第 11 回　データ活用基礎

～二項分布と正規分布～

本多俊一

2021 年 11 月 26 日

公立千歳科学技術大学　理工学部　共通教育科

連絡事項

- **レポート課題の不正への対応**
- **【第 08 回 _ レポート課題】の返却＆解説**

Rem. 採点結果に関する異議申し立ては，１週間（12 月 2 日まで）受け付けます．

Rem. 提出方法の不備（ファイル名など）に関しては「３回目以降の不備は１点減点」とします．

イントロダクション

ある人はカードのマークの赤，黒を言いあてられるという.

この能力を試したとき，52 枚中 40 枚をあてた.

「この人は偶然あてたのか ?」

40 の確率論的意味を考える.

中心極限定理で二項分布 $B(52, 1/2)$ を近似すると,

$$\frac{40 - 52 \cdot (1/2)}{\sqrt{52 \cdot (1/2) \cdot (1/2)}} \approx 3.888$$

となる．これは標準正規分布の上側確率 0.0005 の点である.

コンテンツ

1. 二項分布
2. 正規分布
3. レポート課題

1. 二項分布

ベルヌーイ試行とベルヌーイ分布

試行結果が 2 種類の場合について考える.
(e.g. 「コインの表裏」,「くじ引きが当たりか否か」など)

【定義 1】ベルヌーイ試行

ベルヌーイ試行 とは，取り得る結果が「成功」「失敗」の２つのみであり，各試行において成功の確率が一定であるランダム試行である.

成功を $X = 1$, 失敗を $X = 0$ で表し，$P(X = 1) = p$ とする.

１回のベルヌーイ試行で得られる結果の確率分布をベルヌーイ分布と呼び，その確率分布（確率関数）は

$$P(X = 1) = p, \quad P(X = 0) = 1 - p$$

である.

ベルヌーイ試行とベルヌーイ分布

ベルヌーイ分布の期待値と分散を計算すると，

$$E(X) = 1 \times p + 0 \times (1-p) = p \ (=: \mu),$$
$$V(X) = E(X^2) - \mu^2 = 1^2 \times p + 0^2 \times (1-p) - p^2 = p(1-p)$$

となる．

Rem. 成功確率 p のベルヌーイ分布は，後述の二項分布において $n=1$ の場合と同じである．

二項分布

【定義 2】二項分布（binomial distribution）

二項分布 とは，ベルヌーイ試行を n 回行ったときの成功回数を確率変数とする離散型確率分布である.

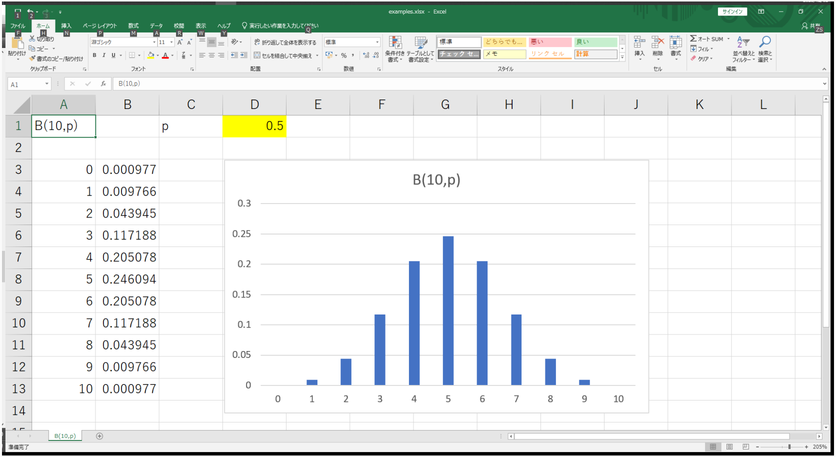
n 回の試行で成功が x 回，失敗が $n-x$ 回であるとすると，確率分布（確率関数）は

$$P(X=x) = {}_n\mathrm{C}_x p^x (1-p)^{n-x}$$

である．${}_n\mathrm{C}_x$ が n 回中 x 回成功する場合の数，$p^x(1-p)^{n-x}$ が成功 x 回と失敗 $n-x$ 回が生じる確率である.

