

離散型確率分布

数学と医療・第 4 回

数学・橋本貴宏

2024 年 9 月 25 日 1 限

本講義資料の無断使用・無断転載・SNS 等への投稿を固く禁じます

本日の授業では

- 前回のおさらい
 - (1) 確率変数と確率分布
 - (2) 期待値と分散
- 2変量の確率分布
 - (1) 同時確率分布
 - (2) 確率変数の独立性
- 離散型確率分布
 - (1) 二項分布
 - (2) ポアソン分布

前回のおさらい (その1)

確率分布 (離散型)

確率変数: 根元事象から実数への対応.

(離散型) 確率分布: $X = x_i$ から $P(X = x_i) = p_i$ への対応

確率分布表

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

$$(0 \leq p_i \leq 1, \sum_{i=1}^n p_i = 1)$$

期待値 (平均とも呼ばれる) $E(X) := \sum_{i=1}^n x_i p_i$ ($= \mu$ とおく),

分散 $V(X) := E((X - \mu)^2) = \sum_{i=1}^n (x_i - \mu)^2 p_i$

前回のおさらい (その2)

確率分布 (連続型)

連続型確率分布とは, 確率密度関数 $f(x)$: $f(x) \geq 0$ で $\int_{-\infty}^{\infty} f(x)dx = 1$, を用いて

$$P(a \leq X \leq b) = \int_a^b f(x)dx \quad (a < b)$$

と表される確率分布のことである. $f(x)$ が確率を表すのではなく, その積分が確率 $P(a \leq X \leq b)$ を表す.

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

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

無限区間の積分 (その1)

一般的には**広義積分**という. a を実数として, f は $[a, \infty)$ で定義されているとする. 区間 $[a, N]$ で積分 $\int_a^N f(x) dx$ が存在するとする. このとき, 有限な極限值 $\lim_{N \rightarrow \infty} \int_a^N f(x) dx$ が存在するならば, f は $[a, \infty)$ で広義積分可能であるといい, この極限値を $\int_a^\infty f(x) dx$ で表す.

- $(-\infty, a]$ の場合は次で定義する;

$$\int_{-\infty}^a f(x) dx = \lim_{M \rightarrow \infty} \int_{-M}^a f(x) dx.$$

無限区間の積分 (その2)

- $(-\infty, \infty)$ の場合は

$$\int_a^{\infty} f(x) dx = \lim_{N \rightarrow \infty} \int_a^N f(x) dx \quad \text{と} \quad \int_{-\infty}^a f(x) dx = \lim_{M \rightarrow \infty} \int_{-M}^a f(x) dx$$

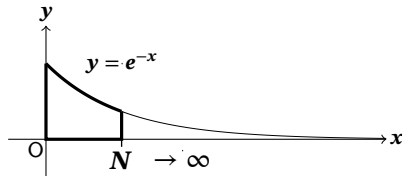
が共に存在するとき広義積分可能であるとする. これらの和を

$$\int_{-\infty}^{\infty} f(x) dx \text{ で表す.}$$

(例) $\int_0^N e^{-x} dx = [-e^{-x}]_0^N = 1 - e^{-N}$

となるので,

$$\int_0^{\infty} e^{-x} dx = \lim_{N \rightarrow \infty} (1 - e^{-N}) = 1.$$



みなさんからのコメント

- 期待値が高校の新課程で導入されたのは知っていて、自分がしていないことなので難しそうなイメージを持っていましたが、実際に計算方法を知ってみると、結構できるなと思いました
期待値は数学 A から出たり入ったりしています
- 授業内での演習課題では連続型の確率分布は扱わなかったため、少し物足りなく感じてしまいました
連続型は、積分で表されるため、避けました
- 講義では国家試験の問題を適宜紹介されているのが、勉強するモチベーションに繋がりととても良かったです
国試で出てくるのは、確率の計算くらいでしょうか。あとは、統計解析の方法が出ます

9月18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 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) X の確率分布を求め確率分布表をかけ。
- (2) 期待値 $E(X)$ 及び分散 $V(X)$ を求めよ。

