

# 1 確率分布

## 1.1 確率変数と確率分布

各値に対応して確率が定まるような変数を **確率変数** という。たとえば、2 枚のコインを同時に投げるとき、表のコインの枚数  $X$  は確率変数である。

確率変数  $X$  が  $a$  である確率を  $P(X = a)$ 、 $a$  以上  $b$  以下である確率を  $P(a \leq X \leq b)$  のように表す。たとえば、2 枚のコインを同時に投げるとき、表のコインの枚数を  $X$  とすると  $P(X = 1) = \frac{{}_2C_1}{2^2} = \frac{1}{2}$  である。

確率変数  $X$  のとりうる値  $x_1, x_2, \dots, x_n$  とそれぞれの値をとる確率  $p_1, p_2, \dots, p_n$  の対応関係を **確率分布** といい、以下のような表で書き表す。

$X$	$x_1$	$x_2$	$\cdots$	$x_n$	計
$P$	$p_1$	$p_2$	$\cdots$	$p_n$	1

確率分布が上の表で与えられるとき、 $\sum_{k=1}^n x_k p_k$  を確率変数  $X$  の **期待値** または **平均** といい、 $E(X)$  または  $m$  で表す。

確率変数  $X$  の各値と平均  $m$  の離れ具合を表す確率変数  $(X - m)^2$  の期待値  $E((X - m)^2)$  を確率変数  $X$  の **分散** といい、 $V(X)$  で表す。 $V(X)$  は、 $X^2$  の期待値  $E(X^2)$  と  $X$  の期待値  $E(X)$  を用いて

$$V(X) = E(X^2) - \{E(X)\}^2$$

とも表される。また、分散の正の平方根  $\sqrt{V(X)}$  を確率変数  $X$  の **標準偏差** といい、 $\sigma(X)$  で表す。

## 1.2 確率変数の和や積

互いに独立な確率変数  $X, Y$  に対して

$$E(X + Y) = E(X) + E(Y)$$

$$E(XY) = E(X)E(Y)$$

$$V(X + Y) = V(X) + V(Y)$$

が成り立つ。

## 1.3 二項分布

1 回の試行で事象  $A$  が起こる確率が  $p$  であるとき、この試行を  $n$  回行う反復試行において  $A$  がちょうど  $r$  回起こる確率は

$${}_nC_rp^r q^{n-r} \quad (\text{ただし, } q = 1 - p)$$

である。つまり、 $A$  の起こる回数を  $X$  とすると、確率変数  $X$  の確率分布は以下の表で与えられる。

$X$	0	1	...	$r$	...	$n$	計
$P$	${}_nC_0q^n$	${}_nC_1pq^{n-1}$	...	${}_nC_rp^r q^{n-r}$	...	${}_nC_np^n$	1

このような確率分布を **二項分布** といい、 $B(n, p)$  で表す。確率変数  $X$  が二項分布  $B(n, p)$  に従うとき、その期待値、分散、標準偏差に対して

$$E(X) = np$$

$$V(X) = npq$$

$$\sigma(X) = \sqrt{npq}$$

が成り立つ。

## 1.4 正規分布

連続した値をとる確率変数を **連続型確率変数** という。連続型確率変数の確率分布は表で書き表せないなので、グラフで表す。

常に  $f(x) \geq 0$  で、どんな値  $a, b$  に対しても  $P(a \leq x \leq b) = \int_a^b f(x) dx$  であるような関数  $f(x)$  を **確率密度関数** という。確率密度関数をグラフに表した曲線  $y = f(x)$  を **分布曲線** という。

$m$  を実数,  $\sigma$  を正の実数とするととき, 関数

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

は確率変数  $X$  の確率密度関数となっている。このような確率分布を **正規分布** といい,  $N(m, \sigma^2)$  で表す。確率変数  $X$  が正規分布  $N(m, \sigma^2)$  に従うとき, その期待値, 分散, 標準偏差に対して

