講義ノート

1　確率分布の作成

数学 A では, ある事象 A が起こる確率を考え,$P(A)$ のように表した. 今から考えるのは, ある値 X に対しての確率である.

例えば, どの目の出る確率も等しいさいころを 1 個投げる試行を考える. 出た目を X とするとき,$X=k$ である確率を $P(X=k)$ と表す. また,X が a 以上 b 以下である確率を $P(a \leqq X \leqq b)$ と表す. このように, それぞれの値に対し確率が定まるような変数を確率変数という. 確率変数 X と確率の対応関係を確率分布といい,X はこの分布に従うという.

上の試行における X の確率分布は

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

例えば,$P(X=1)=\dfrac{1}{6}$ であり,$P(2 \leqq X \leqq 4)=\dfrac{1}{2}$ である.

例題 1

次の各問いに答えよ.

(1) 表裏の出る確率が等しいコインを 2 枚投げる試行において, 表の出る枚数を X とするとき,X の確率分布を求めよ.

(2) 表裏の出る確率が等しいコインを 3 枚投げる試行において, 表の出る枚数を X とするとき,X の確率分布を求めよ.

解答

(1)

X	0	1	2
P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

(2)

X	0	1	2	3
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$\boxed{2}$ 期待値

確率変数 X が下の分布に従うとする.

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

このとき,

$$\sum_{k=1}^{n} x_k p_k = x_1 p_1 + x_2 p_2 + \cdots\cdots + x_n p_n$$

を X の期待値 (平均) といい,$E(X)$ と表す. つまり,

$$\boxed{E(X) = \sum_{k=1}^{n} x_k p_k}$$

である.

例題 2

さいころ 1 つを 1 回投げる試行において, 出た目の数を X とする.X の期待値を求めよ.

解答

X の確率分布は

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

よって,X の期待値は

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2} \quad \cdots\cdots (答)$$

$\boxed{3}$ 分散と標準偏差

確率変数 X が下の分布に従うとする.

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

このとき,X の分散は $V(X)$ と表し,

$$V(X) = E((X - E(X))^2)$$

で定義される. この式を変形すると, 次の公式が得られる.

$$V(X) = E(X^2) - (E(X))^2$$

X の期待値 (平均) を m とし, まとめると

$$V(X) = \begin{cases} \displaystyle\sum_{k=1}^{n}(x_k - m)^2 p_k = (x_1 - m)^2 p_1 + (x_2 - m)^2 p_2 + \cdots\cdots + (x_n - m)^2 p_n \\ \displaystyle\sum_{k=1}^{n} {x_k}^2 p_k - m^2 = ({x_1}^2 p_1 + {x_2}^2 p_2 + \cdots\cdots + {x_n}^2 p_n) - m^2 \end{cases}$$

分散 $V(X)$ の正の平方根を標準偏差といい $\sigma(X)$ と表す.

$$\boxed{\sigma(X) = \sqrt{V(X)}}$$

例題 3

さいころ 1 つを 1 回投げる試行において, 出た目の数を X とする.X の分散と標準偏差を求めよ.

解答

X の確率分布は

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

X の期待値は

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

よって,X の分散は

$$\left(1 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(3 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(4 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(5 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(6 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} = \frac{35}{12} \quad \cdots\cdots(\text{答})$$

標準偏差は $\sqrt{\dfrac{35}{12}} = \dfrac{\sqrt{105}}{6} \quad \cdots\cdots(\text{答})$

別 分散の公式の利用
改めて確率分布を整理すると

X	1	2	3	4	5	6
X^2	1	4	9	16	25	36
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

よって,X の分散は

$$\left(1\cdot\frac{1}{6}+4\cdot\frac{1}{6}+9\cdot\frac{1}{6}+16\cdot\frac{1}{6}+25\cdot\frac{1}{6}+36\cdot\frac{1}{6}\right)-\left(\frac{7}{2}\right)^2=\frac{35}{12}\quad\cdots\cdots(\text{答})$$

$\boxed{4}$ 変数変換

定数 a, b に対し,

$$\begin{cases} E(aX+b) = aE(X) + b \\ V(aX+b) = a^2V(X) \\ \sigma(aX+b) = |a|\sigma(X) \end{cases}$$

が成り立つ.

例題 4

さいころ 1 つを 1 回投げ, 出た目に応じて得点が得られるゲームを行う. 出た目の数を X とし, 得点は $2X+1$ 点で与えられるとする. 得点を Y とするとき,Y の平均, 分散, 標準偏差を求めよ.

解答

$Y = 2X+1$ である.

X の平均, 分散, 標準偏差は例題 2,3 で求めてあるのでそれを使う.

$$E(X) = \frac{7}{2} \quad , \quad V(X) = \frac{35}{12} \quad , \quad \sigma(X) = \frac{\sqrt{105}}{6}$$

したがって,

$$E(Y) = E(2X+1) = 2E(X) + 1 = 8 \quad \cdots\cdots (答)$$

$$V(Y) = V(2X+1) = 2^2V(X) = \frac{35}{3} \quad \cdots\cdots (答)$$

$$\sigma(Y) = \sigma(2X+1) = 2\sigma(X) = \frac{\sqrt{105}}{3} \quad \cdots\cdots (答)$$

$\boxed{5}$ 複数の確率変数

X, Y を確率変数とするとき $X=a$ かつ $Y=b$ である確率を $P(X=a, Y=b)$ と表す.