解説 (その1)

(1) 1回引いたときに当りである確率は $\frac{3}{12} = \frac{1}{4}$ なので, 試行回数 3, $p = \frac{1}{4}$ のベルヌーイ試行となり, x 回当たる確率は $P(X = x) = {}_3C_x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{3-x}$.

$$P(X = 0) = \left(\frac{3}{4}\right)^3 = \frac{27}{64}, \quad P(X = 1) = 3 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^2 = \frac{27}{64},$$

$$P(X = 2) = 3 \cdot \left(\frac{1}{4}\right)^2 \cdot \frac{3}{4} = \frac{9}{64}, \quad P(X = 3) = \left(\frac{1}{4}\right)^3 = \frac{1}{64}.$$

となるので確率分布表は次のようになる.

確率分布表

X	0	1	2	3
P	27/64	27/64	9/64	1/64

解説 (その2)

X	0	1	2	3
X^2	0	1	4	9
P	27/64	27/64	9/64	1/64

(2) (X^2 を付け加えた) 確率分布表より

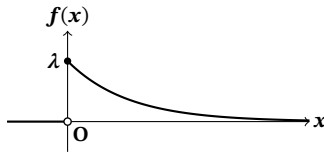
$$E(X) = \sum_{i=0}^3 i p_i = \frac{27}{64} + 2 \cdot \frac{9}{64} + 3 \cdot \frac{1}{64} = \frac{27 + 18 + 3}{64} = \frac{3}{4},$$

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 = \sum_{i=0}^3 i^2 p_i - \left(\frac{3}{4}\right)^2 \\ &= \frac{27}{64} + 4 \cdot \frac{9}{64} + 9 \cdot \frac{1}{64} - \frac{9}{16} = \frac{72}{64} - \frac{9}{16} = \frac{9}{16}. \end{aligned}$$

指数分布 (その1)

確率密度関数

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$



により表される連続分布 X を指数分布という. $f(x) \geq 0$ で

$$\begin{aligned} P(-\infty < X < \infty) &= \int_0^{\infty} \lambda e^{-\lambda x} dx = \lim_{N \rightarrow \infty} [-e^{-\lambda x}]_0^N \\ &= \lim_{N \rightarrow \infty} (-e^{-\lambda N} + 1) = 1 \end{aligned}$$

であるから確率分布の条件をみたす. 期待値は

指数分布 (その2)

$$E(X) = \int_0^{\infty} \lambda x e^{-\lambda x} dx = \lambda \left(\left[-\frac{x e^{-\lambda x}}{\lambda} \right]_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} e^{-\lambda x} dx \right) = \left[-\frac{e^{-\lambda x}}{\lambda} \right]_0^{\infty} = \frac{1}{\lambda}$$

であり, 分散は

$$\begin{aligned} E(X^2) &= \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \lambda \left(\left[-\frac{x^2 e^{-\lambda x}}{\lambda} \right]_0^{\infty} + \frac{1}{\lambda} \int_0^{\infty} 2x e^{-\lambda x} dx \right) \\ &= \frac{2}{\lambda} \int_0^{\infty} \lambda x e^{-\lambda x} dx = \frac{2}{\lambda^2} \quad \text{より} \end{aligned}$$

$$V(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

2つの離散型確率変数 X, Y に対して $X = x_i$ かつ $Y = y_j$ である確率が

$$P(X = x_i, Y = y_j) = p_{ij} \quad (1 \leq i \leq n, 1 \leq j \leq m)$$

で与えられ, $0 \leq p_{ij} \leq 1$, $\sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1$ であるとき, (X, Y) は同時確率分布 に従うという. また, X, Y 一方に着目した

$$p_{i\cdot} = P(X = x_i) = \sum_{j=1}^m p_{ij}, \quad p_{\cdot j} = P(Y = y_j) = \sum_{i=1}^n p_{ij}$$

