P.70~P.72

♠ 代表的な分布としての二項分布

定義

成功確率を p とし，成功か失敗かの二者択一の独立な試行を n 回繰り返し，成功回数を X とする．X の従う分布を**二項分布**といい，$B(n,p)$ と書く．また，X が二項分布 $B(n,p)$ に従うことを，

$$X \sim B(n,p)$$

と書く．

補足

成功確率 p, 試行回数 n の場合，成功回数 X の従う分布は，

X	0	1	$\dots$	k	$\dots$	n	計
P	${}_n\mathrm{C}_0 q^n$	${}_n\mathrm{C}_1 pq^{n-1}$	$\dots$	${}_n\mathrm{C}_k p^k q^{n-k}$	$\dots$	${}_n\mathrm{C}_n p^n$	1

練習 17

1 個のサイコロを 5 回投げて，2 以下の目が出る回数を X とする．X はどのような二項分布に従うか．また，次の確率を求めよ．

(3) $P(2 \leqq X \leqq 4)$

解答

ヒント　5 回中 2 回成功 (2 以下の目が出る) する確率は，

$${}_5\mathrm{C}_2 \underbrace{\left(\frac{1}{3}\right)^2}_{\text{成功確率}} \underbrace{\left(\frac{2}{3}\right)^3}_{\text{失敗確率}}$$

♣ 二項分布の期待値と分散

全体課題

あなたはプロのギャンブラーです．2 つの異なるルーレットゲーム（A，B）があり，どちらも **1 日に「平均（期待値）20 回」の当たり**が見込めると分析しています．

あなたは「平均勝利数」だけでなく，「**収支の安定性**（＝大勝ちや大負けが少なく，平均値前後に収まること，すなわち**バラツキが小さいこと**）」も重視しています．

- **ゲーム A**:
  - 2 択のルーレット（赤か黒か）に賭ける．
  - 1 ゲームの勝率は $p_A = 0.5$.
  - 1 日に $n_A = 40$ 回 プレイする．
  - 1 日の当たり回数を X とする．
- **ゲーム B**:
  - 20 択のルーレット（特定の数字）に賭ける．
  - 1 ゲームの勝率は $p_B = 0.05$.
  - 1 日に $n_B = 400$ 回 プレイする．
  - 1 日の当たり回数を Y とする．

解答 (pre)

解答 (post)

♠ エキスパート A「二項分布の選別」

目標 A

確率変数 X の従う分布が，二項分布であるかそうではないかを見極めることができる．また，その根拠を説明できる．

例

- 1 枚のゆがみのないコインを 100 回投げ，表の出た回数を X とする． このとき，X の従う分布は
【 二項分布_____________である / 二項分布ではない 】．
- 1 個のサイコロを 10 回投げるとき，1 の目がでる回数 X とする．このとき，X の従う分布は
【 二項分布_____________である / 二項分布ではない 】．
- 赤玉 6 個, 白玉 4 個が入った袋から玉を一つずつ元に戻さず 3 回取り出し，取り出された赤玉の個数を X とする．このとき，X の従う分布 [a]は
【 二項分布_____________である / 二項分布ではない 】．
- 赤玉 6 個, 白玉 4 個が入った袋から玉を一つ取り出し元に戻すという試行を 3 回繰り返し，取り出された赤玉の個数を X とする．このとき，X の従う分布は
【 二項分布_____________である / 二項分布ではない 】．
- 1 個のサイコロを，1 の目が出るまで繰り返し投げ，投げた回数を X とする． このとき，X の従う分布[b]は
【 二項分布_____________である / 二項分布ではない 】．
- 不良品率が 3% である大量の部品の山から，10 個の 1 個ずつ部品を取り出し[c]，不良品の個数を X とする．このとき，X の従う分布は
【 二項分布_____________である / 二項分布ではない 】．

[a] このような分布を超幾何分布という．
[b] このような分布を幾何分布と呼ぶ．
[c] 1 つ取り出すごとに母集団の個数が 1 つ減るから厳密には反復試行とは言えない．しかし，母集団の総数が十分多い場合，各試行の確率はほとんど変わらないため，反復試行 (独立な試行) とみなしてよい．

まとめ

二項分布とは，二者択一の「独立」な試行 (反復試行) が「同じ条件」の下で繰り返されたときに，「成功回数」の従う分布のことである．ただし，母集団の大きさが十分大きい際には，各試行は独立とみなすことができる．

