確率変数と確率分布

数学と医療・第 3 回

数学・橋本貴宏

2024 年 9 月 18 日 1 限

本日の授業では

- 前回のおさらい
  (1) 確率の定義
  (2) ベイズの定理
- 確率変数と確率分布
  (1) 定義（離散型）
  (2) 定義（連続型）
  (3) 期待値と分散

前回の演習問題

問題

工場 A と工場 B は, ある同じ医療機器を生産していて, 全体のうち 65 % は工場 A で, 35 % は工場 B で生産しているとする。工場 A で不良品を生産してしまう確率を 0.8 %, 工場 B で不良品を生産してしまう確率を 0.6 % とする。ある製品が不良品だったと報告を受けたとき，その製品が工場 B で生産された確率を求めよ。

演習問題解答

2 つの事象を, A: B 工場で生産された, B: 不良品である, とする.

$$P(A) = 0.35,\ P(\overline{A}) = 0.65,\ P(B|A) = 0.006,\ P(B|\overline{A}) = 0.008$$

より求める確率は

$$P(A|B) = \frac{0.35 \cdot 0.006}{0.35 \cdot 0.006 + 0.65 \cdot 0.0008} = \frac{35 \cdot 6}{35 \cdot 6 + 65 \cdot 8} = \frac{21}{73}.$$

みなさんからのコメント

- 証明の問題についてはいろいろな解答がある問題もあると思うので、みんなの異なる解答が見れたらもっといいなと思った
  そのままに見せるのはちょっと問題があるので, やり方を考え中です
- イプシロンデルタ論法に触れられるかと思った
  この授業では扱えないでしょう. 去年のセミナーでちょこっと出てきました
- 講義資料の中で、$P(B|A)$ の定義がいくつか書かれていたり、日本語がややこしくあまり理解できなかった
  定義は $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$ しかないのですが, バリエーションがいっぱいあったということでしょうか.
- 「検査後確率」の意味がよくわからなかったです
  前回の例では, 検査が陽性であったときの罹患有の確率です

2×2 分割表と定性テスト

	疾患あり (A)	疾患なし ($\overline{A}$)
陽性 (B)	a (TP)	b (FP)
陰性 ($\overline{B}$)	c (FN)	d (TN)

事象 A: 疾患あり，　事象 B: 検査で陽性と判定される

感度，特異度，etc. と条件付き確率

$$\text{感度} = \frac{a}{a+c} = P(B|A)$$

$$\text{特異度} = \frac{d}{b+d} = P(\overline{B}|\overline{A})$$

$$\text{陽性的中率} = \frac{a}{a+b} = P(A|B)$$

$$\text{陰性的中率} = \frac{d}{c+d} = P(\overline{A}|\overline{B})$$

練習問題 (その 1)

第 118 回医師国家試験問題 (C063)

55 歳の男性。便秘を主訴に来院した。

– (中略) –

腹部エックス線写真で小腸ガスや鏡面像を認めない。まず便潜血反応を行うこととした。

この患者の大腸に何らかの病変がある検査前確率 (事前確率) を 20% としたとき, 便潜血反応陽性であった場合の検査後確率はいくつか。

ただし, 便潜血反応の感度は 80%, 特異度は 90% とする。

練習問題 (その 2)

A: 病変がある. B: 検査で陽性となる, とする.

$$\boldsymbol{P(A) = 0.2,\ P(B|A) = 0.8,\ P(\overline{B}|\overline{A}) = 0.9}$$

である. このとき $\boldsymbol{P(B|\overline{A}) = 0.1}$ であるからベイズの定理より検査後確率は

$$\boldsymbol{P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})} = \frac{0.2 \cdot 0.8}{0.2 \cdot 0.8 + 0.8 \cdot 0.1} = \frac{2}{3}}$$

よって求める確率はおよそ $\boldsymbol{67}\%$ である.

参考：選択肢は 33%, 53%, 57%, 67%, 97%

前回のおさらい (その 1)

確率の公理

標本空間を U とする. 事象 A, B に対する確率は, 次の性質をみたす. (次をみたすものを確率と定義する)

(1) $0 \leqq P(A) \leqq 1$

(2) $P(U) = 1$, $P(\emptyset) = 0$

(3) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

条件付き確率

A が起きたという条件のもとで B が起きる条件付き確率を

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (P(A) \neq 0)$$

前回のおさらい (その 2)

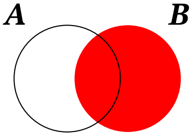
で定義する. $P(A \cap B) = P(A)P(B)$ のとき A と B は互いに独立という.

ベイズの定理

$P(B|A)$: A が起こったときの B の起こる条件付確率

これに対して「結果」が出たときの「原因」の確率 $P(A|B)$ を求める公式は,

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})}$$



分子 $\cdots$ $P(B \cap A)$

分母 $\cdots$ $P(B)$

ベルヌーイ試行

1 回の試行の結果が, 事象 A が「起こる」か「起こらない」のいずれかであるとする. この試行を繰り返して行い, 各回の結果が互いに他の回の結果に左右されず独立のとき, この試行を**ベルヌーイ試行**という.

