5　二項分布の正規近似

5.1　標準正規分布表の使い方

標準正規分布　標準正規分布に従う確率変数 Z の確率密度関数は次のものである。

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

上の確率密度関数の形から，Z の確率は $x = 0$ について対称である。

$$P(-z \leqq Z \leqq 0) = \int_{-z}^{0} f(x)dx = \int_{0}^{z} f(x)dx = P(0 \leqq Z \leqq z)$$

全事象の確率は 1 なので，$P(0 \leqq Z < \infty) = 0.5$ である。[*i] [*i]

$$P(0 \leqq Z < \infty) = \int_{0}^{\infty} f(x)dx = \frac{1}{2}\int_{-\infty}^{\infty} f(x)dx = \frac{1}{2} \times 1 = 0.5$$

以上のことから，標準正規分布に従う Z は次の性質をもつ。

$$P(-z \leqq Z \leqq 0) = P(0 \leqq Z \leqq z), \qquad P(0 \leqq Z < \infty) = 0.5 \tag{1}$$

標準正規分布表　標準正規分布に従う Z と正数 z に対して，確率 $P(0 \leqq Z \leqq z)$ は標準正規分布表から求めることができる。z の小数第 1 位までを表側に，小数第 2 位を表頭に当てはめると，それらの交わる場所に確率が表示されている。[*ii] [*ii]

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767

[*i] 1 点の確率は $P(Z = 0) = 0$ だから $P(0 < Z < \infty) = P(0 \leq Z < \infty)$。

[*ii] 「表側」とは表の左端の見出し,「表頭」とは表の上端の見出しのことである。

1. (a) $P\,(0 \leqq Z \leqq 0.63) = 0.2357$

(b) $P\,(-1.35 \leqq Z \leqq 0) = P\,(0 \leqq Z \leqq 1.35) = 0.4115$

2. (a) $P\,(-1.35 \leqq Z \leqq 0.63) = P\,(-1.35 \leqq Z \leqq 0) + P\,(0 \leqq Z \leqq 0.63)$
$= P\,(0 \leqq Z \leqq 1.35) + P\,(0 \leqq Z \leqq 0.63)$
$= 0.4115 + 0.2357 = 0.6472$

(b) $P\,(0.63 \leqq Z \leqq 1.35) = P\,(0 \leqq Z \leqq 1.35) - P\,(0 \leqq Z \leqq 0.63)$
$= 0.4115 - 0.2357 = 0.1758$

(c) $P\,(-1.35 \leqq Z \leqq -0.63) = P\,(-1.35 \leqq Z \leqq 0) - P\,(-0.63 \leqq Z \leqq 0)$
$= P\,(0 \leqq Z \leqq 1.35) - P\,(0 \leqq Z \leqq 0.63)$
$= 0.4115 - 0.2357 = 0.1758$

3. (a) $P\,(0.63 \leqq Z < \infty) = P\,(0 \leqq Z < \infty) - P\,(0 \leqq Z \leqq 0.63)$
$= 0.5 - 0.2357 = 0.2643$

(b) $P\,(-\infty < Z \leqq -1.35) = P\,(-\infty < Z \leqq 0) - P\,(-1.35 \leqq Z \leqq 0)$
$= P\,(0 \leqq Z < \infty) - P\,(0 \leqq Z \leqq 1.35)$
$= 0.5 - 0.4115 = 0.0885$

4. (a) $P\,(-\infty < Z \leqq 0.63) = P\,(-\infty < Z \leqq 0) + P\,(0 \leqq Z \leqq 0.63)$
$= P\,(0 \leqq Z < \infty) + P\,(0 \leqq Z \leqq 0.63)$
$= 0.5 + 0.2357 = 0.7357$

(b) $P\,(-1.35 \leqq Z < \infty) = P\,(-1.35 \leqq Z \leqq 0) + P\,(0 \leqq Z < \infty)$
$= P\,(0 \leqq Z \leqq 1.35) + P\,(0 \leqq Z < \infty)$
$= 0.4115 + 0.5 = 0.9115$

正規分布の標準化　確率変数 X が平均 μ，分散 σ^2 の正規分布に従うとき，次のように定めた Z のことを X の標準化変数または基準化変数，あるいは簡単に標準化，基準化という。

$$Z = \frac{X - E\,[X]}{\sigma\,[X]} = \frac{X - \mu}{\sigma}$$

たとえば，X が平均 60，分散 15^2 の正規分布 $N(60, 15^2)$ に従うとき，それの標準化 Z は標準正規分布 $N(0,1)$ に従う。このとき，$P\,(66 \leqq X \leqq 78) = P\,(0.4 \leqq Z \leqq 1.2)$ が成り立つ。

$$\begin{aligned} P\,(66 \leqq X \leqq 78) &= P\left(\frac{66-60}{15} \leqq \frac{X-60}{15} \leqq \frac{78-60}{15}\right) \\ &= P\,(0.4 \leqq Z \leqq 1.2) \\ &= P\,(0 \leqq Z \leqq 1.2) - P\,(0 \leqq Z \leqq 0.4) \\ &= 0.3849 - 0.1554 = 0.2295 \end{aligned}$$

