第 IX 部　統計的検定

3　標準正規分布

- 標準化
- 正規分布
- 標準正規分布

3.1　はじめに

- いろいろなことは関わり合いを持っています。
- 確実に説明できることもあれば、何となく信じられていることもあります。

3.1.1　ネイピア数

定義

以下で定義される無理数を e であらわし、ネイピア数という。

$$\mathrm{e} = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \tag{1}$$

ネイピア数 e は底とすることが多い。そこで

$$\mathrm{e}^x = \exp(x) \tag{2}$$

とあらわす。

ネイピア数は、

$$\mathrm{e} = 2.718\,281\,828\,459\,045\,235\,36\cdots \tag{3}$$

と続く無理数である。

3.2　確率

確率とは事象の起こりやすさを定量的に示すものである。

3.2.1　標本空間と事象

- 観測されうる結果を『標本点』とよび ω_i であらわす。
- 標本点の全体の集合を『標本空間』とよび Ω であらわす。
- 標本空間の部分集合を『事象』とよび大文字であらわす。
- 標本点を一つも含まないことも事象とみなし『空(くう)事象』とよび $\varnothing$ であらわす。
- 事象 A の起こる確率を $\Pr(A)$ であらわす。

3.2.2　事象の演算

- 事象 A と 事象 B 二つの事象において A と B が共通の標本点を持たない場合 A と B は『排反(はいはん)事象』であるという。また、A と B が共通の標本点を持たないことを A と B は排反であるという。
- 事象 A と事象 B 二つの事象のうち少なくとも一つがおこる事象を A と B の『和(わ)事象』とよび $A \cup B$ であらわす。
- 事象 A と事象 B 二つの事象が同時に起こる事象を『積(せき)事象』とよび $A \cap B$ であらわす。
- 事象 A が起こらない事象を『補(ほ)事象』とよび A^C であらわす。

定義

事象 A と 事象 B とそれらの積事象 $A \cap B$ の確率の関係が、

$$\Pr(A \cap B) = \Pr(A)\ \Pr(B) \tag{4}$$

である場合、事象 A と 事象 B は『独立』であるという。

積事象の確率を二つの事象の確率の積としてあらわすことができるとき二つの事象は独立である。

複数の事象が同時に起こる事象つまり積事象の確率を『同時確率』という。事象が独立な時、同時確率は事象の確率の積としてあらわすことができる。

3.2.3 確率の公理

確率の公理は次の 3 つからなる。これに合うならどのような数も確率と認められる。

1. すべての事象 A に対して
$$0 \leqq \Pr(A) \leqq 1$$

2. 標本空間 Ω 全体の確率は、
$$\Pr(\Omega) = 1$$

3. 互いに排反な事象 $A_1, A_2, A_3, \cdots$ に対して
$$\Pr\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = \Pr\left(A_1\right) + \Pr\left(A_2\right) + \Pr\left(A_3\right) + \cdots$$

3.3　二項分布

定義

ある試行において事象 E の起こる確率を $\Pr(E) = p$、その補事象の確率を $\Pr(E^C) = q$ とする。この試行を独立に n 回繰り返すとき、事象 E の起こる回数を X とすれば X の確率分布は

$$\Pr(X = x) = {}_nC_x p^x q^{n-x} \tag{5}$$

である。この確率分布を確率 p に対する次数 n の二項分布といい、$B(n, p)$ という記号であらわす。

定理 IX−2 − 1

確率 p に対する次数 n の二項分布 $B(n,p)$ の期待値と分散は

$$E(X) = np \tag{6}$$

$$V(X) = npq \tag{7}$$

である。

二項分布 $B(n,p)$ の標準偏差は

$$D(X) = \sqrt{npq} \tag{8}$$

表 1　$n = 8$ の二項分布の期待値と分散と標準偏差

p	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
期待値	7.2	6.4	5.6	4.8	4.0	3.2	2.4	1.6	0.8
分散	0.72	1.28	1.68	1.92	2.00	1.92	1.68	1.28	0.72
標準偏差	0.8485	1.1314	1.2961	1.3856	1.414	1.386	1.2961	1.1314	0.8485

3.4 標準化

定義

確率変数 (X) の期待値からの偏差を標準偏差で割ることを『標準化』といい、変換された値を『z 値』という。

$$z_i = \frac{x_i - E(X)}{D(X)} \tag{9}$$

確率 p に対する次数 $n = 8$ の二項分布 $B(8, p)$ を標準化する。

表 2　$B(8, 0.5)$ の z 値の分布

x	z 値	$\Pr(Z)$
1	−2.828 427	0.003 906
2	−2.121 320	0.031 250
3	−1.414 213	0.109 375
3	−0.707 106	0.218 750
4	0.000 000	0.273 437
5	0.707 106	0.218 750
6	1.414 213	0.109 375
7	2.121 320	0.031 250
8	2.828 427	0.003 906