(例)　コインやさいころを投げる試行はベルヌーイ試行であり, くじ引き (元に戻さない場合) はベルヌーイ試行ではない.

ベルヌーイ試行の確率

1 回のベルヌーイ試行で事象 A の起こる確率が $\boldsymbol{p}$ であるとする. この試行を $\boldsymbol{n}$ 回繰り返すとき, 事象 A がちょうど $\boldsymbol{x}$ 回起こる確率は

$$\boldsymbol{{}_nC_x p^x (1-p)^{n-x}}$$

で与えられる.

- 実例から

2 枚の硬貨を投げたとき, 表の出た枚数を X とする.

「表」「表」　「表」「裏」　「裏」「表」　「裏」「裏」

$\Downarrow$　　$\Downarrow$　　$\Downarrow$　　$\Downarrow$

$X = 2$　　$X = 1$　　$X = 1$　　$X = 0$

$X = \{0, 1, 2\}$ であって

$$P(X = 0) = \frac{1}{4}, \quad P(X = 1) = \frac{1}{2}, \quad P(X = 2) = \frac{1}{4}$$

表にすると,

表の数 X	0	1	2	計
確率 P	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1

(離散型) 確率分布の定義

ある変数 X について, X のとりうる値 $x_1, x_2, \cdots, x_n$ は決まっているが、何が出るかは決まっていなく、その代わりに出る確率 $p_1, p_2, \cdots, p_n$ が決まっているとき、X を

確率変数 (random variable)

という. そしてその X の一つ一つからその確率への対応を

確率分布 (probability distribution)

という. そして, 変数 $X = x_i$ に対して確率 p_i が対応していることを

$$P(X = x_i) = p_i$$

と表す.

確率変数の期待値と分散 (離散型)

確率分布表:

X の値	x_1	x_2	$\cdots$	x_n	計
$P(X=x_i)$	p_1	p_2	$\cdots$	p_n	1

確率分布 $p_i = P(X = x_i)$ は次の性質をもつ.

(1) $p_i \geqq 0 \quad (i = 1, \cdots, n)$　　(2) $\displaystyle\sum_{i=1}^{n} p_i = 1.$

期待値 (expectation)　$$E(X) := \sum_{i=1}^{n} x_i p_i \ (=: \mu)$$

分散 (variance)　$$V(X) := E((X-\mu)^2) = \sum_i (x_i - \mu)^2 p_i.$$

宝くじの「期待値」

ハロウィンジャンボ宝くじ … 1 ユニット 1000 万枚

	当選金	本数		当選金	本数
1 等	3 億円	1	3 等	100 万円	200
1 等 (前後賞)	1 億円	2	4 等	1 万円	1000
1 等 (組違い)	10 万円	99	5 等	3000 円	10 万
2 等	1000 万円	10	6 等	300 円	100 万

$$
\begin{aligned}
E(X) = \frac{1}{10^7} \Big[& 3 \cdot 10^8 \times 1 + 10^8 \times 2 + 10^5 \times 99 + 10^7 \times 10 \\
& + 10^6 \times 200 + 3 \cdot 10^4 \times 10^3 + 3 \cdot 10^3 \times 10^5 \\
& + 300 \times 10^6 \Big] = 141.99
\end{aligned}
$$

期待値の性質

X の関数 $g(X)$ の期待値 (離散型) を $E(g(X)) = \sum_{i=1}^{n} g(x_i)p_i$ で定義する.

(1) $g(X)$ が 1 次式のとき

$$E(aX+b) = \sum_{i=1}^{n}(ax_i+b)p_i = aE(X)+b$$

(2) $g(X)$ が 2 次式のとき

$$E(aX^2+bX+c) = \sum_{i=1}^{n}(ax_i^2+bx_i+c)p_i = aE(X^2)+bE(X)+c$$

分散の性質

(1) $E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$ より，

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

(2) $E(aX+b) = aE(X) + b = a\mu + b$ であったから

$$\begin{aligned} V(aX+b) &= E([(aX+b) - (a\mu+b)]^2) \\ &= E([a(X-\mu)]^2) = a^2E((X-\mu)^2) = a^2V(X) \end{aligned}$$

標準化

期待値 $\boldsymbol{\mu}$, 分散 $\boldsymbol{\sigma^2}$ $(\boldsymbol{\sigma > 0})$ の確率変数 $\boldsymbol{X}$ を次のようにに変換してみる.

$$\boldsymbol{Z = \frac{X - \mu}{\sigma}}$$

このとき

$$\boldsymbol{E(Z) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = 0}$$

また

$$\boldsymbol{V(Z) = \frac{1}{\sigma^2}V(X) = 1,}$$

つまり期待値 0, 分散 1 となる. これを**標準化**変換という.