を周辺確率分布という. $(\sum_{i=1}^n p_{i\cdot} = \sum_{j=1}^m p_{\cdot j} = 1)$

同時確率分布表

$X \backslash Y$	y_1	$\cdots$	y_j	$\cdots$	y_m	X の周辺分布
x_1	p_{11}	$\cdots$	p_{1j}	$\cdots$	p_{1m}	$p_{1\cdot}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	p_{i1}	$\cdots$	p_{ij}	$\cdots$	p_{im}	$p_{i\cdot}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_n	p_{n1}	$\cdots$	p_{nj}	$\cdots$	p_{nm}	$p_{n\cdot}$
Y の周辺分布	$p_{\cdot 1}$	$\cdots$	$p_{\cdot j}$	$\cdots$	$p_{\cdot m}$	1

(離散型) 確率変数 (X, Y) において

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

の関係が成り立つとき, X と Y は独立であるという. (周辺分布の積)

例題 3.1.3

(p.48)

6 が 1 に書き換えられているので,

$X \backslash Y$	1	2	3	4	5	X の周辺分布
1	4/36	2/36	2/36	2/36	2/36	2/6
2	2/36	1/36	1/36	1/36	1/36	1/6
3	2/36	1/36	1/36	1/36	1/36	1/6
4	2/36	1/36	1/36	1/36	1/36	1/6
5	2/36	1/36	1/36	1/36	1/36	1/6
Y の周辺分布	2/6	1/6	1/6	1/6	1/6	1

例題 3.1.3 (つづき)

すべての x, y に対して,

$$P(X = 1)P(Y = 1) = \frac{2}{6} \cdot \frac{2}{6} = \frac{4}{36} = P(X = 1, Y = 1)$$

$$P(X = 1)P(Y = y) = \frac{2}{6} \cdot \frac{1}{6} = \frac{2}{36} = P(X = 1, Y = y) \quad (y = 2, 3, 4, 5)$$

$$P(X = x)P(Y = 1) = \frac{1}{6} \cdot \frac{2}{6} = \frac{2}{36} = P(X = x, Y = 1) \quad (x = 2, 3, 4, 5)$$

$$P(X = x)P(Y = y) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} = P(X = x, Y = y) \quad (x, y = 2, 3, 4, 5)$$

が成り立つので X と Y は独立である.

2つの確率変数 X, Y の和 $X + Y$ と積 XY も確率変数である. 離散型の場合,

$$E(X + Y) = \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j) p_{ij},$$

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{ij}$$

であり, 次が成り立つ.

$$(1) E(X + Y) = E(X) + E(Y)$$

$$(2) X \text{ と } Y \text{ が独立ならば } E(XY) = E(X)E(Y)$$

期待値 (その2)

$$\begin{aligned} \text{(証)} \quad (1) \quad E(X + Y) &= \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j) p_{ij} = \sum_{i=1}^n \sum_{j=1}^m x_i p_{ij} + \sum_{j=1}^m \sum_{i=1}^n y_j p_{ij} \\ &= \sum_{i=1}^n x_i \sum_{j=1}^m p_{ij} + \sum_{j=1}^m y_j \sum_{i=1}^n p_{ij} = \sum_{i=1}^n x_i p_{i\cdot} + \sum_{j=1}^m y_j p_{\cdot j} = E(X) + E(Y) \end{aligned}$$

(2) X, Y が独立ならば, $p_{ij} = p_{i\cdot} p_{\cdot j}$ となるから

$$\begin{aligned} E(XY) &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{ij} = \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{i\cdot} p_{\cdot j} \\ &= \sum_{i=1}^n x_i p_{i\cdot} \sum_{j=1}^m y_j p_{\cdot j} = E(X)E(Y) \end{aligned}$$

例題 (同時分布の期待値)

右の分布表で与えられる
同時確率分布について,
 $E(X)$, $E(Y)$, $E(XY)$
を求めよ

$X \backslash Y$	1	2	3	X の周辺分布
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{2}{3}$
4	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{3}$
Y の周辺分布	$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{4}$	1