表 3　$B(8, 0.4)$ の z 値の分布

x	z 値	$\Pr(Z)$
1	$-2.309\,401\,1$	$0.016\,796\,2$
2	$-1.587\,713\,2$	$0.089\,579\,5$
3	$-0.866\,025\,4$	$0.209\,018\,9$
3	$-0.144\,337\,6$	$0.278\,691\,8$
4	$0.577\,350\,3$	$0.232\,243\,2$
5	$1.299\,038\,1$	$0.123\,863\,0$
6	$2.020\,725\,9$	$0.041\,287\,7$
7	$2.742\,413\,8$	$0.007\,864\,3$
8	$3.464\,101\,6$	$0.000\,655\,4$

表 4　$B(8, 0.3)$ の z 値の分布

x	z 値	$\Pr(Z)$
1	$-1.851\,640\,2$	$0.057\,648\,0$
2	$-1.080\,123\,4$	$0.197\,650\,3$
3	$-0.308\,606\,7$	$0.296\,475\,5$
3	$0.462\,910\,0$	$0.254\,121\,8$
4	$1.234\,426\,8$	$0.136\,136\,7$
5	$2.005\,943\,5$	$0.046\,675\,4$
6	$2.777\,460\,3$	$0.010\,001\,9$
7	$3.548\,977\,0$	$0.001\,224\,7$
8	$4.320\,493\,8$	$0.000\,065\,6$

3.4.1 確率を面積とするグラフ

それぞれの z_i を中心に、$z_{i+1} - z_i$ の幅を X 軸上にとり、長方形の面積が対応する確率をあらわすように高さをとる。この高さを $f(Z)$ とあらわす。

z_i の区間は

$$\left[z_i - \frac{(z_{i+1} - z_i)}{2}, z_i + \frac{(z_{i+1} - z_i)}{2}\right] \tag{10}$$

である。

表 5　$B(8, 0.5)$ の z 値の区間の幅と高さ

左端	右端	高さ
−3.181 981	−2.474 874	0.005 524
−2.474 874	−1.767 767	0.044 194
−1.767 767	−1.060 660	0.154 680
−1.060 660	−0.353 553	0.309 359
−0.353 553	0.353 553	0.386 699
0.353 553	1.060 660	0.309 359
1.060 660	1.767 767	0.154 680
1.767 767	2.474 874	0.044 194
2.474 874	3.181 981	0.005 524

