4　主要な確率分布

4.1　二項分布

ベルヌーイ試行（Bernoulli trial）　試行の結果として事象 A（成功）とその余事象 $\bar{A}$（失敗）のみを考えるとき，その試行のことをベルヌーイ試行という。成功の確率を $P(A) = p$，失敗の確率を $P\left(\bar{A}\right) = 1 - p$ とし，試行の結果が成功なら $X = 1$，失敗なら $X = 0$ として定めた X の確率分布をベルヌーイ分布という。

$X:$	0	1
$P:$	$1-p$	p

期待値と分散は次のようになる。この分布は $n = 1$ の場合の二項分布に等しい。

$$\begin{aligned} E[X] &= 0 \cdot (1-p) + 1 \cdot p = p \\ V[X] &= (0-p)^2(1-p) + (1-p)^2 p = p(1-p) \end{aligned}$$

二項分布（binomial distribution）　互いに独立で，成功率が p のベルヌーイ試行を n 回繰り返したとき，成功の回数を X とする。X の確率分布のことを二項分布といい，$B(n,p)$ と表す [*i]。このことを「X は $B(n,p)$ に従う」または $X \sim B(n,p)$ のように表す。二項定理から確率の総和は 1 になる。 *i

$X:$	0	$\cdots$	k	$\cdots$	n
$P:$	${}_n\mathrm{C}_0\,(1-p)^n$	$\cdots$	${}_n\mathrm{C}_k\,p^k(1-p)^{n-k}$	$\cdots$	${}_n\mathrm{C}_n\,p^n$

たとえば，3 枚の硬貨を同時に投げるとき，表の出る枚数 X の確率分布は二項分布 $B(3, 1/2)$ になる。

$X:$	0	1	2	3
$P:$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

二項分布の期待値・分散　期待値や分散を求めるために，次の式を利用する。

$$k \cdot {}_n\mathrm{C}_k = n \cdot {}_{n-1}\mathrm{C}_{k-1}, \qquad k(k-1) \cdot {}_n\mathrm{C}_k = n(n-1) \cdot {}_{n-2}\mathrm{C}_{k-2}$$

第 1 式は

$${}_n\mathrm{C}_k = \frac{n!}{k!(n-k)!} = \frac{n}{k} \cdot \frac{(n-1)!}{(k-1)!(n-k)!} = \frac{n}{k} \cdot {}_{n-1}\mathrm{C}_{k-1}$$

第 2 式は

$${}_n\mathrm{C}_k = \frac{n!}{k!(n-k)!} = \frac{n(n-1)}{k(k-1)} \cdot \frac{(n-2)!}{(k-2)!(n-k)!} = \frac{n(n-1)}{k(k-1)} \cdot {}_{n-2}\mathrm{C}_{k-2}$$

[*i] 定数 n，p のことを二項分布 $B(n,p)$ の母数（parameter）という。

より示すことができる。次の二項定理も利用する。

$$\sum_{k=0}^{n} {}_n\mathrm{C}_k\, p^k (1-p)^{n-k} = \{p + (1-p)\}^n = 1$$

*ii 上の性質から，二項分布の期待値を求める。[*ii]

$$\begin{aligned} E\left[X\right] &= \sum_{k=0}^{n} k \cdot {}_n\mathrm{C}_k\, p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} n \cdot {}_{n-1}\mathrm{C}_{k-1}\, p\, p^{k-1} (1-p)^{n-k} \\ &= np \sum_{k=0}^{n-1} {}_{n-1}\mathrm{C}_k\, p^k (1-p)^{n-k-1} = np \end{aligned}$$

次に，分散のために $E\left[X(X-1)\right]$ を求める。

$$\begin{aligned} E\left[X(X-1)\right] &= \sum_{k=0}^{n} k(k-1) \cdot {}_n\mathrm{C}_k\, p^k (1-p)^{n-k} \\ &= \sum_{k=2}^{n} n(n-1) \cdot {}_{n-2}\mathrm{C}_{k-2}\, p^2 p^{k-2} (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} {}_{n-2}\mathrm{C}_k\, p^k (1-p)^{n-k-2} = n(n-1)p^2 \end{aligned}$$

一般に $E\left[X^2\right] = E\left[X(X-1) + X\right] = E\left[X(X-1)\right] + E\left[X\right]$ が成り立つので，

$$\begin{aligned} V\left[X\right] &= E\left[X^2\right] - \{E\left[X\right]\}^2 = E\left[X(X-1)\right] + E\left[X\right] - \{E\left[X\right]\}^2 \\ &= n(n-1)p^2 + np - (np)^2 = np(1-p) \end{aligned}$$