X のとり得る値が $x_1, x_2, \cdots\cdots, x_n$,$Y$ のとり得る値が $y_1, y_2, \cdots\cdots, y_m$ であるとき, $X=x_i$ かつ $Y=y_j$ である確率 $P(X=x_i, Y=y_j)$ を考えることができ, この対応を X と Y の同時分布という. 例えば, 大小 2 つのさいころの出る目をそれぞれ X,Y とすれば同時分布は下の表のようになる.

X \ Y	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

また, 和 $X+Y$ の期待値について,

$$\boxed{E(X+Y)=E(X)+E(Y)}$$

が成り立つ. さらに, $\boxed{4}$ 変数変換の等式から $E(aX+bY)=aE(X)+bE(Y)$ が成り立つ.

例題 5

大小 2 つのさいころを 1 回投げ, 出た目に応じて得点が得られるゲームを行う. 大きいさいころの出た目の数を X, 小さいさいころの出た目の数を Y とする.

(1) 得点が $X+Y$ で与えられるとき, 得点の期待値を求めよ.

(2) 得点が $2X+3Y$ で与えられるとき, 得点の期待値を求めよ.

(3) 大きいさいころだけを 10 回投げ,i 回目に出た目 X_i を得点とする. ただし,$1 \leqq i \leqq 10$ である. 総得点 $X_1+X_2+\cdots\cdots+X_{10}$ の期待値を求めよ.

$\boxed{\text{解答}}$

X の期待値 $E(X)$ は例題 2 より $E(X)=\dfrac{7}{2}$ であり, 同様に,$E(Y)=\dfrac{7}{2}$ である.

(1) $E(X+Y)=E(X)+E(Y)=\dfrac{7}{2}+\dfrac{7}{2}=7 \quad \cdots\cdots$(答)

(2) $E(2X+3Y)=E(2X)+E(3Y)=2E(X)+3E(Y)=2\cdot\dfrac{7}{2}+3\cdot\dfrac{7}{2}=\dfrac{35}{2} \quad \cdots\cdots$(答)

(3) 求める期待値は

$$\begin{aligned} E(X_1+X_2+\cdots\cdots+X_{10}) &= E(X_1)+E(X_2)+\cdots\cdots+E(X_{10}) \\ &= \frac{7}{2}+\frac{7}{2}+\cdots\cdots+\frac{7}{2}=\frac{7}{2}\times 10=35 \quad \cdots\cdots(\text{答}) \end{aligned}$$

$\boxed{6}$ 独立と従属の判定

X, Y を確率変数とする.X のとる任意の値 k と Y のとる任意の値 l に対して,

$$\boxed{\text{確率変数}\ X\ \text{と}\ Y\ \text{が独立} \quad \Leftrightarrow \quad P(X=k, Y=l) = P(X=k)P(Y=l)}$$

X と Y が独立でないとき, 従属であるという.

また,2 つの事象 A, B が起こる確率をそれぞれ $P(A), P(B)$ とし,A かつ B が起こる確率を $P(A \cap B)$ とすると,

$$\boxed{\text{事象}\ A\ \text{と}\ B\ \text{が独立} \quad \Leftrightarrow \quad \begin{cases} P_A(B) = P(B) \\ P(A \cap B) = P(A)P(B) \end{cases}}$$

A と B が独立でないとき, 従属であるという.

例題 6

(1) 大小 2 つのさいころを 1 回投げ, 大きいさいころの出た目の数を X, 小さいさいころの出た目の数を Y とする.

① $X = 2$ となる事象と $Y = 3$ となる事象は独立か.

② $X = 2$ となる事象と $X + Y = 5$ となる事象は独立か.

(2) コインを 2 回投げるとき,1 回目に表が出る事象を A,2 回のうち 1 回表が出る事象を B とする.2 つの事象 A と B は独立か.

解答

(1) ① $P(X=2) = \dfrac{1}{6}, P(Y=3) = \dfrac{1}{6}$ である. また,$P(X=2, Y=3) = \dfrac{1}{36}$ である.
したがって,

$$P(X=2, Y=3) = P(X=2) \cdot P(Y=3)$$

が成り立つので独立である　$\cdots\cdots$(答)

② $P(X+Y=5) = \dfrac{4}{36} = \dfrac{1}{9}$ である.