パラメータ (n,p) で確率分布が決まるので $B(n,p)$ と表す.

二項分布の例


B(10,p)
p
0.5
0 0.000977
1 0.009766
2 0.043945
3 0.117188
4 0.205078
5 0.246094
6 0.205078
7 0.117188
8 0.043945
9 0.009766
10 0.000977
B(10,p)
0.3
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6 7 8 9 10

二項分布の期待値と分散

【事実 3】

二項分布 $B(n,p)$ の期待値と分散を計算すると,

$$E(X) = \cdots = np, \quad V(X) = \cdots = np(1-p)$$

となる.

2. 正規分布

正規分布

【定義 4】正規分布（normal distribution）

期待値 μ，分散 σ^2 の**正規分布**とは，確率密度関数が

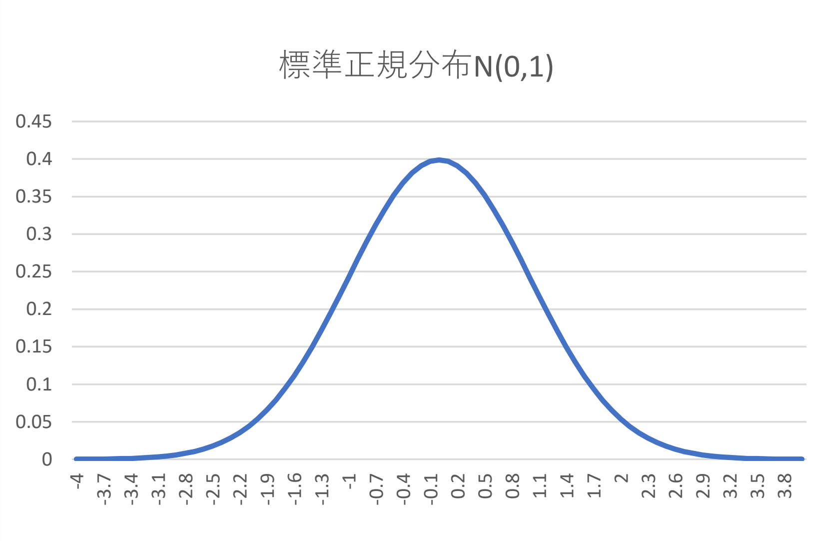
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

で与えられる連続型確率分布である．

正規分布の性質を理解するには，上記関数の広義積分などを考える必要がある．現状では以降の性質を理解しておくとよい．

パラメータ (μ, σ^2) で確率分布が決まるので $N(\mu, \sigma^2)$ と表す．

標準正規分布 $N(0,1)$ のグラフ



正規分布の性質 1

正規分布の性質 1 ：

- 期待値は μ，分散は σ^2 である.
- 単峰，左右対称である.
- $x = \mu$ が最大値，$x = \pm\sigma$ が変曲点に対応する.
- X が
  - $(\mu - \sigma, \mu + \sigma)$ に入る確率は約 68%,
  - $(\mu - 2\sigma, \mu + 2\sigma)$ に入る確率は約 95%,
  - $(\mu - 3\sigma, \mu + 3\sigma)$ に入る確率は約 99.7%

  である.

正規分布の性質２

正規分布の性質２：

- 確率変数 X が正規分布 $N(\mu, \sigma^2)$ に従うとき，X の１次関数 $Y = aX + b$ は正規分布 $N(a\mu + b, a^2\sigma^2)$ に従う．
- 上記の特殊な場合として標準化がある．つまり，

  $$Z = \frac{X - \mu}{\sqrt{\sigma^2}}$$

  と変数変換すると，Z は期待値 0，分散 1 の標準正規分布 $N(0, 1)$ に従う．
- 確率変数 X が二項分布 $B(n, p)$ に従い，n が大きいとき，確率変数 X が従う分布は正規分布 $N(np, np(1-p))$ で近似できる．
  （中心極限定理の特別な場合，cf. 【参考動画】）

標準正規分布 $N(0,1)$ の確率の計算

正規分布は連続型確立分布であるので，確率は確率密度関数の積分によって与えられる．

$\rightarrow$ 計算するのは大変なので，数表やコンピュータを用いる．

$\rightarrow$ スライドでは数表を用いた計算方法を紹介する．

(normal_distribution.xlsx)

