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 $P(a \leqq X \leqq b)$ 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 $P(X=1)=\dfrac{{}_2\mathrm{C}_1}{2^2}=\dfrac{1}{2}$ である。

確率変数 X のとりうる値 $x_1, x_2, \ldots\ldots, x_n$ とそれぞれの値をとる確率 $p_1, p_2, \ldots\ldots, p_n$ 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，$\displaystyle\sum_{k=1}^{n} x_k p_k$ 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 $E\big((X-m)^2\big)$ 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 $E\big(X^2\big)$ と X の期待値 $E(X)$ を用いて

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

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

$${}_n\mathrm{C}_r p^r q^{n-r} \text{（ただし，} q=1-p\text{）}$$

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	$\cdots$	r	$\cdots$	n	計
P	${}_n\mathrm{C}_0 q^n$	${}_n\mathrm{C}_1 pq^{n-1}$	$\cdots$	${}_n\mathrm{C}_r p^r q^{n-r}$	$\cdots$	${}_n\mathrm{C}_n p^n$	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ので，グラフで表す。

常に $f(x) \geqq 0$ で，どんな値 a，b に対しても $P(a \leqq x \leqq b) = \displaystyle\int_a^b f(x)\,\mathrm{d}x$ 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，σ を正の実数とするとき，関数

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

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，$N\left(m,\,\sigma^2\right)$ で表す。確率変数 X が正規分布 $N\left(m,\,\sigma^2\right)$ 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 $N\left(m,\,\sigma^2\right)$ 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

$$x_1,\ x_2,\ \cdots\cdots,\ x_n$$

とし，それぞれの値をとる個体の個数を

$$f_1,\ f_2,\ \cdots\cdots,\ f_n$$

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のようになる。

X	x_1	x_2	$\cdots$	x_n	計
P	$\dfrac{f_1}{N}$	$\dfrac{f_2}{N}$	$\cdots$	$\dfrac{f_n}{N}$	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，σ 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 $X_1,\ X_2,\ \cdots\cdots,\ X_n$ とするとき，

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

を **標本平均** という。

母平均 m，母標準偏差 σ 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。

$T=\displaystyle\sum_{k=1}^{n} X_k$ 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 $\bar{X}=\dfrac{T}{n}$ 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 $N\!\left(p,\dfrac{pq}{n}\right)$ 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

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

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。

例　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。