また,$P(X=2, X+Y=5) = P(X=2, Y=3) = \dfrac{1}{36}$ である.
したがって,

$$P(X=2, X+Y=5) \neq P(X=2) \cdot P(X+Y=5)$$

となるから, 独立でない. すなわち従属である.　$\cdots\cdots$(答)

(2) $P(A) = \dfrac{1}{2}, P(B) = {}_2\mathrm{C}_1 \cdot \dfrac{1}{2} \cdot \dfrac{1}{2} = \dfrac{1}{2}$ である. また,$P(A \cap B) = \dfrac{1}{4}$ であるから,

$$P(A \cap B) = P(A) \cdot P(B)$$

が成り立つ. よって,A と B は独立である.　$\cdots\cdots$(答)

$\boxed{7}$ 独立な確率変数

確率変数 X, Y が互いに独立であるとき,

$$\boxed{E(XY) = E(X)E(Y)}$$

が成り立つ.
さらに,

$$\boxed{V(X+Y) = V(X) + V(Y)}$$

が成り立つ. $\boxed{4}$ 変数変換の等式を使えば, $V(aX+bY) = a^2V(X) + b^2V(Y)$ となる.

例題 7

大小 2 つのさいころを 1 回投げ, 出た目に応じて得点が得られるゲームを行う. 大きいさいころの出た目の数を X, 小さいさいころの出た目の数を Y とする.

(1) 得点が XY で与えられるとき, 得点の平均を求めよ.

(2) 得点が $X+Y$ で与えられるとき, 得点の分散を求めよ.

(3) 得点が $2X+3Y$ で与えられるとき, 得点の分散を求めよ.

$\boxed{\text{解答}}$

確率変数 X, Y は独立であり, 平均は例題 2 からともに $\dfrac{7}{2}$, 分散は例題 3 からともに $\dfrac{35}{12}$ である.

(1) $E(XY) = E(X)E(Y) = \dfrac{7}{2} \cdot \dfrac{7}{2} = \dfrac{49}{4} \quad \cdots\cdots$ (答)

(2) $V(X+Y) = V(X) + V(Y) = \dfrac{35}{12} + \dfrac{35}{12} = \dfrac{35}{6} \quad \cdots\cdots$ (答)

(3) X, Y が独立のとき $2X, 3Y$ も独立である. したがって,

$$\begin{aligned} V(2X+3Y) &= V(2X) + V(3Y) \\ &= 2^2V(X) + 3^2V(Y) \\ &= 4 \cdot \frac{35}{12} + 9 \cdot \frac{35}{12} \\ &= \frac{455}{12} \quad \cdots\cdots \text{(答)} \end{aligned}$$

$\boxed{8}$ 二項分布

1 回の試行で事象 A の起こる確率 $P(A)$ を p とおく. この試行を n 回繰り返したとき,A の起こる回数を X とすると, 確率変数 X の分布は

$$P(X=k)={}_n\mathrm{C}_k p^k(1-p)^{n-k}\ (k=0,1,2,\cdots\cdots,n)$$

となる. この確率分布を二項分布といい, **X は二項分布 $B(n,p)$ に従う** という.

例えば, 硬貨を 3 回投げるとき, 表の出る回数を X とすると,

$$P(X=k)={}_3\mathrm{C}_k\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\right)^{3-k}\quad(k=0,1,2,3)$$

であるから,X は二項分布 $B\left(3,\dfrac{1}{2}\right)$ に従う.

X が二項分布 $B(n,p)$ に従うとき, 平均, 分散, 標準偏差は

$$\boxed{\begin{cases}E(X)=np\\ V(X)=np(1-p)\\ \sigma(X)=\sqrt{np(1-p)}\end{cases}}$$

となる.

例題 8

1 個のさいころを 3 回投げて,3 以上の目が出る回数を X とする.X の期待値と分散を求めよ.

解答

1 回の試行において,3 以上の目が出る確率は $\dfrac{2}{3}$ である.

X は二項分布 $B\left(3,\dfrac{2}{3}\right)$ に従うから,

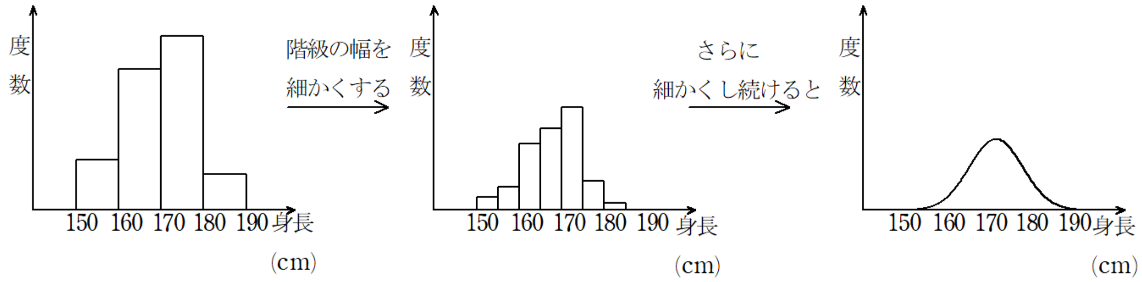
$$E(X)=3\cdot\frac{2}{3}=2\quad\cdots\cdots(答)$$

$$V(X)=3\cdot\frac{2}{3}\cdot\left(1-\frac{2}{3}\right)=\frac{2}{3}\quad\cdots\cdots(答)$$