5.2 中心極限定理

確率変数の列 $X_1, X_2, \ldots, X_n$ は互いに独立で同一の分布をもち，$E[X_i] = \mu$，$V[X_i] = \sigma^2$ とする。それらの和を $X = X_1 + X_2 + \cdots + X_n$ とおくと，$E[X] = n\mu$，$V[X] = n\sigma^2$ である。X の標準化を Z とする。

$$Z = \frac{X - E[X]}{\sigma[X]} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

中心極限定理（central limit theorem） 確率変数の列 $X_1, X_2, \ldots, X_n, \ldots$ は互いに独立で同一の分布をもち，$E[X_i] = \mu$，$V[X_i] = \sigma^2$ とすると，次式が成り立つ。これを中心極限定理という。

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \leqq z\right) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

左辺は Z の分布関数の極限，右辺は標準正規分布 $N(0,1)$ の分布関数である。n が十分大きいとき [*iii]，Z は近似的に $N(0,1)$ に，X は近似的に $N(n\mu, n\sigma^2)$ に従う。 *iii

$$Z \sim N(0,1), \qquad X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2) \tag{2}$$

Laplace の定理 中心極限定理を二項分布に適用したものは Laplace の定理とよばれる。$X_1, X_2, \ldots, X_n$ は互いに独立な，成功率 p のベルヌーイ変数とする。

$$E[X_i] = \mu = p, \qquad V[X_i] = \sigma^2 = p(1-p)$$

これらの和 $X = X_1 + X_2 + \cdots + X_n$ は二項分布 $B(n,p)$ に従う。定理により，n が十分大きいとき，X は近似的に $N(n\mu, n\sigma^2)$ すなわち $N(np, np(1-p))$ に従う。

$$X \sim N(np, np(1-p)), \qquad Z = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1) \tag{3}$$

5.3 二項分布の正規近似

1 個のさいころを 180 回投げ，1 の目の出る回数が 24 回以上 39 回以下となる確率を求める。この確率を正確に求めるには次のような計算が必要である。

$${}_{180}\mathrm{C}_{24}\left(\frac{1}{6}\right)^{24}\left(1-\frac{1}{6}\right)^{180-24} + \cdots + {}_{180}\mathrm{C}_{39}\left(\frac{1}{6}\right)^{39}\left(1-\frac{1}{6}\right)^{180-39}$$

これの計算結果は $P(24 \leqq X \leqq 39) = \mathbf{0.8741}$ であるが，コンピュータを利用せずに求めるのは困難である。

[*iii] $n \geq 30$ であれば,「n は十分大きい」と考えてよいとされている。

二項分布の正規分布による近似　1 個のさいころを 180 回投げ，1 の目の出る回数が 24 回以上 39 回以下となる確率を求める。1 の目の出る回数を X とすると，X は試行数 $n = 180$，成功率 $p = 1/6$ の二項分布になる。平均と標準偏差を求め，

$$E[X] = np = 180 \cdot \frac{1}{6} = 30, \quad \sigma[X] = \sqrt{np(1-p)} = \sqrt{180 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 5$$

それをもとに X を標準化し，境界値の 24，39 も同様に計算する。

$$Z = \frac{X - 30}{5}, \qquad \frac{24 - 30}{5} = -1.2, \qquad \frac{39 - 30}{5} = 1.8$$

試行数 n は十分大きいので，Z は近似的に標準正規分布に従う。1 の目の出る回数が 24 回以上 39 回以下となる確率は $\mathbf{0.8490}$ となる。

$$\begin{aligned} P(24 \leqq X \leqq 39) &= P(-1.2 \leqq Z \leqq 1.8) \\ &= P(-1.2 \leqq Z \leqq 0) + P(0 \leqq Z \leqq 1.8) \\ &= P(0 \leqq Z \leqq 1.2) + P(0 \leqq Z \leqq 1.8) \\ &= 0.3849 + 0.4641 = 0.8490 \end{aligned}$$

半整数補正　上の近似では $24 \leqq X \leqq 39$ の境界値として 24 と 39 を用いたが，境界値を 23.5 と 39.5 に置き換えると，近似の精度が向上する。この手法を半整数補正または連続補正という。境界値として 23.5 と 39.5 を用いると，

$$Z = \frac{X - 30}{5}, \qquad \frac{23.5 - 30}{5} = -1.3, \qquad \frac{39.5 - 30}{5} = 1.9$$

1 の目の出る回数が 24 回以上 39 回以下となる確率は $\mathbf{0.8745}$ となる。

$$\begin{aligned} P(23.5 \leqq X \leqq 39.5) &= P(-1.3 \leqq Z \leqq 1.9) \\ &= P(-1.3 \leqq Z \leqq 0) + P(0 \leqq Z \leqq 1.9) \\ &= P(0 \leqq Z \leqq 1.3) + P(0 \leqq Z \leqq 1.9) \\ &= 0.4032 + 0.4713 = 0.8745 \end{aligned}$$

一般には，次のように置き換える（a，b は整数とする）。

$$\begin{aligned} P(a \leqq X \leqq b) &= P\left(\frac{a - 0.5 - E[X]}{\sigma[X]} \leqq Z \leqq \frac{b + 0.5 - E[X]}{\sigma[X]}\right) \\ P(a < X < b) &= P\left(\frac{a + 0.5 - E[X]}{\sigma[X]} \leqq Z \leqq \frac{b - 0.5 - E[X]}{\sigma[X]}\right) \end{aligned}$$

本稿の参考文献

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

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