標準正規分布 $N(0,1)$ の確率の計算

【例題 5】

確率変数 X が正規分布 $N(0,1)$ に従うとき，次の値を求めよ．

$$P(X \geq 1.96),\ P(X \leq 1.96),\ P(X \leq -1.96),\ P(|X| \leq 1.96).$$

$P(X \geq 1.96)$：

Step1 数表の左側の列から 1.9 を探す．

Step2 横に（上側の 0.06 のところまで）進める．

Step3 その値 0.0250 が $P(X \geq 1.96)$ である．

標準正規分布 $N(0,1)$ の確率の計算

$P(X \leq 1.96)$：

全体の面積が 1 であるから，

$$P(X \leq 1.96) = 1 - P(X \geq 1.96) = 0.9750.$$

$P(X \leq -1.96)$：

正規分布は左右対称であるから，

$$P(X \leq -1.96) = P(X \geq 1.96) = 0.0250.$$

$P(|X| \leq 1.96)$：

正規分布は左右対称であるから，

$$P(|X| \leq 1.96) = 1 - 2 \times P(X \geq 1.96) = 0.950.$$

標準正規分布 $N(0,1)$ の確率の計算

【例題 6】

確率変数 X が正規分布 $N(0,1)$ に従うとき，
$P(X \geq x) = 0.10$ を満たす x を求めよ．

Step1 数表の中で 0.10 に一番近いセルを探す（0.1003）．

Step2 左端の数値 1.2 と上端の 0.08 を足す．

Step3 $P(X \geq 1.28) \approx 0.10$ である．

一般の正規分布 $N(\mu, \sigma^2)$ の確率の計算

確率変数 X が一般の正規分布 $N(\mu, \sigma^2)$ に従うとき，標準化と標準正規分布 $N(0, 1)$ の表を用いて，確率を求めることが出来る.

【例題 7】

確率変数 X が正規分布 $N(50, 10^2)$ に従うとき，$P(X \geq 65)$ の値を求めよ.

標準化変数

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - 50}{10}$$

を考えると，

$$P(X \geq 65) = P(Z \geq 1.5) = 0.0668.$$

二項分布の正規分布近似

【事実 8】

確率変数 X が二項分布 $B(n, p)$ に従い，n が大きいとき，確率変数 X が従う分布は正規分布 $N(np, np(1-p))$ で近似できる（中心極限定理の特別な場合）.

<u>Rem.</u> $(np, np(1-p)) = (\mu, \sigma^2)$ である.

イントロダクション（再掲）

ある人はカードのマークの赤，黒を言いあてられるという．

この能力を試したとき，52 枚中 40 枚をあてた．

「この人は偶然あてたのか？」

40 の確率論的意味を考える．

中心極限定理で二項分布 $B(52, 1/2)$ を近似すると，

$$\frac{40 - 52 \cdot (1/2)}{\sqrt{52 \cdot (1/2) \cdot (1/2)}} \approx 3.888$$

となる．これは標準正規分布の上側確率 0.0005 の点である．

二項分布の正規分布近似

【例題 9】

確率変数 X が二項分布 $B(52, 0.5)$ に従うとき，正規分布近似を用いて $P(X \geq 40)$ を求めよ.

確率変数 X の期待値は $E(X) = 26$，分散は $V(X) = 13$ である．したがって，X の分布は正規分布 $N(26, 13)$ で近似できる．標準化変数

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - 26}{\sqrt{13}}$$

を考えると，

$$P(X \geq 40) \approx P(Z \geq \underline{3.883}) = \underline{0.0001}.$$

3. レポート課題

第 11 回　データ活用基礎　レポート課題

残りの時間は演習時間です．ポータルの指示に従いレポート課題に取り組んで下さい.

補足

1. Zoom：
   メインルーム，質問対応（TASA），グループ学習
2. **授業で紹介していない操作方法（Excel など）は・・・**
   ウェブなどで調べる → 質問するなど・・・
   各自で工夫して下さい.

二項分布 $B(100, 0.2)$ の正規分布近似