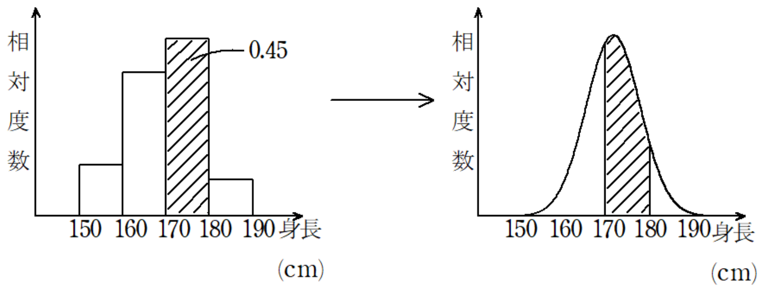
9　連続型確率変数

ここまでの話は, 確率変数がとびとびの値をとるものを考えていた. 確率変数の定義によれば, 身長や体重, 時間などの切れ目のないものを考えることができる.

例えばある集団の身長をヒストグラムにまとめ, 階級の幅を細かくしていくと滑らかな曲線に近づいていく.



縦軸を度数ではなく, 相対度数にすると, その階級に属する人の割合が分かる. そして, その割合の合計は 1 となる. 各階級の長方形の面積が相対度数となるヒストグラムを書き, 上と同様に階級の幅を細かくしていく.



170cm 以上 180cm 未満の階級の相対度数が 0.45 であったとき, 上の右図において, 曲線と 2 直線 $x = 170, x = 180$ および x 軸で囲まれる囲まれる **面積** は 0.45 となり, これは身長が 170cm 以上 180cm 未満である **確率** を表すことになる. この曲線を分布曲線といい, 分布曲線が $y = f(x)$ で表されているとき,$f(x)$ を確率密度関数という.

以上のことから, 確率変数 X の取る値の範囲が $a \leqq X \leqq b$, 確率密度関数を $f(x)$ とするとき,

$$\boxed{\int_a^b f(x)dx = 1}$$

であり,$\alpha \leqq X \leqq \beta$ である確率は,

$$\boxed{P(\alpha \leqq X \leqq \beta) = \int_\alpha^\beta f(x)dx}$$

となる.

例題 9
確率変数 X のとり得る値 x の範囲が $0 \leqq x \leqq 5$ であり, 確率密度関数が $f(x) = kx(0 \leqq x \leqq 5)$ で与えられているとき, 定数 k の値を求めよ. また,$P(1 \leqq x \leqq 3)$ を求めよ.

$\boxed{\text{解答}}$

$\displaystyle\int_0^5 kxdx = k\left[\frac{1}{2}x^2\right]_0^5 = \frac{25}{2}k = 1$ より $k = \dfrac{2}{25}$ $\quad\cdots\cdots$(答)

また,$P(1 \leqq x \leqq 3) = \displaystyle\int_1^3 \frac{2}{25}xdx = \frac{1}{25}\left[x^2\right]_1^3 = \frac{8}{25}\cdots\cdots$(答)

$\boxed{10}$ 連続型確率変数の平均と分散

確率変数 X の取る値の範囲が $a \leqq X \leqq b$, 確率密度関数を $f(x)$ とする.
このとき, 平均 (期待値)$E(X)$ は

$$\boxed{E(X) = \int_a^b xf(x)dx}$$

で定義され, 分散 $V(X)$ はとびとびの値をとる離散型分布のときと同様に,
$V(X) = E((X - E(X))^2)$ で定義される. したがって,

$$\boxed{V(X) = \int_a^b (x - E(X))^2 f(x)dx}$$

となる. また, 公式 $V(X) = E(X^2) - (E(X))^2$ より

$$\boxed{V(X) = \int_a^b x^2 f(x)dx - (E(X))^2}$$

でも求められる.
また, 標準偏差 $\sigma(X) = \sqrt{V(X)}$ で定義される.

例題 10
確率変数 X の確率密度関数が $f(x) = 2x\ (0 \leqq x \leqq 1)$ で与えられているとき,X の期待値 $E(X)$ と分散 $V(X)$ を求めよ.

解答

$$E(X) = \int_0^1 xf(x)dx = \int_0^1 2x^2 dx = \frac{2}{3}\left[x^3\right]_0^1 = \frac{2}{3} \quad \cdots\cdots(答)$$

$$V(X) = \int_0^1 \left(x - \frac{2}{3}\right)^2 f(x)dx = \int_0^1 \left(2x^3 - \frac{8}{3}x^2 + \frac{8}{9}x\right) dx$$

$$= \left[\frac{1}{2}x^4 - \frac{8}{9}x^3 + \frac{4}{9}x^2\right]_0^1 = \frac{1}{18} \quad \cdots\cdots(答)$$

