \Longleftrightarrow \lim_{n\to\infty} \phi_n(t) = \phi(t)$$

が成り立つ．ただし，x は F の連続点，t は任意である.

第 3 章

確率分布の代表的モデル

3.1　離散分布モデル

ここでは，離散一様分布，二項分布，ポアソン分布について，その分布の特性値を求める．

3.1.1　離散一様分布 $DU(n)$

確率変数 X が値 $1, \ldots, n$ を等確率 $1/n$ でとるとき，その確率関数は

$$f(x) = P(X = x) = \frac{1}{n}, \ \ x = 1, \ldots, n$$

である．このような分布を離散一様分布 (discrete uniform distribution) といい，記号 $DU(n)$ で表す．

$DU(n)$ について，平均，分散はそれぞれ，

$$\begin{aligned} E[X] &= \sum_{x=1}^{n} x\frac{1}{n} = \frac{n+1}{2} \\ V[X] &= \sum_{x=1}^{n} x^2\frac{1}{n} - (E[X])^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12} \end{aligned}$$

で与えられる．

3.1.2　二項分布 $B_N(n, p)$

成功の確率が $p\,(0 < p < 1)$，失敗の確率が $1-p$ の試行をベルヌーイ試行 (Bernoulli trial) という．その試行の成功を 1，失敗を 0 とするとき，2 値 (bivariate) r.v. ε が得られる:

$$\varepsilon = \begin{cases} 1 & (\text{確率 } p), \\ 0 & (\text{確率 } 1-p). \end{cases}$$

この分布をベルヌーイ分布といい，記号 $Ber(p)$ で表す．確率関数は，

$$f(\varepsilon|p) = p^{\varepsilon}(1-p)^{\varepsilon}, \ \varepsilon = 0, 1$$

となる．特に，成功の確率と失敗の確率の比 $p/(1-p)$ をオッズ (odds) という．

ベルヌーイ分布の平均と分散は，

$$E[\varepsilon] = 1 \times p + 0 \times (1-p) = p,$$
$$V[\varepsilon] = E[\varepsilon^2] - (E[\varepsilon])^2 = p - p^2 = p(1-p).$$

次に，n 回の独立なベルヌーイ試行を $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ とし，そのときの成功の回数を $X = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_n$ とおくと，X は離散値 $0, 1, \ldots, n$ をとる r.v. で，その確率関数は次のようになる:

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \ \ x = 0, 1, \ldots, n.$$

二項定理より，

$$\sum_{x=0}^{n} f(x|p) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-p))^n = 1.$$

この分布を二項分布 (binomial distribution) といい，記号 $B_N(n,p)$ で表す．ただし，$B_N(n,p) = Ber(p)$ である．

p.g.f. は，

$$\begin{aligned} P(t) = E[t^X] &= \sum_{x=0}^{\infty} t^x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{\infty} \binom{n}{x} (tp)^x (1-p)^{n-x} = (pt + 1 - p)^n \end{aligned}$$

である．この微分を計算することで，k 次階乗モーメントが計算できる:

$$P'(t) = np(pt + 1 - p)^{n-1}, \ \ \therefore E[X] = P'(1) = np.$$
$$P''(t) = n(n-1)p^2(pt + 1 - p)^{n-2}, \ \ \therefore E[X(X-1)] = P''(1) = n(n-1)p^2.$$

これから分散も計算でき，

$$V[X] = np(1-p)$$

が得られる．

3.1.3 ポアソン分布 $P_O(\lambda)$

「まれな現象の大量観測」によって発生する事象の個数は，ポアソン分布 (Poisson distribution) に従う．例えば，1 台の自動車が 1 日に交通事故を起こす確率は小さいが，自動車の台数は多いので，1 日の交通事故の件数はポアソン分布に従うことが知られている．ポアソン分布は記号 $P_O(\lambda)$ で表し，その確率関数は，

$$f(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, \ \ x = 0, 1, 2, \ldots,$$
$$\Theta = \{\lambda : 0 < \lambda < \infty\}$$

で与えられる．母数 λ のことを強度 (intensity) という．指数関数のべき級数展開により，

$$\sum_{x=0}^{\infty} f(x|\lambda) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