図 1 $\boldsymbol{B(8, 0.5)}$ の $\boldsymbol{Z}$ の分布

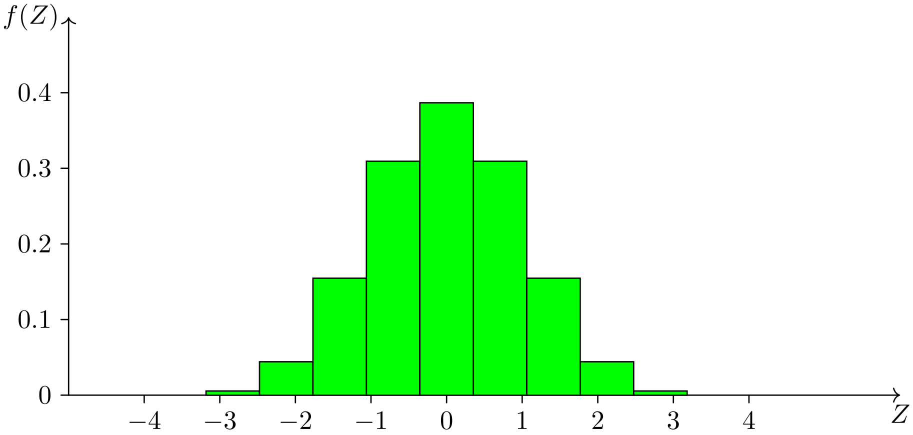


図 2　$\boldsymbol{B(8, 0.4)}$ **の** $\boldsymbol{Z}$ **の分布**

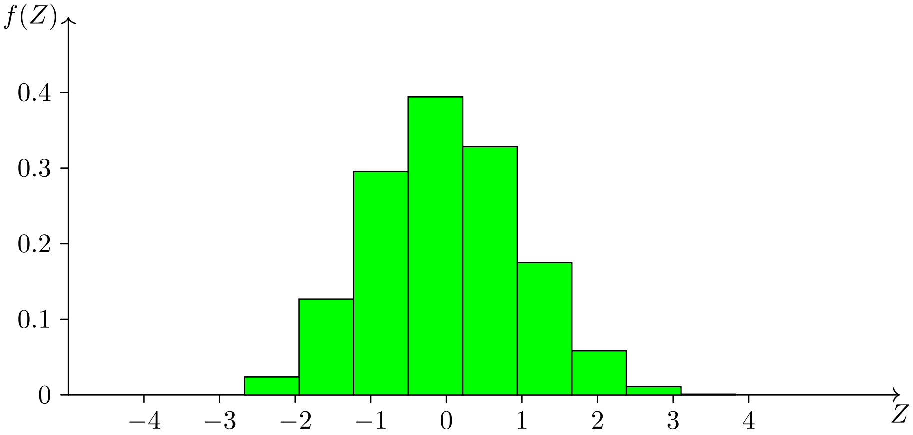
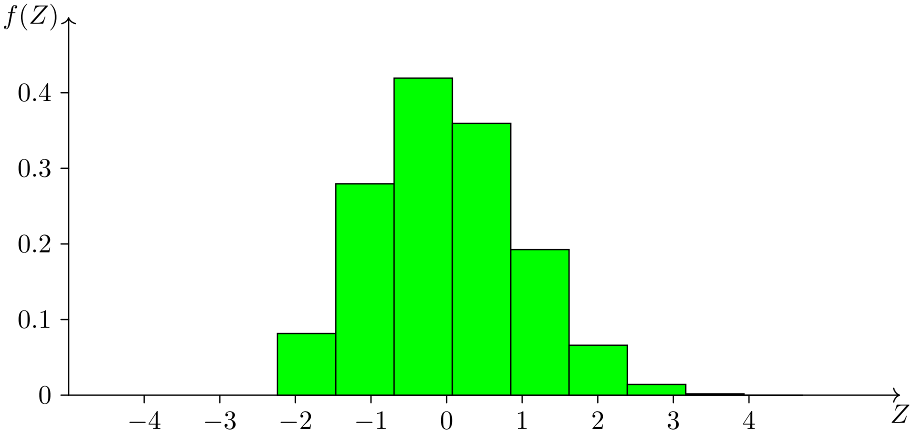