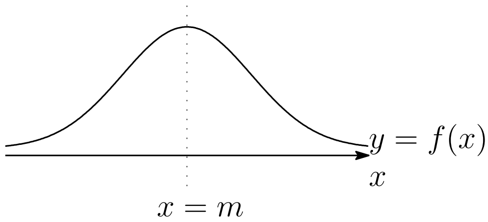
別 分散の公式の利用

$$V(X) = \int_0^1 x^2 f(x)dx - \left(\frac{2}{3}\right)^2 = \int_0^1 2x^3 dx - \frac{4}{9}$$

$$= \frac{1}{2}\left[x^4\right]_0^1 - \frac{4}{9} = \frac{1}{18} \quad \cdots\cdots(答)$$

$\boxed{11}$ 正規分布

確率変数 X の確率密度関数が $f(x)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$ であるとき,



X は正規分布 $N(m,\sigma^2)$ に従うという.

ここで,m は平均,σ^2 は分散 (σ は標準偏差) である.

※ e はネイピア数という特別な定数で,$e=2.71828\cdots\cdots$ である.

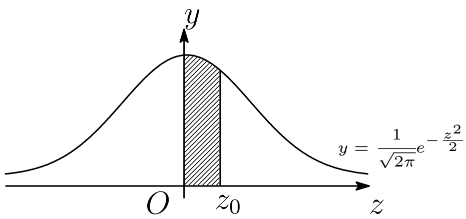
曲線 $y=f(x)$ のグラフは直線 $x=m$ に関して対称となる.

特に,$m=0,\sigma=1$ とした分布を標準正規分布 $N(0,1)$ という. 関数密度関数は $f(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ となり, 曲線 $y=f(x)$ と x 軸,y 軸および y 軸平行な直線で囲まれる面積は表にまとめられている. これを正規分布表という.

また, このグラフが y 軸対称であることもよく利用される.

以下, 正規分布表を利用するために, 変数変換をする.

確率変数 X が正規分布 $N(m,\sigma^2)$ に従うとき,



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

とおくと,Z は標準正規分布 $N(0,1)$ に従う.

正規分布表の一部

z_0	0.00	0.01	$\cdots$
0.0	0.0000	0.0040	$\cdots$
$\vdots$	$\vdots$	$\vdots$	
1.0	0.3413	0.3438	$\cdots$

$P(0\leqq Z\leqq 1)=0.3413$

$P(0\leqq Z\leqq 1.01)=0.3438$

例題 11

　ある集団のテストの得点 X は平均 60 点, 標準偏差 15 点の正規分布に従っていることが分かっている.45 点以上 75 点以下を取った人の割合は何 % か. また,90 点以上取った人の割合を求めよ.

解答

X は正規分布 $N(60,15^2)$ に従う. このとき,$Z=\dfrac{X-60}{15}$ とすると,

Z は標準正規分布 $N(0,1)$ に従う.

$45\leqq X\leqq 75$ のとき,$-1\leqq Z\leqq 1$ であるから,(グラフの対称性を考えて)

$$\begin{aligned}P(45\leqq X\leqq 75)&=P(-1\leqq Z\leqq 1)=2P(0\leqq Z\leqq 1)\\&=2\times 0.3413=0.6826\end{aligned}$$

よって,68.26%　$\cdots\cdots$(答)

また,$X\geqq 90$ のとき,$Z\geqq 2$ であるから,

$$\begin{aligned}P(X\geqq 90)&=P(Z\geqq 2)=0.5-P(0\leqq Z\leqq 2)\\&=0.5-0.4772=0.0228\end{aligned}$$

よって,2.28%　$\cdots\cdots$(答)

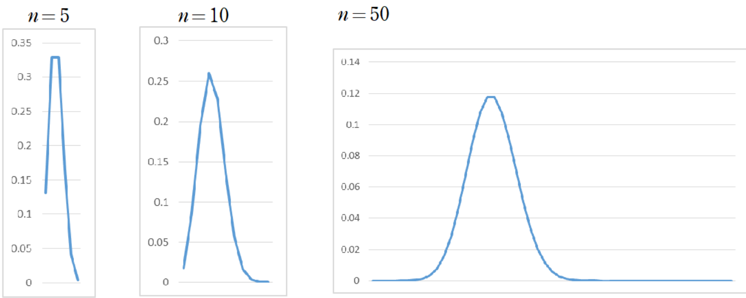
正規分布表の一部

z_0	0.00	$\cdots$
0.0	0.0000	$\cdots$
$\vdots$	$\vdots$	
2.0	0.4772	$\cdots$

$P(0\leqq Z\leqq 2)=0.4772$

$\boxed{12}$　二項分布の正規分布による近似

二項分布 $B(n,p)$ について,$B(5,\frac{1}{3}), B(10,\frac{1}{3}), B(50,\frac{1}{3})$ の分布を折れ線グラフにしてみると,



正規分布の形に近づくことが分かる.$\boxed{8}$ で見たように, 確率変数 X が二項分布 $B(n,p)$ に従うとき,$E(X)=np, V(X)=np(1-p)$ となる.
そして,n が大きいとき **X は正規分布 $N(np, np(1-p))$ に従う**とみなせる.

