5 代表的な確率分布

5.1 事象の独立性

(1) 事象 A と B が独立であるとは、$P(A\cap B)=P(A)P(B)$ が成り立つことである。
(2) 事象 $A_1, A_2, \ldots, A_n$ が独立であるとは、n 個の中から重複を許さずに任意個 $(=m)$ 選んで、それらを

$$B_1, B_2, \ldots, B_m$$

と表したとき、$P(B_1\cap B_2\cap\cdots\cap B_m)=P(B_1)P(B_2)\ldots P(B_m)$ が成り立つことである。

例 5.1. **(待ち時間と ATM の数)** 人が ATM を 2 分以上にわたって使用する確率は 0.37 であるとする。P 君が銀行に入ったところ、ATM は 10 台あり、全部使用中であった。P 君が待ち行列の先頭であるとき、P 君が 2 分以上待つ確率は幾らか。

ATM が 1 台ならば、P 君が 2 分以上待つ確率は明らかに 0.37 である。ATM が 10 台のときは、$A_i=\{$ 第 i 番目の ATM で 2 分以上かかる $\}$ と置くと、$P(A_i)=0.37$ でかつ、

$$\begin{aligned} P(\{2\text{ 分以上待つ }\}) &= P(A_1\cap A_2\cap\cdots\cap A_{10})=P(A_1)\ P(A_2)\ \ldots P(A_{10}) &(1)\\ &= (0.37)^{10}=0.0000481. &(2)\end{aligned}$$

ATM が n 台ならば 0.37 の n 乗となる。0.37 の計算法は指数分布のところで勉強する。

例 5.2. **(幾何分布入門)** 表が出る確率が p であるような歪んだコインを続けて投げるとする。n 回目に初めて表が出る確率を求める。

$$A_i=\{i\text{ 回目に表が出る }\} \qquad (3)$$

と置くと、各 $A_1,\ldots,A_n$ は互いに独立であり、$P(A_i)=p$ かつ $P(A_i^c)=1-p$ であることは明らかである。

$$\begin{aligned} P(\{n\text{ 回目に初めて表が出る }\}) &= P(A_1^c\cap A_2^c\cap\cdots\cap A_{n-1}^c\cap A_n) &(4)\\ &= P(A_1^c)\ P(A_2^c)\ \times\cdots\times P(A_{n-1}^c)\ P(A_n) &(5)\\ &= (1-p)^{n-1}p &(6)\end{aligned}$$

5.2 離散型分布：2 項分布、ポアソン分布、幾何分布

定理 1. **(2 項分布 $B(n,p)$ の導出)** 表が出る確率が p のコインを n 回投げる試行において、$X=$「n 回中の表 (成功) の回数」とおくと、X は 2 項分布 $B(n,p)$ に従う。即ち、

$$P(X=x)={}_nC_xp^x(1-p)^{n-x}\quad (x=0,1,\cdots,n).$$

(証明) 各回の試行は独立であることに注意しよう。

$$E_i=\Big\{i\text{ 回目に表が出る}\Big\}\ \ (i=1,\cdots,n)$$

とおく。

$$\begin{aligned} P(X=0) &= P\Big(\{\text{ 全て裏 }\}\Big)=P\Big(E_1^c\cap E_2^c\cap\cdots\cap E_n^c\Big)=P(E_1^c)\times P(E_2^c)\times\cdots\times P(E_n^c)\ \text{(独立性より)} &(1)\\ &= (1-p)\times(1-p)\times\cdots\times(1-p)=(1-p)^n={}_nC_0p^0(1-p)^{n-0},\text{(何故なら}{}_nC_0=1) &(2)\end{aligned}$$

となる。次に

$$\begin{aligned} P(X=1) &= P\Big(\{1\text{ 回目のみ表 }\}\cup\{2\text{ 回目のみ表 }\}\cup\cdots\cup\{n\text{ 回目のみ表 }\}\Big)\\ &= P\Big(\Big[E_1\cap E_2^c\cap\cdots\cap E_n^c\Big]\cup\Big[E_1^c\cap E_2\cap E_3^c\cdots\cap E_n^c\Big]\cup\cdots\cup\Big[E_1^c\cap\cdots\cap E_{n-1}^c\cap E_n\Big]\Big) &(3)\\ &= P(E_1\cap E_2^c\cap\cdots\cap E_n^c)+P(E_1^c\cap E_2\cap E_3^c\cdots\cap E_n^c)+\cdots+P(E_1^c\cap\cdots\cap E_{n-1}^c\cap E_n)\ \text{(排反)} &(4)\end{aligned}$$