が得られる．

ポアソン分布の確率母関数は，

$$P(t) = E[t^X] = e^{-\lambda} \sum_{x=0}^{\infty} \frac{t^x \lambda^x}{x!} = e^{-\lambda} e^{t\lambda} = e^{\lambda(t-1)}.$$

であり，

$$\begin{aligned} P'(t) &= \lambda e^{\lambda(t-1)} \quad \therefore E[X] = P'(1) = \lambda \\ P''(t) &= \lambda^2 e^{\lambda(t-1)} \quad \therefore E[X(X-1)] = P''(1) = \lambda^2 \end{aligned}$$

から分散は $V[X] = \lambda$ と求められる．つまり，ポアソン分布の平均と分散は等しく強度 λ である．ポアソン分布は，ある期間に平均 λ 回起こる事象がある期間に起こる回数 X が従う分布である．成功確率が小さいが，試行回数が大きいとき，二項分布に対する近似分布として，ポアソン分布が導出される．

$n \to \infty$ のとき，$np_n \to \lambda$ とする．すなわち，

$$p_n = \frac{\lambda + o(1)}{n}$$

とする．このとき，二項分布の確率関数は，

$$\begin{aligned} f_n(x|p_n) &= \binom{n}{x} p_n^x (1-p_n)^{n-x} \\ &= \frac{n!}{x!(n-x)!} \left(\frac{\lambda + o(1)}{n}\right)^x \left(1 - \frac{\lambda + o(1)}{n}\right)^{n-x} \\ &= \frac{(\lambda + o(1))^x}{x!} \frac{n(n-1)\cdots(n-x+1)}{n^x} \left(1 - \frac{\lambda + o(1)}{n}\right)^{n-x} \\ &\to \frac{\lambda^x}{x!} e^{-\lambda} = f(x|\lambda) \end{aligned}$$

となり，ポアソン分布の確率関数が得られる．

3.2 連続分布モデル

ここでは，一様分布，正規分布および指数分布について，その分布の特性値を求める．

3.2.1 一様分布 $U(\alpha, \beta)$

確率変数 X が区間 (α, β) 上の値を等確率でとるとき，その密度関数は，

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & (\alpha < x < \beta) \\ 0 & (\text{otherwise}). \end{cases}$$
$$\Theta = \{\boldsymbol{\theta} = (\alpha, \beta) : -\infty < \alpha < \beta < \infty\}$$

と記述される．このような分布を一様分布 (uniform distribution) といい，記号 $U(\alpha, \beta)$ で表す．

一様分布の平均および分散は，

$$E[X] = \frac{\alpha+\beta}{2}$$
$$V[X] = \frac{(\beta-\alpha)^2}{12}$$

と求められる．

3.2.2　正規分布 $N(\mu, \sigma^2)$

正規分布 (normal distribution) / ガウス分布 (Gaussian distribution) は，次の密度関数を持つ．

$$f(x|\boldsymbol{\theta}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \ \ -\infty < x < \infty,$$
$$\Theta = \{\boldsymbol{\theta} = (\mu, \sigma^2) : -\infty < \mu < \infty,\, 0 < \sigma^2 < \infty\}.$$

正規分布は母数 μ, σ^2 にのみ依存するので，記号 $N(\mu, \sigma^2)$ で表す．この母数 μ, σ^2 はそれぞれ平均と分散である．特に，$\mu = 0, \sigma^2 = 1$ のとき，$N(0,1)$ を標準正規分布 (standard normal distribution) という．標準正規分布の密度関数は

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

という記号で表す．一般の正規分布 $N(\mu, \sigma^2)$ の密度関数は，

$$f(x|\boldsymbol{\theta}) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

のように，$\varphi(z)$ の線形変換として記述できる．

標準化　r.v. X が平均 μ，分散 σ^2 を持つとする．このとき，

$$Z := \frac{X-\mu}{\sigma}$$

とすると，$E[Z] = 0, V[Z] = 1$ となる．この変換を標準化という．

標準化を用いると，

$$X \sim N(\mu, \sigma^2) \Longrightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

とすることができる．

つまり，確率変数 $X \sim N(\mu, \sigma^2)$ の性質は標準化によって，$Z \sim N(0,1)$ の性質に変換することができる．逆に，$Z \sim N(0,1)$ から $X \sim N(\mu, \sigma^2)$ を構成することができる．