以上のことから，二項分布 $B(n,p)$ の期待値と分散は次のようになる。

$$E\left[X\right] = np, \qquad V\left[X\right] = np(1-p) \tag{1}$$

ベルヌーイ分布の和としての二項分布 互いに独立で，成功率が p のベルヌーイ試行を n 回繰り返すとき，i 番目の試行の結果が成功なら $X_i = 1$，失敗なら $X_i = 0$ とする。それらの和 X は成功の回数だから，X は二項分布 $B(n,p)$ に従う。

$$X = X_1 + X_2 + \cdots + X_n, \qquad X \sim B(n,p)$$

ベルヌーイ変数 X_i は，$E\left[X_i\right] = p$，$V\left[X_i\right] = p(1-p)$ を満たす。独立な確率変数の和についての公式（3.3 節，3.4 節）を用いると，二項分布の期待値，分散は，

$$\begin{aligned} E\left[X\right] &= E\left[X_1\right] + E\left[X_2\right] + \cdots + E\left[X_n\right] = np \\ V\left[X\right] &= V\left[X_1\right] + V\left[X_2\right] + \cdots + V\left[X_n\right] = np(1-p) \end{aligned}$$

として導くこともできる。

[*ii] 二項分布の期待値や分散は, 結果だけ覚えておけばよい。

4.2 その他の離散型確率分布

離散一様分布（uniform distribution, discrete） 次の表で定義される X の確率分布を離散一様分布といい，$DU(n)$ と表す。たとえば，硬貨投げの結果は $n=2$, さいころ投げの結果は $n=6$ の離散一様分布になる。

$X:$	1	2	$\cdots$	n
$P:$	$\dfrac{1}{n}$	$\dfrac{1}{n}$	$\cdots$	$\dfrac{1}{n}$

期待値は

$$E[X]=\sum_{k=1}^{n}k\cdot\frac{1}{n}=\frac{1}{n}\cdot\frac{n(n+1)}{2}=\frac{n+1}{2}$$

分散は

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

幾何分布（geometric distribution） 独立で，成功率 p のベルヌーイ試行を繰り返すとき，初めて成功するまでに行った試行の回数 X の確率分布を幾何分布といい，$Ge(p)$ と表す *iii。たとえば，さいころ投げを繰り返すとき，初めて 1 の目が出るまでの回数の分布は $p=1/6$ の幾何分布になる。 *iii

$X:$	1	2	3	$\cdots$
$P:$	p	$(1-p)p$	$(1-p)^2p$	$\cdots$

期待値と分散は

$$E[X]=\frac{1}{p},\qquad V[X]=\frac{1-p}{p^2}$$

であるが，これの導出はやや面倒である。

ポアソン分布（Poisson distribution） 次の表で定義される X の確率分布をポアソン分布といい，$Po(\lambda)$ と表す。ただし $\lambda>0$ で，$e=2.71828\cdots$ はネイピア数である。たとえば，タクシーがランダムに到着するとき，一定時間に到着するタクシーの台数はポアソン分布になる。

$X:$	0	1	2	$\cdots$
$P:$	$\dfrac{e^{-\lambda}}{0!}$	$\dfrac{e^{-\lambda}\lambda}{1!}$	$\dfrac{e^{-\lambda}\lambda^2}{2!}$	$\cdots$

期待値と分散はともに λ である。

$$E[X]=\lambda,\qquad V[X]=\lambda$$

*iii 初めて成功するまでに行った失敗の回数を X とする定義もある。

超幾何分布（hypergeometric distribution）　M 個の成功を含む N 個の母集団から，非復元抽出で n 個を取り出すとき，取り出した成功の数 X の確率分布を超幾何分布といい，$HG(N, M, n)$ と表す。ただし $N > M > 0$，$M \geqq n$，$N - M \geqq n$，$n \geqq 0$ とする。たとえば，M 個の赤玉と $N - M$ 個の白玉が入った袋から異なる n 個を取り出すとき，取り出した赤玉の個数は超幾何分布になる。

$X:$	0	1	$\cdots$	n
$P:$	$\dfrac{{}_M\mathrm{C}_0\ {}_{N-M}\mathrm{C}_n}{{}_N\mathrm{C}_n}$	$\dfrac{{}_M\mathrm{C}_1\ {}_{N-M}\mathrm{C}_{n-1}}{{}_N\mathrm{C}_n}$	$\cdots$	$\dfrac{{}_M\mathrm{C}_n\ {}_{N-M}\mathrm{C}_0}{{}_N\mathrm{C}_n}$