2項分布 B(100, p) の2シグマ区間の確率

n	100	100	100	100	100	100	100	100	100
p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
平均	10	20	30	40	50	60	70	80	90
標準偏差	3.0	4.0	4.6	4.9	5.0	4.9	4.6	4.0	3.0
分散	9.0	16.0	21.0	24.0	25.0	24.0	21.0	16.0	9.0
平均+2標準偏差	16	28	39.165	49.798	60	69.798	79.165	88	96
平均－2標準偏差	4	12	20.835	30.202	40	50.202	60.835	72	84
2シグマ区間の確率	0.9557	0.9547	0.9502	0.9585	0.954	0.9581	0.9496	0.9533	0.952

これは確率の n 個の足し算だが、実はおのおのの確率は全て $p(1-p)^{n-1}$ に等しい。例えば、$P(E_1 \cap E_2^c \cap \cdots \cap E_n^c) = p \times (1-p) \times \cdots \times (1-p) = p(1-p)^{n-1}$. 従って、

$$P(X=1) = n \times p(1-p)^{n-1} = {}_nC_1 p^1 (1-p)^{n-1} \quad (\text{何故なら}\,{}_nC_1 = n) \tag{5}$$

が得られる。同様に考えて、$P(X=x)$ を求めよう。まず

$$P\Big(\{\text{ 最初の } x \text{ 回は表、あとは裏 }\}\Big) = P\Big(E_1 \cap \cdots E_x \cap E_{x+1}^c \cap \cdots \cap E_n^c\Big) = p^x(1-p)^{n-x} \tag{6}$$

は明らか。従って

$$P\Big(X=x\Big) = \text{「}n \text{ 回中 } x \text{ 回 } 1 \text{ が出る場合の数」} \times\ p^x(1-p)^{n-x} = {}_nC_x p^x (1-p)^{n-x} \tag{7}$$

が得られる。(証明終)

定理 2. (ポアソン分布 $Po(\lambda)$ の平均と分散) $X \sim Po(\lambda)$ ならば、

$$E(X) = \lambda,\ V(X) = \lambda,\ D(X) = \sqrt{\lambda} \tag{8}$$

(証明). 次の等式（テイラー展開）

$$e^a = 1 + a + a^2/2 + a^3/3! + a^4/4! + \cdots = \sum_{i=0}^{\infty} \frac{a^i}{i!} \tag{9}$$

を用いる。中辺の第 1 項は $1 = a^0/0!$ と考えると右辺とつながる。

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x\, e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} x\, \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=1}^{\infty} x\, \frac{\lambda^x}{x!} \qquad (x=0 \text{ の項は明らかに } 0) && (10)\\
&= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!} \quad \left(\frac{x}{x!} = \frac{1}{(x-1)!}\right) && (11)\\
&= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda \times \lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \quad (i = x-1 \text{ と置いた}) && (12)\\
&= \lambda e^{-\lambda} \times e^{\lambda} = \lambda && (13)
\end{aligned}$$

最終行でテイラー展開を使った。同様にして、$E\{X(X-1)\} = \lambda^2$ が得られる。これを使うと、

$$V(X) = E\{X(X-1)\} + E(X) - \{E(X)\}^2 \tag{14}$$

から分散が求まる。 □

例 1. (交通事故負傷者数) 人口 10 万人のある年は平均して 1 日当たり 6 人の交通事故負傷者が出る。明日の負傷者数を X とすれば、$X \sim B(100000,\ 6/100000)$ と考えることが出来るが、ポアソン分布 $Po(6)$ を用いる方が合理的である。

$$P(X=x) = e^{-6}\frac{6^x}{x!} \quad (x = 0, 1, 2, \ldots). \tag{15}$$

このとき、$E(X) = 6$、$D(X) = \sqrt{6} = 2.45$, $P(X \le 6) = 0.606$ である。 □

例 2.(タイプミスの数). あるタイピストは 1000 字に 1 字の割合でミスをするものとする。1 ページ 2000 字を打つとき、次の問に答えよ。(1)1 ページにミスがせいぜい 1 字だけである確率を求めよ。(2) 1 ページに 3 字以上のミスがある確率を求めよ。(3)50 ページを打つとき、ミスの平均と分散、標準偏差はそれぞれいくらか。(現時点での知識では解けないので直感で)

ポアソン分布Po(λ)の2シグマ区間の確率

平均	2	3	4	5	6	8	10	12	14	16	18	20
標準偏差	1.41	1.73	2.00	2.24	2.45	2.83	3.16	3.46	3.74	4.00	4.24	4.47
平均+2標準偏差	4.83	6.46	8.00	9.47	10.90	13.66	16.32	18.93	21.48	24.00	26.49	28.94
平均−2標準偏差	−0.83	−0.46	0.00	0.53	1.10	2.34	3.68	5.07	6.52	8.00	9.51	11.06
2シグマ区間の確率	0.848	0.917	0.960	0.928	0.963	0.969	0.944	0.958	0.940	0.956	0.941	0.957