標準正規分布 $N(0,1)$ の積率母関数は，

$$M_z(t) = E[e^{tZ}] = \int_{-\infty}^{\infty} e^{tz}\varphi(z)dz = e^{\frac{t^2}{2}}$$

であり，正規分布 $N(\mu, \sigma^2)$ の積率母関数は，

$$M_x(t) = E[e^{tX}] = E[e^{t(\mu+\sigma Z)}] = e^{\mu t}E[e^{(t\sigma)}Z] = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

である．

3.2.3　指数分布 $E_X(\lambda)$

指数分布 (exponential distribution) は，以下のような密度関数を持つ．

$$f(x|\lambda) = \lambda e^{-\lambda x}, 0 < x < \infty,$$
$$\Theta = \{\lambda : 0 < \lambda < \infty\}.$$

母数 λ をもつ指数分布を $E_X(\lambda)$ と表す．

積率母関数は，

$$M(t) = E[e^{tX}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \int_0^\infty \lambda e^{-(\lambda-t)x}dx = \frac{\lambda}{\lambda - t}$$

であり，積率母関数を微分することで，

$$M'(t) = \frac{\lambda}{(\lambda-t)^2} \quad \therefore E[X] = M'(0) = \frac{1}{\lambda}$$
$$M''(t) = \frac{2\lambda}{(\lambda-t)^3} \quad \therefore E[X^2] = M''(0) = \frac{2}{\lambda^2}$$

がえられ，分散は $V[X] = 1/\lambda^2$ と求められる．

第 4 章

2 変量 (多変量) 確率ベクトルの分布

4.1　n 次元確率ベクトルの同時分布

一般に，n 個の 1 次元確率変数 $X_1, X_2, \ldots, X_n$ に対して，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を n 次元確率ベクトル (random vector, r.vec.)，$P(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2, \cdots, a_n \leq X_n \leq b_n)$, $a_i \leq b_i$ $(i = 1, 2, \ldots, n)$ を $\boldsymbol{X}$ の同時確率分布という．$F_{\boldsymbol{X}}(x_1, \cdots x_n) = P(X_1 \leq x_1, \cdots, X_n \leq x_n)$ を $\boldsymbol{X}$ の同時分布関数という．以降，主に連続型について述べる．

定義 4.1（同時確率密度関数）　$\boldsymbol{X} = (X_1, \ldots, X_n)$ を n 次元確率ベクトルとする．

$$P(a_1 \leq X_1 \leq b_1, \cdots, a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\boldsymbol{X}}(x_1, \cdots, x_n) dx_1 \cdots dx_n$$

で与えられ，次を満たす $f_{\boldsymbol{X}}(x_1, \cdots, x_n)$ を $\boldsymbol{X}$ の (n 次元) 同時確率密度関数 (joint p.d.f.) という．

1. $f_{\boldsymbol{X}}(x_1, \cdots, x_n) \geq 0.$
2. $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{X}}(x_1, \cdots, x_n) dx_1 \cdots dx_n = 1.$

同時確率密度関数は，

$$f_{\boldsymbol{X}}(x_1, \cdots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\boldsymbol{X}}(x_1, \cdots, x_n)$$

および

$$F_{\boldsymbol{X}}(x_1, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\boldsymbol{X}}(y_1, \cdots, y_n) dy_1 \cdots dy_n$$

を満たす．

各成分だけに注目した分布関数

$$\begin{aligned} F_i(x_i) &= F_{\boldsymbol{X}}(\infty, \cdots, \infty, x_i, \infty, \cdots, \infty) \\ &= P(X_1 \in \mathbb{R}, \cdots, X_{i-1} \in \mathbb{R}, X_i \leq x_i, X_{i+1} \in \mathbb{R}, \cdots, X_n \in \mathbb{R}) \\ &= P(X_i \leq x_i) \end{aligned}$$

を X_i の周辺分布関数という．その分布関数を

$$f_{X_i}(x_i) = \frac{\partial F_i(x_i)}{\partial x_i} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

を周辺確率密度関数という．周辺確率密度関数は，

$$\int_{-\infty}^{x_i} f_{X_i}(x)dx = F_i(x_i)$$

を満たす．