図 3　$B(8, 0.3)$ の Z の分布



3.5 $B(25,\ p)$ の z 値

3.5.1 $B(25, 0.5)$

X が $B(25,\ 0.5)$ に従うとき期待値, 分散, 標準偏差は

$$E(X) = np = 25 \times 0.5 = 12.5 \tag{11}$$

$$V(X) = npq = 25 \times 0.5 \times 0.5 = 6.25 \tag{12}$$

$$D(X) = \sqrt{V(x)} = \sqrt{6.25} = 2.500 \tag{13}$$

である。

従って z 値は

$$z = \frac{x - 12.5}{2.500} \tag{14}$$

である。

3.5.2 $B(25, 0.4)$

$$E(X) = 25 \times 0.4 = 10.0 \tag{15}$$

$$V(X) = 25 \times 0.4 \times 0.6 = 6.00 \tag{16}$$

$$D(X) = \sqrt{6.00} = 2.449 \tag{17}$$

$$z = \frac{x - 10.0}{2.449} \tag{18}$$

3.5.3 $B(25, 0.3)$

$$E(X) = 25 \times 0.3 = 7.5 \tag{19}$$

$$V(X) = 25 \times 0.3 \times 0.7 = 5.25 \tag{20}$$

$$D(X) = \sqrt{5.25} = 2.291 \tag{21}$$

$$z = \frac{x - 7.5}{2.291} \tag{22}$$

図 4 $\boldsymbol{B(25, 0.5)}$ の Z の分布

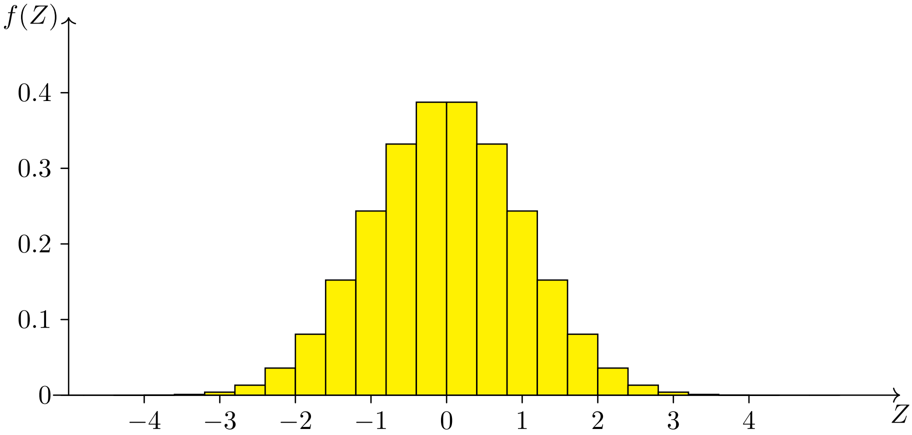


図 5　$B(25, 0.4)$ の Z の分布

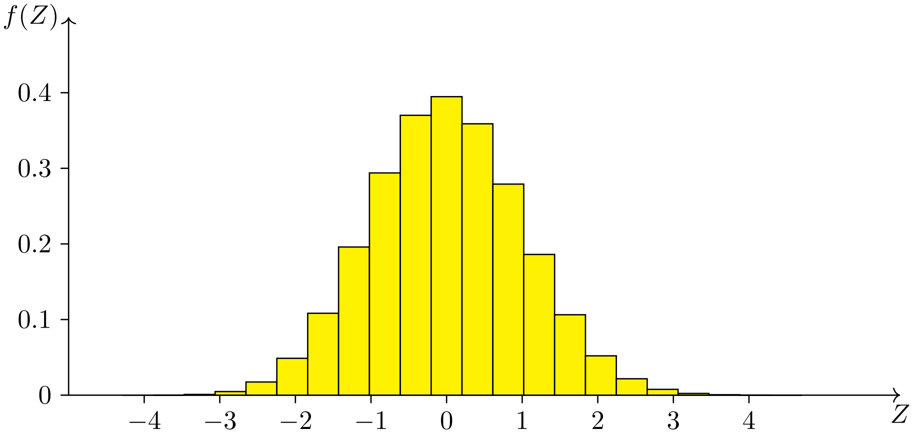
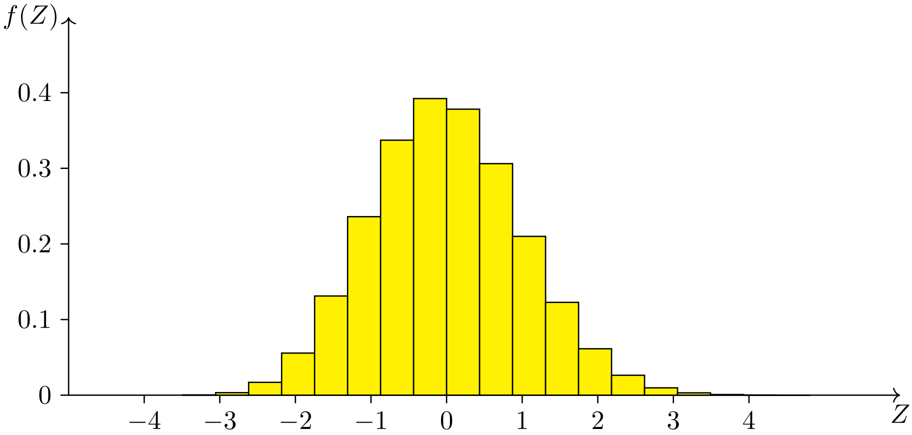


図 6　$B(25, 0.3)$ の Z の分布



3.6 $B(50,p)$ の z 値

同様に $n=50$ とし、$p=0.5, 0.4, 0.3$ のヒストグラムを作成する。

表 6 $B(50,p)$ の期待値と分散

	$p=0.5$	$p=0.4$	$p=0.3$
$E(X)$	25.0	20.0	15.0
$V(X)$	12.50	12.00	10.50
$D(X)$	3.535	3.464	3.240