期待値と分散は

$$E[X] = n\frac{M}{N}, \qquad V[X] = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\left(1 - \frac{n-1}{N-1}\right)$$

であるが，これの導出は非常に面倒である。

超幾何分布・二項分布・ポアソン分布の確率関数　X は N，M，n で定まる超幾何分布に従うとする。

$$P(X = k) = \frac{{}_M\mathrm{C}_k\ {}_{N-M}\mathrm{C}_{n-k}}{{}_N\mathrm{C}_n}$$

$p = M/N$ とおき，n と p を固定したまま，$N \to \infty$（$M = Np \to \infty$）の極限を考えると，$P(X = k) \to {}_n\mathrm{C}_k\, p^k(1-p)^{n-k}$ が成り立つ。その極限を分布とする Y は n と p で定まる二項分布になる。

$$P(Y = k) = {}_n\mathrm{C}_k\, p^k(1-p)^{n-k}$$

さらに $\lambda = np$ とおき，λ を固定したまま，$n \to \infty$（$p = \lambda/n \to 0$）の極限を考えると，$P(Y = k) \to e^{-\lambda}\lambda^k/k!$ が成り立つ。その極限を分布とする Z は λ で定まるポアソン分布になる。

$$P(Z = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

4.3　主な連続型確率分布

一様分布（uniform distribution, continuous）　$a \leqq x \leqq b$ の範囲のどの値をとる確率も均一となるような確率分布を一様分布といい，$U(a, b)$ と表す。たとえば，コンピュータで発生させた通常の乱数は 0 以上 1 未満の実数で，一様分布 $U(0, 1)$ になる。確率密度関数は次のようになる。

$$f(x) = \frac{1}{b-a} \quad (a \leqq x \leqq b)$$

$a \leqq x \leqq b$ でない場合は $f(x) = 0$ とする。期待値と分散は次のようになる。

$$E[X] = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{a+b}{2}$$

$$V[X] = \int_a^b x^2 \cdot \frac{1}{b-a}\,dx - \{E[X]\}^2 = \frac{(b-a)^2}{12}$$

指数分布（exponential distribution）　次の確率密度関数で定義される確率分布を指数分布といい，$Ex(\lambda)$ と表す。ただし $\lambda > 0$，$e = 2.71828\cdots$ である。たとえば，タクシーがランダムに到着するとき，到着する時間の間隔は指数分布になる。

$$f(x) = \lambda e^{-\lambda x} \quad (x \geqq 0)$$

$x < 0$ の場合は $f(x) = 0$ とする。期待値と分散は次のようになる。

$$E[X] = \frac{1}{\lambda}, \qquad V[X] = \frac{1}{\lambda^2}$$

正規分布（normal distribution）　次の確率密度関数で定義される確率分布を正規分布といい，$N(\mu, \sigma^2)$ と表す。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

期待値と分散は次のようになる。

$$E[X] = \mu, \qquad V[X] = \sigma^2$$

特に，$\mu = 0$，$\sigma^2 = 1$ の正規分布を**標準正規分布** $N(0, 1)$ という。

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

正規分布は他の分布の極限として現れる重要な連続型確率分布である。

参考文献

統計学入門（基礎統計学）
東京大学教養学部統計学教室（編）東京大学出版会 978-4-13-042065-5

統計学
久保川 達也 （著） 東京大学出版会 978-4-13-062921-8

はじめての統計学
道家 暎幸（共著） コロナ社 978-4-339-06113-0

確率統計 新版 （新版数学シリーズ）
岡本 和夫 （ほか著） 実教出版 978-4-407-32171-5

統計学序論 改訂版
山本 義郎 （著） 東海大学出版部 978-4-486-02133-9

確率統計（高専テキストシリーズ）
上野 健爾（監修） 森北出版 978-4-627-05561-2

基本統計学 第４版
宮川 公男 （著） 有斐閣 978-4-641-16455-0

新統計入門
小寺 平治 （著） 裳華房 978-4-7853-1099-8

Schaum's Outline of Introduction to Probability and Statistics
Seymour Lipschutz （著） McGraw-Hill Education 978-0-07-176249-6

A Dictionary of Statistics
Graham Upton （著） Oxford Univ Pr 978-0-19-967918-8

www5e.biglobe.ne.jp/~emm386/statistics/