例題 12
　1 つのさいころを 450 回投げる試行について,3 の倍数が出る回数が 170 回以下である確率の近似値を求めよ.

解答

3 の倍数が X 回出るとすると,X は二項分布 $B\left(450, \dfrac{1}{3}\right)$ に従う.

このとき,$E(X)=450\cdot\dfrac{1}{3}=150, V(X)=450\cdot\dfrac{1}{3}\cdot\left(1-\dfrac{1}{3}\right)=100$ となる.450 回の試行回数は十分多いので,X は正規分布 $N(150,100)$ に従うとみなせる.

$Z=\dfrac{X-150}{\sqrt{100}}$ とすると,Z は標準正規分布 $N(0,1)$ に従う.
よって,

$$\begin{aligned} P(X \leqq 170) &= P(Z \leqq 2) = 0.5 + P(0 \leqq Z \leqq 2) \\ &= 0.5 + 0.4772 = 0.9772 \quad \cdots\cdots (\text{答}) \end{aligned}$$

正規分布表の一部

z_0	0.00	$\cdots$
0.0	0.0000	$\cdots$
$\vdots$	$\vdots$	
2.0	0.4772	$\cdots$

$P(0 \leqq Z \leqq 2)=0.4772$

$\boxed{13}$ 標本平均の平均と標準偏差

ある集団の特徴を調べたい. 集団全員を調査することが難しいとき, 集団から無作為に抽出したものを調査する方法がある. この方法を標本調査という.

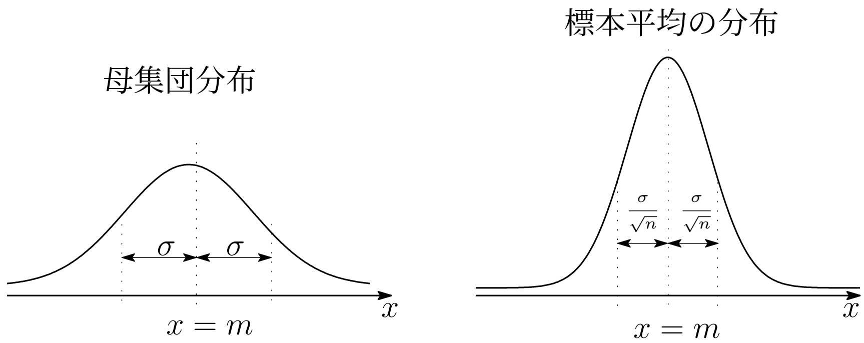
もとの集団を**母集団**といい, 母集団の平均を**母平均**, 標準偏差を**母標準偏差**という. 母集団から n 個の要素を無作為に抽出した集団を**大きさ n の標本**といい, この平均を**標本平均**, 標準偏差を**標本標準偏差**という.

母平均 m, 母標準偏差 σ の母集団から, 大きさ n の標本 X_i $(n = 1, 2, \cdots\cdots, n)$ を抽出したとき, 標本平均を $\overline{X}$ と表す. すなわち,$\overline{X} = \dfrac{X_1 + X_2 + \cdots\cdots + X_n}{n}$ である. 各 X_i は確率変数であり,$\overline{X}$ も標本を抽出するという試行から定まる確率変数である.

標本平均 $\overline{X}$ の平均 $E(\overline{X})$ と標準偏差 $\sigma(\overline{X})$ は,

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

となる.



母平均 m, 母標準偏差 σ の母集団から, 大きさ n の標本を抽出したとき, 標本平均 $\overline{X}$ は $E(\overline{X}) = m, \sigma(\overline{X}) = \dfrac{\sigma}{\sqrt{n}}$ であり,n が大きいとき,

$$\boxed{\overline{X}\ \text{は正規分布}\ \left(m, \frac{\sigma^2}{n}\right)\ \text{に従う}}$$

とみなせる.

また, 母平均 m の集団から大きさ n の無作為標本を抽出するとき, その標本平均は,n が大きくなるにつれて, 母平均 m に近づく. このことを**大数の法則**という.

例題 13

母平均 50, 母標準偏差 10 をもつ母集団から大きさ 100 の標本を抽出する.

(1) 標本平均 $\overline{X}$ の平均 $E(\overline{X})$, 標準偏差 $\sigma(\overline{X})$ を求めよ.

(2) 標本平均が 51 以上である確率を求めよ.