図 7　$\boldsymbol{B(50, 0.5)}$ の Z の分布

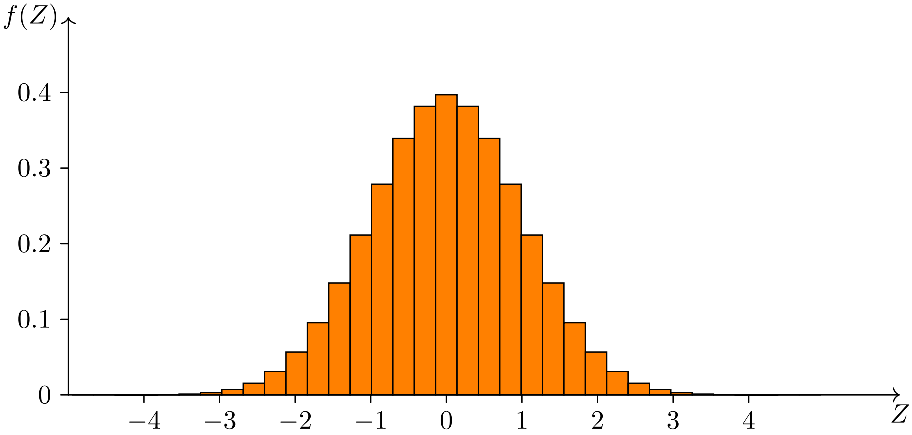


図 8 $\boldsymbol{B(50, 0.4)}$ **の** Z **の分布**

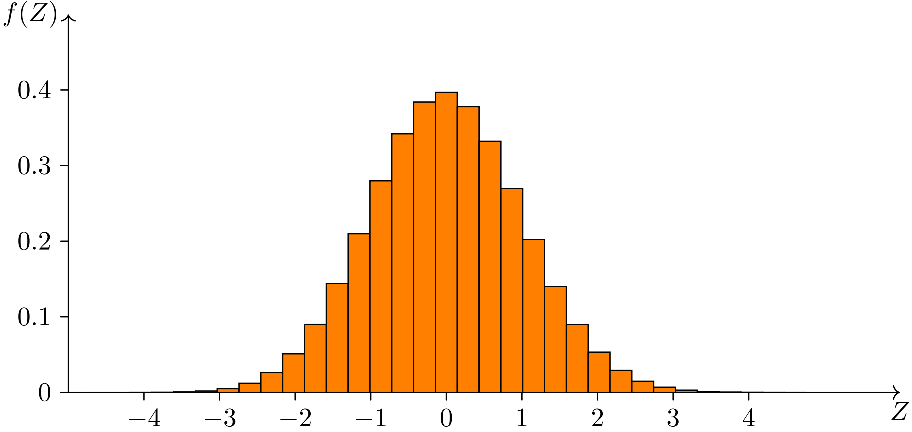
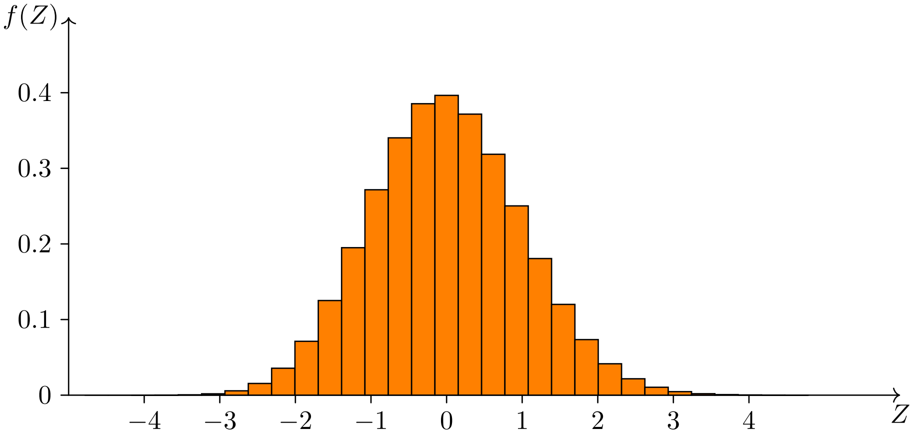


図 9 $\boldsymbol{B(50, 0.3)}$ の Z の分布



3.7 $B(100,\ p)$ の z 値

問題 IX$-\mathbf{3}-\mathbf{2}$

確率変数 X が以下の分布に従うとき、X の期待値, 分散, 標準偏差を求めなさい。

1. $B(100, 0.5)$
2. $B(100, 0.4)$
3. $B(100, 0.3)$

表 7　$\boldsymbol{B(100, p)}$ の期待値と分散

	$p = 0.5$	$p = 0.4$	$p = 0.3$
$E(X)$	50.0	40.0	30.0
$V(X)$	25.00	24.00	21.00
$D(X)$	5.000	4.898	4.582