$$E(X) = m$$

$$V(X) = \sigma^2$$

$$\sigma(X) = \sigma$$

が成り立つ。

正規分布  $N(0, 1)$  をとくに **標準正規分布** という。確率変数  $X$  が正規分布  $N(m, \sigma^2)$  に従うとき,

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

とおくと, 確率変数  $Z$  は標準正規分布  $N(0, 1)$  に従う。

## 2 統計的な推測

### 2.1 母集団と標本

特性を表す数量を **変量** という。大きさ  $N$  の母集団において、変量  $x$  のとりうる異なる値を

$$x_1, x_2, \dots, x_n$$

とし、それぞれの値をとる個体の個数を

$$f_1, f_2, \dots, f_n$$

とする。この母集団から 1 個の個体を無作為に抽出し、変量  $x$  の値を  $X$  とするとき、確率変数  $X$  の確率分布は以下の表ようになる。

$X$	$x_1$	$x_2$	$\cdots$	$x_n$	計
$P$	$\frac{f_1}{N}$	$\frac{f_2}{N}$	$\cdots$	$\frac{f_n}{N}$	1

この  $X$  の確率分布を **母集団分布** という。また、確率変数  $X$  の期待値、標準偏差をそれぞれ **母平均**、**母標準偏差** といい、 $m$ 、 $\sigma$  で表す。

母集団から大きさ  $n$  の無作為標本を抽出し、それらの変量  $x$  の値を  $X_1, X_2, \dots, X_n$  とするとき、

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

を **標本平均** という。

母平均  $m$ 、母標準偏差  $\sigma$  の母集団から大きさ  $n$  の無作為標本を抽出するとき、その標本平均  $\bar{X}$  の期待値  $E(\bar{X})$  と標準偏差  $\sigma(\bar{X})$  は

$$E(\bar{X}) = m, \quad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

である。

母集団の中で特性  $A$  をもつものの割合を、その特性  $A$  の **母比率** といい、抽出された標本の中で特性  $A$  をもつものの割合を **標本比率** という。

標本の中で特性  $A$  をもつものの個数を考えるには、特性  $A$  をもつとき  $X_k = 1$ 、特性  $A$  をもたないとき  $X_k = 0$  として考えればよい。

$T = \sum_{k=1}^n X_k$  が特性  $A$  をもつものの個数を表す確率変数となっている。

標本の大きさを  $n$ 、母比率を  $p$  とすると、確率変数  $T$  は二項分布  $B(n, p)$  に従う。標本平均  $\bar{X} = \frac{T}{n}$  は特性  $A$  の標本比率  $R$  を表すので、 $n$  が大きいとき、標本比率  $R$  は近似的に正規分布  $N\left(p, \frac{pq}{n}\right)$  に従う。ただし、 $q = 1 - p$  とする。

＊ **標本平均**の標準偏差であるという点に注意する。

## 2.2 推定

母集団から抽出した大きさ  $n$  の無作為標本について平均値が  $\bar{X}$ 、標準偏差が  $S$  であるとき、母平均の信頼度 95% の信頼区間は

$$\left[ \bar{X} - \frac{1.96S}{\sqrt{n}}, \quad \bar{X} + \frac{1.96S}{\sqrt{n}} \right]$$

である。また、その標本比率が  $R$  であるとき、母比率の信頼度 95% の信頼区間は

$$\left[ R - 1.96\sqrt{\frac{R(1-R)}{n}}, \quad R + 1.96\sqrt{\frac{R(1-R)}{n}} \right]$$

である。

＊ 1.96 という値は、正規分布において平均から両側 95% の範囲、すなわち、片側 47.5% の区間を表す。信頼度 90% で考えたければ 1.65、信頼度 99% で考えたければ 2.58 のようにすればよい。

## 2.3 仮説検定

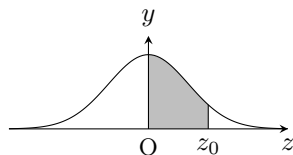
ある仮説が正しいか否かを統計学的に検証することを **仮説検定** という。「特別なことが起こっていない」という仮説  $H_0$  を **帰無仮説** といい、それと対立する「特別なことが起こっている」という仮説  $H_1$  を **対立仮説** という。帰無仮説  $H_0$  が採択されると、「特別なことが起こっていない」という結論に帰着し、帰無仮説  $H_0$  が棄却されると、対立仮説が支持される。

㊦ コインを 400 回投げて、表が 218 回出た。有意水準を 5% として、次のことがいえると判断してよいかをそれぞれ検定せよ。

- (1) このコインは表が出やすい。
- (2) このコインは、表と裏の出方に偏りがある。

# 正規分布表

次の表は、標準正規分布の分布曲線における右図の灰色部分の面積の値をまとめたものである。



$z_0$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1.0	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2.0	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3.0	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900