例題 (同時分布の期待値) 解答

$$X \text{ の期待値: } E(X) = 2 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3} = \frac{8}{3}.$$

$$Y \text{ の期待値: } E(Y) = 1 \cdot \frac{5}{12} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{4} = \frac{5+8+9}{12} = \frac{11}{6}.$$

$x_i y_j$ の表は

X \ Y	1	2	3
	2	4	6
4	4	8	12

であるから, XY の期待値は

$$\begin{aligned} E(XY) &= 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} + 6 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 8 \cdot \frac{1}{12} + 12 \cdot \frac{1}{12} \\ &= \frac{6+12+12+8+8+12}{12} = \frac{29}{6}. \end{aligned}$$

X と Y の共分散 (covariance) $\mathbf{Cov}(X, Y)$ を

$$\mathbf{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \quad (\mu_X = E(X), \mu_Y = E(Y))$$

で定義する. 離散型同時確率分布については

$$\mathbf{Cov}(X, Y) = \sum_{i=1}^n \sum_{j=1}^m (x_i - \mu_X)(y_j - \mu_Y) p_{ij}$$

である. また, X と Y の相関係数 $\rho(X, Y)$ を

$$\rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \quad \text{で定義する.}$$

共分散 (その2)

定理 3.2.3 定数 a, b, c, d と確率変数 X, Y について次が成り立つ

- (1) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
- (2) X と Y が独立のとき $\text{Cov}(X, Y) = 0$
- (3) $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$

(証) (1) $\text{Cov}(X, Y) = E(XY - \mu_X Y - X\mu_Y + \mu_X \mu_Y)$

$$= E(XY) - \mu_X E(Y) - E(X)\mu_Y + \mu_X \mu_Y = E(XY) - E(X)E(Y)$$

(2) X と Y が独立のとき $E(XY) = E(X)E(Y)$ なので成り立つ.

(3) $E(aX + b) = a\mu_X + b$, $E(cY + d) = c\mu_Y + d$ であるから,

$$\begin{aligned}\text{Cov}(aX + b, cY + d) &= E[\{(aX + b) - (a\mu_X + b)\}\{(cY + d) - (c\mu_Y + d)\}] = \\ &= E[a(X - \mu_X)c(Y - \mu_Y)] = acE[(X - \mu_X)(Y - \mu_Y)] = ac\text{Cov}(X, Y)\end{aligned}$$

(これより, $a, c > 0$ のとき $\rho(aX + b, cY + d) = \rho(X, Y)$ も示される)

共分散 (その3)

$E(aX + bY) = a\mu_X + b\mu_Y$ より,

$$\begin{aligned} V(aX + bY) &= E((aX + bY - a\mu_X - b\mu_Y)^2) = E([a(X - \mu_X) + b(Y - \mu_Y)]^2) \\ &= a^2E((X - \mu_X)^2) + 2abE((X - \mu_X)(Y - \mu_Y)) + b^2E((Y - \mu_Y)^2) \end{aligned}$$

であるから

$$V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab\text{Cov}(X, Y)$$

が成り立ち, X, Y が独立ならば,

$$V(aX + bY) = a^2V(X) + b^2V(Y).$$

※ $E(X - Y) = E(X) - E(Y)$ であるが, $V(X - Y) = V(X) + V(Y)$

二項分布 (その1)

(p.61)

確率変数 X が $0, 1, \dots, n$ の値をとるとき,

$$P(X = x) = {}_nC_x p^x (1 - p)^{n-x} \quad (0 < p < 1)$$

で与えられる確率分布を二項分布といい, $B(n, p)$ で表す. $B(1, p)$ はベルヌーイ試行という. 二項定理より

$$\sum_{x=0}^n P(X = x) = \sum_{x=0}^n {}_nC_x p^x (1 - p)^{n-x} = (p + 1 - p)^n = 1.$$

確率分布表は,

X	0	...	x	...	n
$P(X = x)$	$(1 - p)^n$	...	${}_nC_x p^x (1 - p)^{n-x}$	...	p^n

となる.