これらを用いて、$B(100, p)$ の z 値の分布をグラフにあらわす。

図 10　$B(100, 0.5)$ の Z の分布

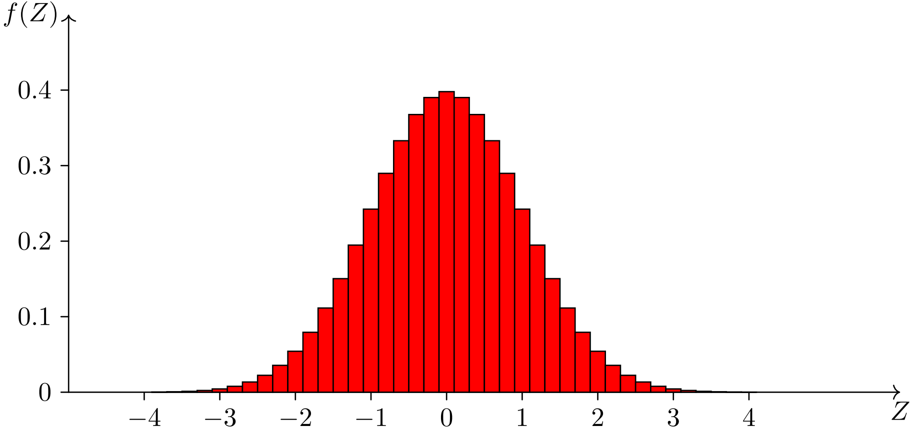


図 11 $B(100, 0.4)$ の Z の分布

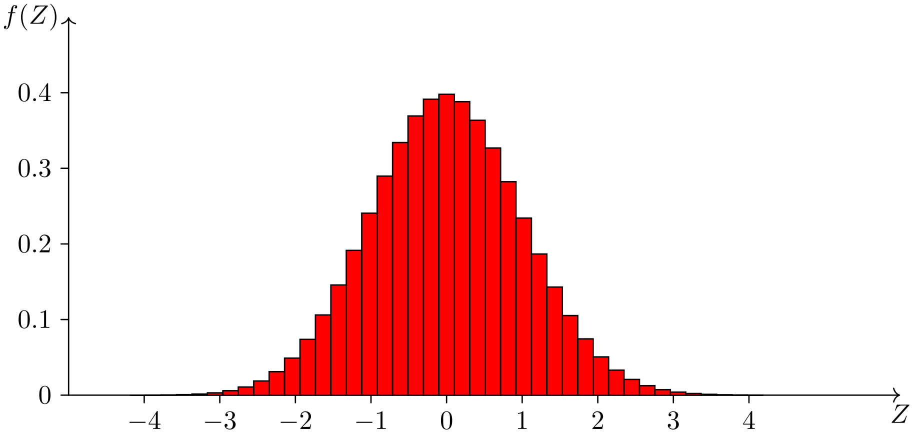
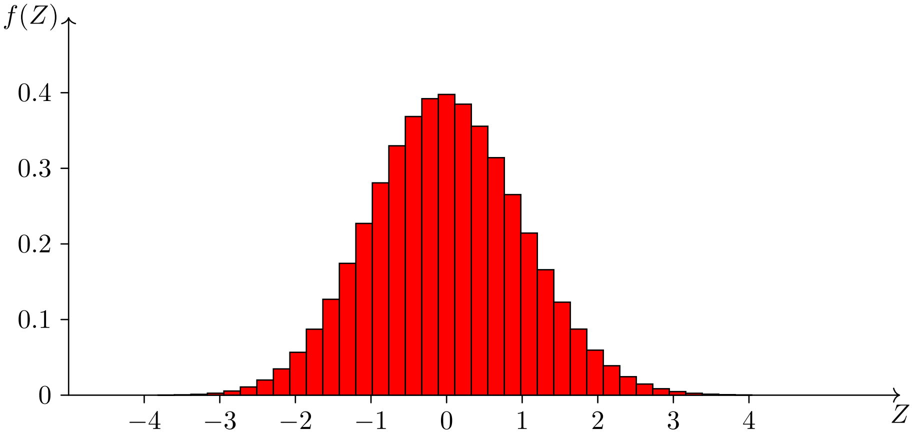


図 12 $\boldsymbol{B(100, 0.3)}$ の Z の分布



問題 IX−**3 − 3**

確率変数 X が以下の二項分布に従うとき、確率変数 X を標準化した変数 Z の分布を以下の仕様に基づき図示しなさい。グラフ作成に関しては何らかのアプリを使うものとする。

- 棒グラフとする。
- 横軸の範囲は

$$-4 \leqq Z \leqq 4$$

とする。

- それぞれの長方形は、

$$\left[z_i - \frac{1}{2\sigma_X}, z_i + \frac{1}{2\sigma_X}\right]$$

を横軸上の幅とし、長方形の面積が

$$\Pr(Z = z_i)$$

となるように高さをとる。

1. $B(100, 0.2)$
2. $B(100, 0.9)$

図 13 $\boldsymbol{B(100, 0.2)}$ の Z の分布

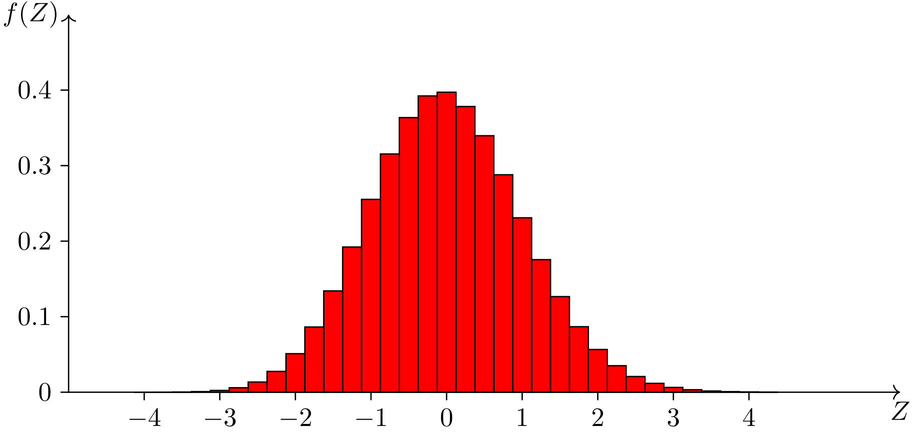
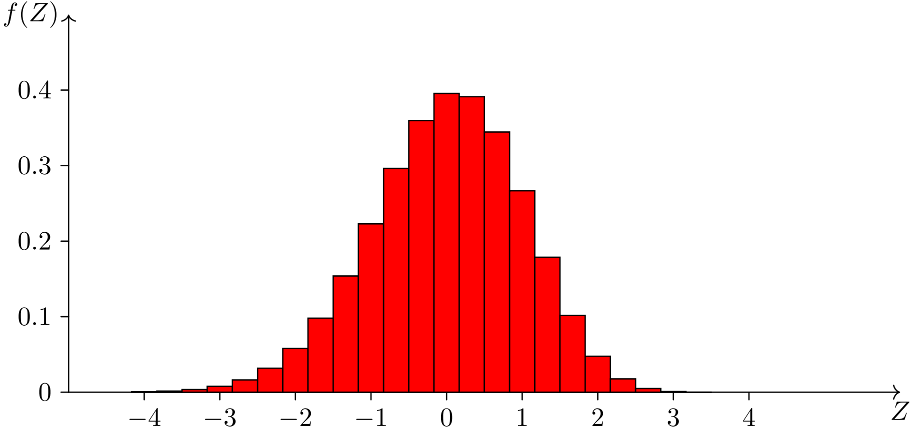