解答

(1) 標本平均の平均は母平均と一致するから $E(\overline{X}) = 50 \quad \cdots\cdots$(答)

また,$\sigma(\overline{X}) = \dfrac{10}{\sqrt{100}} = 1 \quad \cdots\cdots$(答)

(2) $\overline{X}$ は正規分布 $N(50, 1^2)$ に従うとみなせるから,$Z = \dfrac{\overline{X} - 50}{1}$ とおくと,
Z は標準正規分布 $N(0, 1)$ に従う. よって,

$$\begin{aligned} P(\overline{X} \geqq 51) &= P(Z \geqq 1) = 0.5 - P(0 \leqq Z \leqq 1) \\ &= 0.5 - 0.3413 = 0.1587 \quad \cdots\cdots(\text{答}) \end{aligned}$$

正規分布表の一部

z_0	0.00	$\cdots$
0.0	0.0000	$\cdots$
$\vdots$	$\vdots$	
1.0	0.3413	$\cdots$

$P(0 \leqq Z \leqq 1) = 0.3413$

$\boxed{14}$ 母平均の推定

標本平均の平均は母平均と一致するが, 標本平均は確率変数であるから母平均と一致するとは限らない. そこで, 得られた標本平均から母平均が入る区間を推定してみよう.

例えば, 確率変数 Z が標準正規分布 $N(0,1)$ に従うとき,$-1 \leqq Z \leqq 1$ となる確率は正規分布表より 0.6826 である. これはデータの 68.26% が $-1 \leqq Z \leqq 1$ の範囲に入っていることを意味する.

ではデータの 95% が入る範囲を求めよう. それは表より,$-1.96 \leqq Z \leqq 1.96$ である. つまり, この範囲にデータの 95% が含まれる.

母平均 m, 母標準偏差 σ をもつ母集団から大きさ n の標本を抽出すると, $\boxed{13}$ より $\overline{X}$ は平均 m, 標準偏差 $\dfrac{\sigma}{\sqrt{n}}$ となり,n が大きければ正規分布 $N\left(m, \dfrac{\sigma^2}{n}\right)$ に従うとみなせる.

$Z = \dfrac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}}$ とすると,Z は標準正規分布 $N(0,1)$ に従う.

データの 95% は $-1.96 \leqq Z \leqq 1.96$ の範囲にあることから,

$$-1.96 \leqq \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \leqq 1.96$$

$$-1.96 \cdot \frac{\sigma}{\sqrt{n}} \leqq \overline{X} - m \leqq 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

よって,

$$\boxed{\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leqq m \leqq \overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}}$$

となる. この範囲を信頼度 95% の**信頼区間**といい, $\left[\overline{X} - 1.96 \cdot \dfrac{\sigma}{\sqrt{n}}, \overline{X} + 1.96 \cdot \dfrac{\sigma}{\sqrt{n}}\right]$ と表す.

※標本を抽出して信頼区間を求めるという試行を繰り返したとき, 母平均が入っている標本の現れる確率が 95% である, という意味である.

例題 14

320000 人を対象にある試験を行った. 試験の点数は正規分布に従うとする. 標準偏差が 20.0 点であったが, 平均点が公表されなかったため, 無作為に選ばれた 196 人の試験の点数をもとに平均点 m を推定する.196 人の平均点が 60.0 点であったとき, 信頼度 95% の信頼区間を求めよ.

解答

信頼区間は

$$60 - 1.96 \cdot \frac{20}{\sqrt{196}} \leqq m \leqq 60 + 1.96 \cdot \frac{20}{\sqrt{196}} \text{ より,}$$

$57.2 \leqq m \leqq 62.8$ $\quad\cdots\cdots$(答)

$\boxed{15}$　母比率の推定

母集団が, ある特性 A をもっているものともっていないものからなると考える. 母集団の中で特性 A をもつものの割合を**母比率**という. 母集団から大きさ n の標本を抽出し, その中で特性 A をもっているものの割合を**標本比率**という. 今回は標本比率 p' から母比率 p の推定を目指す.

母集団から大きさ n の標本を抽出し, この標本の中で特性 A をもつものが X 個あったとき, 標本比率 p' は $p' = \dfrac{X}{n}$ である. また,n 個中 X 個, 確率 p で特性 A をもつと考えられるから,X は二項分布 $B(n, p)$ に従うとみなせる. したがって,n が大きいとき X は近似的に正規分布 $N(np, np(1-p))$ に従う. よって,$Z = \dfrac{X - np}{\sqrt{np(1-p)}}$ とすると Z は標準正規分布 $N(0, 1)$ に従うから, 信頼度 95% の信頼区間は, $-1.96 \leqq Z \leqq 1.96$ より,

$$-1.96 \leqq \frac{X - np}{\sqrt{np(1-p)}} \leqq 1.96$$

$$-1.96 \cdot \frac{1}{n} \leqq \frac{\frac{X}{n} - p}{\sqrt{np(1-p)}} \leqq 1.96 \cdot \frac{1}{n}$$

$$-1.96 \cdot \frac{\sqrt{np(1-p)}}{n} \leqq \frac{X}{n} - p \leqq 1.96 \cdot \frac{\sqrt{np(1-p)}}{n}$$

$$-1.96 \cdot \sqrt{\frac{p(1-p)}{n}} \leqq \frac{X}{n} - p \leqq 1.96 \cdot \sqrt{\frac{p(1-p)}{n}}$$

$$\frac{X}{n} - 1.96 \cdot \sqrt{\frac{p(1-p)}{n}} \leqq p \leqq \frac{X}{n} + 1.96 \cdot \sqrt{\frac{p(1-p)}{n}}$$

$\dfrac{X}{n} = p'$ であるから,

$$p' - 1.96 \cdot \sqrt{\frac{p(1-p)}{n}} \leqq p \leqq p' + 1.96 \cdot \sqrt{\frac{p(1-p)}{n}}$$

n が大きいとき, 標本比率と母比率は等しいと考えられるから,$p = p'$ とすると,

$$\boxed{p' - 1.96 \cdot \sqrt{\frac{p'(1-p')}{n}} \leqq p \leqq p' + 1.96 \cdot \sqrt{\frac{p'(1-p')}{n}}}$$

となる.