離散型分布の例 (その1)

さいころを 1 回投げて, 出た目を確率変数 X とする. 確率分布表は次のようになる.

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

期待値は

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}.$$

分散を定義通り求めると, $V(X) = \sum_{i=1}^{6} (x_i - \frac{7}{2})^2 p_i$ であるが, 先ほどの公式を使う.

離散型分布の例 (その 2)

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

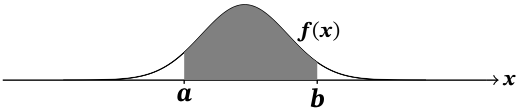
$$E(X^2) = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6}$$

であるから

$$V(X) = E(X^2) - E(X)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

連続型確率分布 (その1)

X: 日本人全体の身長, とすると X は連続的な値をとるので確率分布表は作れない. **確率密度関数** f を用いてグラフで表す必要がある.



f は次をみたす

$$f(x) \geqq 0, \quad \int_{-\infty}^{\infty} f(x)\, dx = 1.$$

$a < b$ として $a \leqq X \leqq b$ である確率は次で定義される:

$$P(a \leqq X \leqq b) = \int_a^b f(x)\, dx.$$

つまり, **「確率はグラフで囲まれた部分の面積」**である

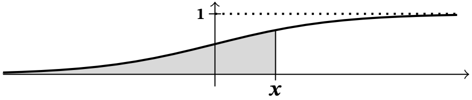
連続型確率分布 (その 2)

性質

- 任意の a について $P(X = a) = P(a \leqq X \leqq a) = \int_a^a f(x)dx = 0$
- $P(a < X < b) = P(a \leqq X < b) = P(a < X \leqq b) = P(a \leqq X \leqq b)$

連続型確率変数 X の取る値が x 以下である確率 $P(X \leqq x)$ を考え

$$F(x) = P(X \leqq x) = \int_{-\infty}^{x} f(t)dt$$



を X の**分布関数**と定義する.

- Excel で確率を求めるときは, **分布関数**を用いる.

(例) NORM.S.DIST(x の値, TRUE) で標準正規分布の $P(X \leqq x)$ の値.

連続型確率変数の期待値と分散

期待値 $\quad E(X) := \int_{-\infty}^{\infty} x f(x) dx \ (=: \mu)$

分散 $\quad V(X) := E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$

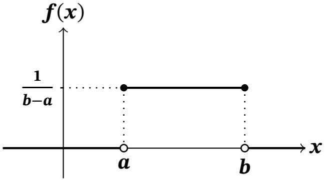
$$V(X) = \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx$$

$$- 2\mu \int_{-\infty}^{\infty} x f(x) dx + \mu^2 = E(X^2) - (E(X))^2$$ が成り立つ.

連続型確率分布の例

確率密度関数が

$$
f(x) = \begin{cases} \dfrac{1}{b-a} & (a \leqq x \leqq b) \\ 0 & (\text{それ以外}) \end{cases}
$$



と表される連続分布 X を **(連続) 一様分布**という．期待値と分散は

$$
E(X) = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{a+b}{2},
$$

$$
V(X) = \int_a^b \frac{x^2}{b-a} dx - E(X)^2 = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b - \frac{(a+b)^2}{4}
$$

$$
= \frac{4(a^2+ab+b^2) - 3(a^2+2ab+b^2)}{12} = \frac{(b-a)^2}{12}.
$$

まとめ

- 確率変数 X と確率分布

  $X = x_i$ からそのときの確率 p_i への対応が (離散型) 確率分布

- 離散型確率分布

  確率分布表をかけば良い

- 連続型確率分布

  確率密度関数の**グラフにより囲まれた部分の面積**が確率である

- 期待値と分散

  期待値は平均ともいう．基本的な統計量でもある

本日の演習問題

問題

箱の中に当りくじ 3 本と外れくじ 9 本が入っている。この中から 1 本ずつ 3 回くじを引いて当りくじが出た回数を確率変数 $\boldsymbol{X}$ とするとき, 次の問いに答えよ. ただし, 1 回ごとにくじを箱の中に戻すとする。

(1) $\boldsymbol{X}$ の確率分布を求め確率分布表をかけ。

(2) 期待値 $\boldsymbol{E}(\boldsymbol{X})$ 及び分散 $\boldsymbol{V}(\boldsymbol{X})$ を求めよ。

次回

二項分布, ポアソン分布をやった後に, 同時確率分布をやります.

課題提出

AIDLE-K にアクセスし, 「0918 課題提出」から提出する. 締め切りは 9/21 23:59 とする. **フィードバックも忘れずに** (こちらは今日中).

注意

- ファイル形式に注意し, 拡張子をつける (jpg, pdf 等)
- 解答用紙にも名前を書くこと
- **過去の課題プリントで出していないものがある場合は, 超過でもかまわないので出すこと**