図 14　$B(100, 0.9)$ の Z の分布



3.8 標準正規分布

確率変数 X が二項分布 $B(n,p)$ に従うとき確率変数 X を標準化した変数 Z

$$z_i = \frac{X - np}{\sqrt{np(1-p)}} \tag{23}$$

は n を十分に大きくすると、そのヒストグラムは、左右対称な釣り鐘型の図形に近づいてゆく。

このとき、ヒストグラムの上部を点で結んだ曲線は

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z\right) \tag{24}$$

に近似する。(24) を『正規曲線』という。この正規曲線と z 軸で囲われる範囲の面積は 1 である。正規曲線は $z=0$ の時、最大値

$$f(z) = 0.39894228\cdots \tag{25}$$

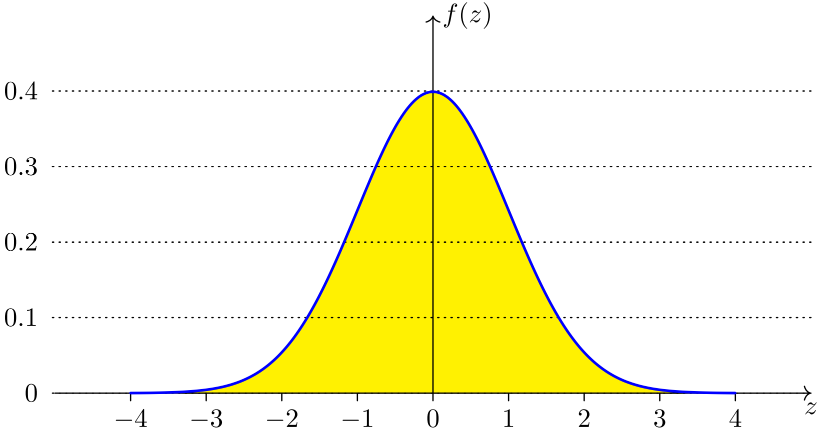
をとる f 軸に対して対称な単峰型の正値関数である。

問題 IX−3 − 4

　次の曲線をグラフにあらわしなさい。グラフ作成に際しては何らかのアプリを使うものとする。

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

図 15　正規曲線



3.8.1 正規分布

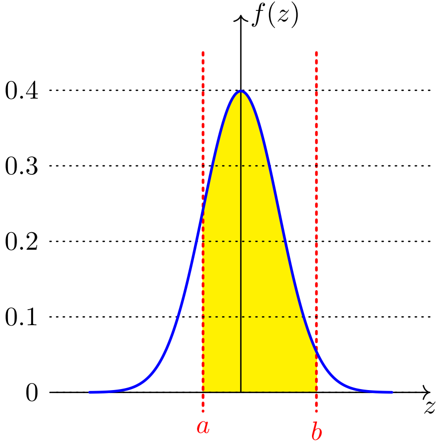
定義

確率変数 Z が a と b の間にある確率 $\Pr(a \leqq Z \leqq b)$ が区間 $[a,\ b]$ において、正規曲線と z 軸に挟まれた領域の面積として与えられるとき、確率変数 Z の分布は『標準正規分布』であるといい

$$N(0,1) \tag{26}$$

であらわす。

図 16　$\Pr\,(a \leqq z \leqq b)$



3.8.2 標準正規分布の特徴

確率変数 Z が標準正規分布に従うとき、以下の特徴がある。

- $\Pr(Z \geqq 1) = 0.158\,655\,254$ である。
- $\Pr(Z \leqq -1) = 0.158\,655\,254$ である。
- $\Pr(-1 \leqq Z \leqq 1) = 0.682\,689\,492$ である。
- $\Pr(Z \geqq 2) = 0.022\,750\,132$ である。
- $\Pr(Z \leqq -2) = 0.022\,750\,132$ である。
- $\Pr(-2 \leqq Z \leqq 2) = 0.954\,499\,736$ である。