例題 15

ある工場で生産される製品の不良品率を調べるため, 無作為に選んだ 400 個の製品を調査した. 不良品が 8 個見つかったとき, 不良品率 p を信頼度 95% で推定せよ.

$\boxed{\text{解答}}$

標本比率は $\dfrac{8}{400} = \dfrac{1}{50}$ であるから,

$$1.96 \cdot \sqrt{\frac{\frac{1}{50} \cdot \left(1 - \frac{1}{50}\right)}{400}} = 0.01372$$

よって, 求める信頼区間は $0.02 - 0.01372 \leqq p \leqq 0.02 + 0.01372$ より,

$0.00628 \leqq p \leqq 0.03372$　$\cdots\cdots$(答)

Appendix　仮説検定

ある硬貨を3回投げたとき,3回とも表が出たとする. この硬貨は表と裏の出る確率は $\dfrac{1}{2}$ といえるのだろうか. このことを仮説を立てて検証することを仮説検定という.

今, 考えられる結論は2通りで,

· 結論1:硬貨の表裏の出る確率は等しく $\dfrac{1}{2}$

· 結論2:表の方が出やすい

この場合, 結論2であること疑っている, つまり主張したい. これを否定した仮説「硬貨の表裏の出る確率は等しく $\dfrac{1}{2}$ (結論1)」を**帰無仮説**といい H_0 と表す. 一方で主張したい仮説 (結論2) を**対立仮説**といい H_1 と表す. 実際に仮説を立てる.
表の出る確率を p とすると,

$\mathrm{H}_0 : p = \dfrac{1}{2}$

$\mathrm{H}_1 : p > \dfrac{1}{2}$

H_0 のもとで3回とも表が出る確率は

$$\left(\frac{1}{2}\right)^3 = 0.125$$

となる. これは稀なことなのかどうかを判断する基準が必要になる. この基準を**有意水準**, または危険率と呼ぶ. 例えば有意水準5% で検定すると,

$0.125 > 0.05$

であるから, めったに起こらないとは判断できない. したがって,H_0 を否定することができない. このような場合,H_0 を**受容する**, または**採択する**という.

では, この硬貨を10回投げ,9回表が出た場合はどうだろうか. 仮説は上と同じで,

$\mathrm{H}_0 : p = \dfrac{1}{2}$

$\mathrm{H}_1 : p > \dfrac{1}{2}$

この場合は, 硬貨を10回投げ9回以上表出る確率を計算する. すると,

$${}_{10}\mathrm{C}_9 \left(\frac{1}{2}\right)^9 \cdot \frac{1}{2} + \left(\frac{1}{2}\right)^{10} = 0.0107421875$$

となり0.05より小さいので, 非常に珍しいことが起こったと判断できる. このとき,H_0 を**棄却する**といい,H_1 を採択するという.

※確率が小さくても起こる確率は0でないので,$p = \dfrac{1}{2}$ であったとしてもたまたま珍しいことが起こってしまった可能性もある. つまり,H_0 が正しかったとしても, 棄却される危険がある. 危険率5% とはそういった意味である.

この硬貨を 64 回投げて表が 40 回出た場合, この硬貨がいかさまかどうかを有意水準 5% で検定しよう.
表の出る確率を p とする. この場合の仮説は,

$$\mathrm{H}_0 : p = \frac{1}{2}$$

$$\mathrm{H}_1 : p \neq \frac{1}{2}$$

として考える.
この検定の場合, 極端に大きい値と極端に小さい値の合計が 5% になる点を考える. つまり, 上から 2.5% までと下から 2.5% までの間は, 珍しいことが起こったと考えることにする.
表の出る回数 X は二項分布 $B\left(64, \dfrac{1}{2}\right)$ に従う. これは近似的に正規分布 $N(32, 16)$ に従うとみなせるから,$Z = \dfrac{X-32}{4}$ とおくと,

$$P(X \geqq 40) = P(Z \geqq 2) = 0.5 - P(0 \leqq Z \leqq 2) = 0.0228$$

これは 0.025 より小さいので, 稀なことが起こったと判断できる. したがって,H_0 を棄却し,H_1 を採択する.

また, これまで見てきた通り, 正規分布表より $-1.96 \leqq Z \leqq 1.96$ を満たす確率が 95% である. つまり,$|Z| \geqq 1.96$ を満たすとき,H_0 は棄却される. この範囲を**棄却域**という.