(略解)(1) X を 1 ページ当たりのタイプミスの字数とすると、$X \sim Po(2)$. $P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-2}\left\{\frac{2^0}{0!} + \frac{2^1}{1!}\right\} = 0.406$. (2)$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \{P(X = 0) + P(X = 1) + P(X = 2)\} = 1 - 0.677 = 0.323$.
(3) 第 i ページのミスの字数を X_i と置くと ($i = 1, 2, \cdots, 50$)、求めるミスの数 X は $X = X_1 + X_2 + \cdots + X_{50} \sim Po(50 \times 2) = Po(100)$. よって、$E(X) = 100$、$D(X) = \sqrt{100} = 10$。 □

公式: $X \sim Ge(p)$ ならば、$q = 1 - p$ と置けば

$$P(X \leq x) = 1 - q^x, \quad P(X > x) = q^x. \tag{16}$$

となる。 □

定理 3 ((幾何分布 $Ge(p)$ の平均と分散)). $X \sim Ge(p)$ ならば、$q = 1 - p$ と置くと

$$E(X) = 1/p, \quad V(X) = q/p^2, \quad D(X) = \sqrt{q}/p. \tag{17}$$

(証明). 平均を求めると、

$$E(X) = \sum_{x=1}^{\infty} xpq^{x-1} = p\sum_{x=1}^{\infty} xq^{x-1}. \tag{18}$$

ところで、初項 c、公比 r の幾何数列 c, cr, cr^2, $\cdots$ の和は次の通り:

$$c\sum_{i=0}^{\infty} r^i = c(1 + r + r^2 + \cdots) = \frac{c}{1-r}. \tag{19}$$

両辺を r の関数と見て微分すると次式の通り:

$$c\sum_{i=1}^{\infty} ir^{i-1} = c(1 + 2r + 3r^2 + \cdots) = \frac{c}{(1-r)^2}. \tag{20}$$

ここで、$i = x$、$c = p$、$r = q$ と置き換えると、上式左辺が $E(X)$ に等しいことが分かるから、

$$E(X) = p\frac{1}{(1-q)^2} = \frac{1}{p}. \tag{21}$$

同様の考え方でもう一回微分すると、分散を求めることが出来る。 □

例 3. (幾何分布の応用と近似計算). 100 年に一度というような出来事が 5 年以内に起こる確率は幾らか。
(解) X 年目に起こるとすれば、$X \sim Ge(0.01)$。従って、$q = 0.99$ とすれば、

$$P(X \leq 5) = 1 - q^5 = 1 - (0.99)^5 = 0.049 \approx 0.05. \tag{22}$$

一般に、n 年に一度というような出来事が m 年以内に起こる確率は $1 - \{1 - (1/n)\}^m$。$x = 0$ の回りでのテイラー展開により、

$$1 - (1-x)^m \approx mx. \ \text{故に} \ 1 - \{1 - (1/n)\}^m \approx m/n \tag{23}$$

なので、n が大の時、この確率は m/n で近似できる。 □

定義: X が負の 2 項分布 $NB(r, p)$ に従うとは、

$$P(X = x) = {}_{r+x-1}C_x p^r (1-p)^x \quad (x = 0, 1, 2, \cdots) \tag{24}$$

が成り立つことである。 □

定理 4 ((負の 2 項分布の平均と分散)). $X \sim NB(r,p)$ のとき、

$$E(X) = r\left(\frac{1}{p} - 1\right), \quad V(X) = rq/p^2 \tag{25}$$

例 4 .(カードが無効になるまでの引き落とし回数). 暗証番号を 3 回間違えると無効になるキャッシュカードがある。ある年齢層の男性は 5 回に 1 回の割合で暗証番号を間違えるものとする。この年齢層の男性がキャッシュカードを使うとき、無効になるまで引き落としを行う回数を X と置くと、$X \sim NB(3, 1/5) = NB(3, 0.2)$ である。従って、

$$E(X) = 3 \times (5-1) = 12 \text{ 回}, \quad V(X) = 3 \times (4/5)/(1/5)^2 = 36, \quad D(X) = 6 \text{ 回} \tag{26}$$

また、X の分布は次表の通りである。例えば、

$$P(X = 12) = {}_{3+12-1}C_{12} 0.2^3 (1-0.2)^{12} = 0.05$$

である。 □