図 17　$\mathbf{Pr}\,(-1 \leqq z \leqq 1)$

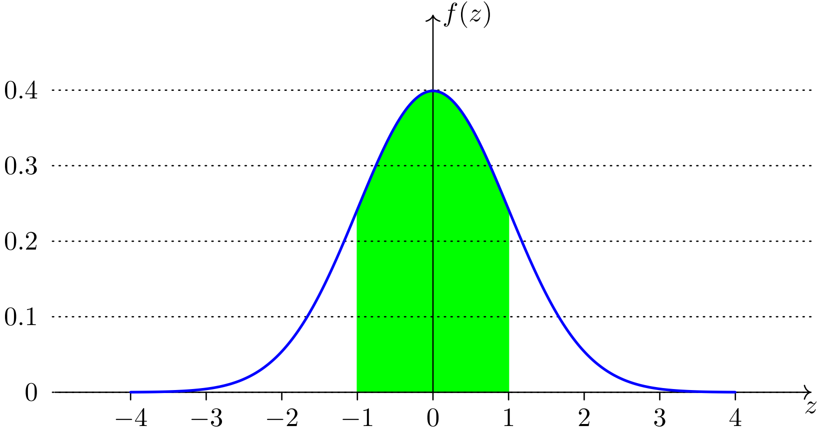
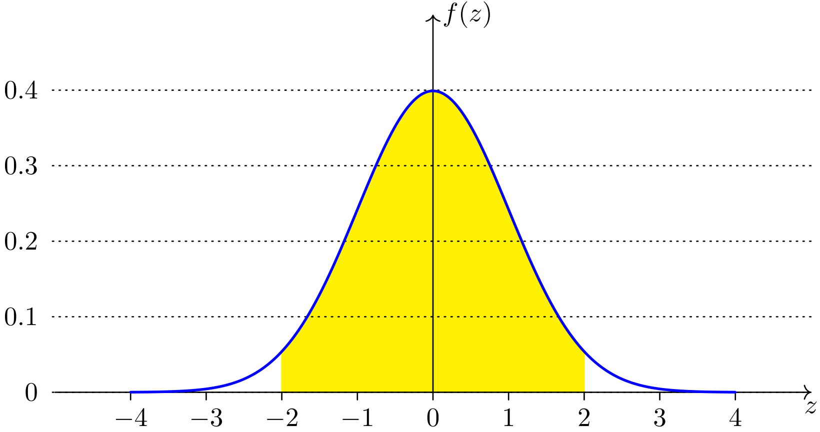


図 18 $\mathbf{Pr}\,(-2 \leqq z \leqq 2)$



問題 IX–3 – 5

確率変数 Z が標準正規分布に従うとき、それぞれの確率を求めなさい。なお、確率の算出に際しては適切な統計表を用いるか何らかのアプリを使用すること。

1. $\Pr(Z \geqq 1.96)$
2. $\Pr(Z \leqq -1.96)$
3. $\Pr(-1.96 \leqq Z \leqq 1.96)$

定義

確率変数 X が標準正規分布 Z を用いて

$$X = m + \sigma Z \quad ; ただし \sigma \neq 0 \tag{27}$$

とあらわせるとき、確率変数 X は平均 m、分散 σ^2 の正規分布であるといい、$N(m, \sigma^2)$ であらわす。

確率変数 X が正規分布 $N(m, \sigma^2)$ に従うとき、X を標準化した確率変数

$$Z = \frac{X - m}{\sigma} \tag{28}$$

は標準正規分布 $N(0, 1)$ に従う。

3.9　まとめ

- 確率変数 $X = x_i$ の期待値 $E(X)$ からの偏差を標準偏差 $D(X)$ で割った値を z 値とよぶ。

$$z_i = \frac{x_i - E(X)}{D(X)}$$

- 確率変数 X から確率変数 Z を求めることを『標準化』という。
- 確率変数 X が $B(n, p)$ に従うとき、X を標準化した確率変数 Z は、n が十分に大きいとき、標準正規分布 $N(0, 1)$ に従う。
- 平均 m、分散 σ^2 の正規分布を $N(m, \sigma^2)$ とあらわす。

3.9.1 参考文献

(1) 「行動科学における統計解析法」芝祐順　南風原朝和　著 『東京大学出版会』1990 年 3 月 20 日　初版

(2) 「完全独習　統計学入門」小島　寛之　著 『ダイヤモンド社』2006 年 9 月 28 日　初版