♠ エキスパート B「困難は分割せよ！」

目標 B

二項分布 $B(n,p)$ に従う確率変数 X の期待値 $E(X)$, 分散 $V(X)$, 標準偏差 $\sigma(X)$ がそれぞれ

$$E(X)=np,\quad V(X)=npq,\quad \sigma(X)=\sqrt{npq} \qquad (\text{ただし } q=1-p)$$

と書けることを理解し説明できる．

説明

- **直接計算したい?**
「定義通り」計算しようとすると, 全頁「補足」から,
$E(X)=\sum_{k=0}^{n} k\times P(X=k)=\sum_{k=1}^{n} k\times {}_n\mathrm{C}_k p^k(1-p)^{n-k}$ となり，計算が困難.
- **困難は分割せよ!!**
『n 回の合計成功回数 X』を，『**1 回ごとの試行**』に分解する．
  - (例) $n=5$ 回のフリースローで，「成功，失敗，成功，失敗，成功」→ $X=3$ 回
  - 「i 回目（$i=1$,2 ,3, 4, 5）の試行」だけに着目し，その結果を「1 か 0 か」の点数で表す変数を X_i とおく．
→「$X_1=1$, $X_2=0$, $X_3=1$, $X_4=0$, $X_5=1$」
  - 全体を部分で表せる．→ $X=X_1+X_2+X_3+X_4+X_5=3$.
  - このような, 量的データ「成功か失敗か」を質的データ「1 か 0 か」へ置き換える変数 X_i のことをダミー変数と呼ぶ.
- **ダミー変数の期待値は** p
「1 回だけの試行 X_i」の確率分布は

X_i	0	1	計
P	$1-p$	p	1

(p は各回の成功確率.)
  - 期待値は $E(X_i)=0\times(1-p)+1\times p=p$
  - 「1 回だけの試行 X_i」なら期待値を簡単に計算でき，その値は成功確率 p に等しい．
- **小まとめ**
ダミー変数を用いたことで,「$X=X_1+X_2+\cdots+X_n$」と「$E(X_i)=p$」が得られた．
- **期待値には加法性がある.**
さらに，前回の授業で期待値の加法性 $E(X+Y)=E(X)+E(Y)$ を示した．
- **困難は分割せよ!!**
分割した困難をまとめることで，次のように結論を得る．

$$\begin{aligned} E(X) &= E(X_1+X_2+\cdots+X_n) \\ &= E(X_1)+E(X_2)+\cdots+E(X_n) \\ &= \underbrace{p+p+\cdots+p}_{n\text{ 回}}=np. \end{aligned}$$

- 分散 $V(X)$ についてもダミー変数を用いて計算すれば $V(X)=npq$ を得られる．(今回は省略.)

まとめ

ダミー変数を用いて，全体の「成功回数」を「1 回ごと」の成功回数 (1 か 0 か) の和とみて計算した．
困難は分割せよ[a]!!

[a] これは数学者ルネ・デカルトの言葉である．

性質

2 つの確率変数 X, Y が互いに独立であるとき,

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

証明

使える事実

$$V(X+Y)=\underbrace{E((X+Y)^2)-(E(X+Y))^2}_{(2\text{ 乗の期待値})-(\text{期待値の 2 乗})}$$

$$E(XY)=E(X)E(Y)\quad(\leftarrow\text{ 独立なら})$$

証明

上の事実を使うと, X と Y が独立なら,

$$\begin{aligned}
V(X+Y)&=E((X+Y)^2)-(E(X+Y))^2\\
&=\underbrace{E(X^2+2XY+Y^2)}_{\text{中身を展開}}-\underbrace{(E(X)+E(Y))^2}_{\text{加法性 } E(X+Y)=E(X)+E(Y)}\\
&=\underbrace{E(X^2)+2E(XY)+E(Y^2)}_{\text{期待値の加法性}}\\
&\qquad-(E(X))^2-2E(X)E(Y)+(E(Y))^2\\
&=\underbrace{E(X^2)-(E(X))^2}_{=V(X)}+\underbrace{E(Y^2)-(E(Y))^2}_{=V(Y)}\\
&\quad+2\underbrace{\{E(XY)-E(X)E(Y)\}}_{\text{上の事実から}=0}\\
&=V(X)+V(Y).
\end{aligned}$$