定理 3.3.1 X が二項分布 $B(n, p)$ に従うとき,

$$(1) E(X) = np \quad (2) V(X) = np(1-p) \quad (3) \sigma(X) = \sqrt{np(1-p)}$$

(証) まず $n = 1$ のときを求める. 確率分布表をかくと

X	0	1
$P(X = k)$	$1 - p$	p

となるので, 期待値と分散は,

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p,$$

$$V(X) = 0^2 \cdot (1-p) + 1^2 \cdot p - p^2 = p(1-p)$$

二項分布の期待値・分散 (その2)

となる. これらの試行を $X_1, X_2, \dots, X_n$ として n 回繰り返したものを X とすると, $X = X_1 + \dots + X_n$ であって, 各試行 X_i は $B(1, p)$ に従うので $E(X_i) = p$ である. よって

$$E(X) = E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) = np.$$

分散については, 各試行 X_i については, $V(X_i) = p(1 - p)$ であって, X_i が互いに独立であることに注意すると,

$$V(X) = V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n) = np(1 - p).$$

最後に $\sigma(X) = \sqrt{V(X)} = \sqrt{np(1 - p)}$.

例題 3.3.1

(p.62)

1 回で治療が終了する確率は **54%** である. 8 人の患者に治療を行ったとき, 1 回で治療が終了する人数を X とする.

(1) $P(X = x) = {}_8C_x(0.54)^x(0.46)^{8-x}$

(2) 二項分布 $B(8, 0.54)$ に従う

(3) $P(X = 6) = {}_8C_6(0.54)^6(0.46)^2 \doteq 0.147$

X は二項分布 $B(8, 0.54)$ に従うので, 期待値と分散は

$$E(X) = 8 \cdot 0.54 = 4.32,$$

$$V(X) = 8 \cdot 0.54 \cdot 0.46 = 1.9872.$$

確率変数 X が $0, 1, 2, \dots, n, \dots$ の値をとるとき,

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (\lambda > 0)$$

で与えられる確率分布をポアソン分布といい, $Po(\lambda)$ で表す.

- 指数関数 e^λ のマクローリン展開より

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

(指数関数のマクローリン展開とは, $e^x = 1 + x + \frac{1}{2}x^2 + \dots + \frac{1}{n!}x^n + \dots$ という級数展開のこと)

ポアソン分布の期待値・分散

ポアソン分布は, 二項分布 $B(n, p)$ において $np = \lambda$ (一定) とおいて $n \rightarrow \infty$ としたものと考えられる.

$$n \rightarrow \infty \text{ のとき, } {}_nC_x p^x (1-p)^{n-x} \rightarrow \frac{\lambda^x}{x!} e^{-\lambda} \text{ より示される}$$

二項分布の期待値は np , 分散は $np(1-p)$ であったから, ポアソン分布は,

$$\text{期待値: } E(X) = \lambda, \quad \text{分散: } V(X) = \lambda$$

となる.

まとめ

- 2変量の分布

独立性は（統計的）仮説検定において再び登場します

- 二項分布 $B(n, p)$

$E(X) = np$, $V(X) = np(1 - p)$ は重要事項

- ポアソン分布 $Po(\lambda)$

二項分布の極限. X の取りうる値は無限個である

9月25日の演習問題

下の分布表で与えられる同時確率分布について、次の問いに答えよ。

$X \backslash Y$	1	2	4	X の周辺分布
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{2}{3}$
3	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{3}$
Y の周辺分布	$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{4}$	1

- (1) 期待値 $E(XY)$ を求めよ。
- (2) 共分散 $\text{Cov}(X, Y)$ を求めよ。

次回

「連続型確率分布」の代表格である「正規分布」をやります. この分布の理解が, 統計学Iでやることの理解につながります.

課題提出

AIDLE-K にアクセスし, 「0925 課題提出」から提出する. 提出期限は **9/28 23:59** とする. フィードバックも忘れずに.

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**