

Lecture 4

離散型確率分布

確率変数(random variable)とは？

現象
調査対象



観測値 x

確率変数として扱う

変数：ある範囲の値を代表する. $x, y, z, t, \dots$

確率変数：確率を伴ってある範囲の値をとる. $X, Y, Z, T, \dots$

➤ Discrete random variables (離散型確率変数)

- (1) コインを3回投げるとき表の出る回数.
- (2) 授業開始時の出席者数.

数える

➤ Continuous random variables (連続型確率変数)

- (1) 円の内部から1点をランダムに選んだとき, その点と中心との距離.
- (2) 新生児の体重.

測る/量る

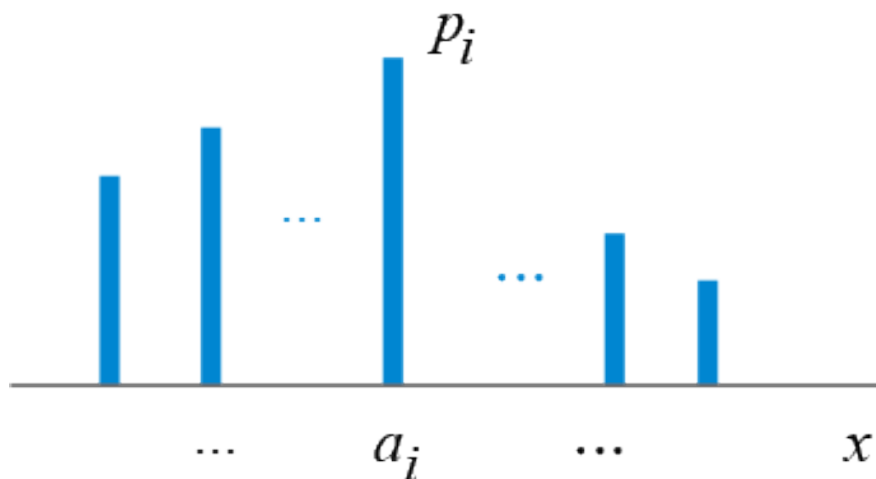
(注意) 確率変数 X は特定の数を表すのではない.
その取りうる個別の値を X の実現値という.

離散型確率変数の分布(distribution)

取りうる値 $a_1, a_2, \dots, a_i, \dots$

その確率 $p_1, p_2, \dots, p_i, \dots$

x	a_1	$\dots$	a_i	$\dots$	合計
$P(X = x)$	p_1	$\dots$	p_i	$\dots$	1



数式による表記

$$P(X = a_i) = p_i$$

基本的な性質

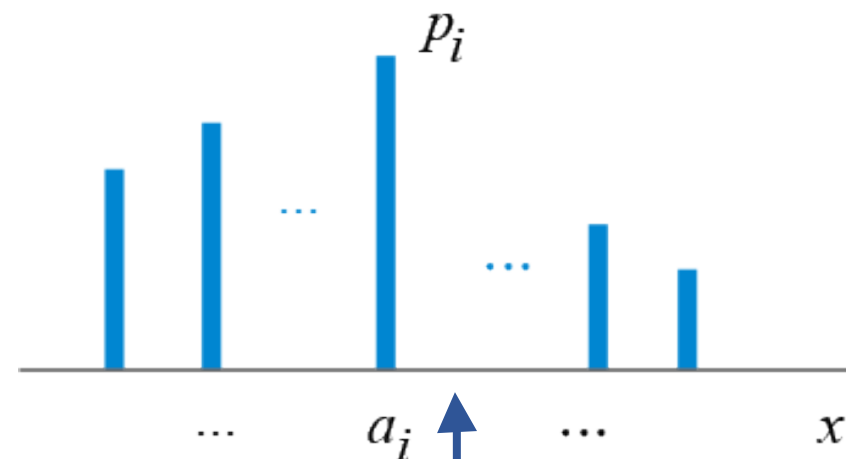
$$(1) \quad p_i \geq 0$$

$$(2) \quad \sum p_i = 1$$

離散型確率変数の平均値と分散

離散型確率変数 X の確率分布

x	a_1	...	a_i	...	合計
$P(X = x)$	p_1	...	p_i	...	1



➤ 平均値 $\mathbf{E}[X] = m_X = \sum a_i p_i = \sum a_i P(X = a_i)$

重心

➤ 分散 $\mathbf{V}[X] = \sigma_X^2 = \sum (a_i - m_X)^2 p_i = \sum a_i^2 p_i - m_X^2$

$$= \mathbf{E}[(X - m_X)^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 \quad \text{【分散公式】}$$

例題 4.1 サイコロを 1 回振って, 出た目を X は $\{1,2,3,4,5,6\}$ の範囲を動く確率変数である. X の確率分布, 平均値, 分散を求めよ.

例題 4.1 サイコロを1回振って、出た目を X は $\{1,2,3,4,5,6\}$ の範囲を動く確率変数である. X の確率分布, 平均値, 分散を求めよ.

確率分布

$$P(X = k) = \frac{1}{6}, \quad k = 1, 2, \dots, 6$$

x	1	2	3	4	5	6	合計
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

平均値

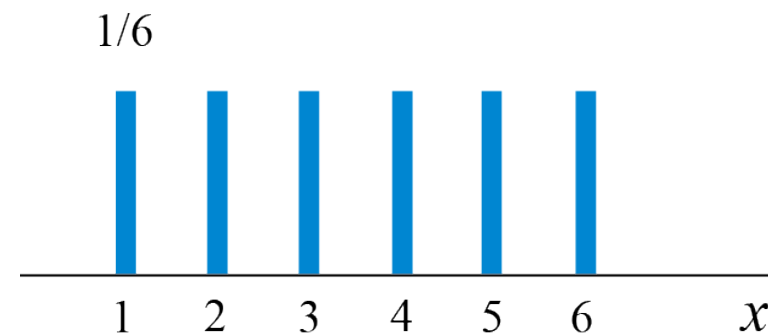
$$\mathbf{E}[X] = m_X = \sum a_i p_i = \sum_{k=1}^6 k \frac{1}{6} = \frac{7}{2}$$

$$\mathbf{E}[X^2] = \sum a_i^2 p_i = \sum_{k=1}^6 k^2 \frac{1}{6} = \frac{91}{6}$$

分散

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

【分散公式】



名前の付いた重要な離散分布

- ✓ ベルヌイ分布 二項母集団の分布
- ✓ 二項分布 ベルヌイ試行列における成功回数の分布
- ✓ 幾何分布 ベルヌイ試行列における成功までの待ち時間
- ✓ ポアソン分布 二項分布の極限（少数の法則）

教科書等で見しておく

ベルヌイ分布 $B(1, p)$

2 値確率変数 = ベルヌイ型確率変数

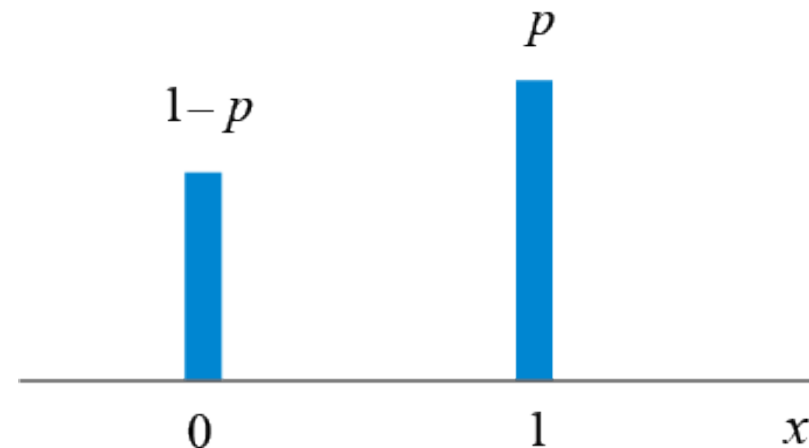
$$Z = \begin{cases} 1 & (\text{確率 } p \text{ で}) \\ 0 & (\text{確率 } 1 - p \text{ で}) \end{cases}$$

x	0	1	計
$P(Z = x)$	$1 - p$	p	1

数式だけで書くなら

$$P(Z = 0) = 1 - p$$

$$P(Z = 1) = p$$



平均値 $\mathbf{E}[Z] = 0 \times (1 - p) + 1 \times p = p$

$$\mathbf{E}[Z^2] = 0^2 \times (1 - p) + 1^2 \times p = p$$

分散 $\mathbf{V}[Z] = \mathbf{E}[Z^2] - \mathbf{E}[Z]^2 = p - p^2 = p(1 - p)$

平均値 $m = p$

分散 $\sigma^2 = p(1 - p)$

二項分布 $B(n, p)$

X : 表の出る確率 p のコインを n 回投げたときの表の回数

確率分布の計算

$P(X = k)$: $\bigcirc$ (表) が k 回, $\times$ (裏) が $n - k$ 回の確率

その一例 $\bigcirc \times \cdots \bigcirc \times \cdots \times \bigcirc \times$

その確率 $p q \cdots p q \cdots q p q = p^k q^{n-k}$ (ただし, $q = 1 - p$)

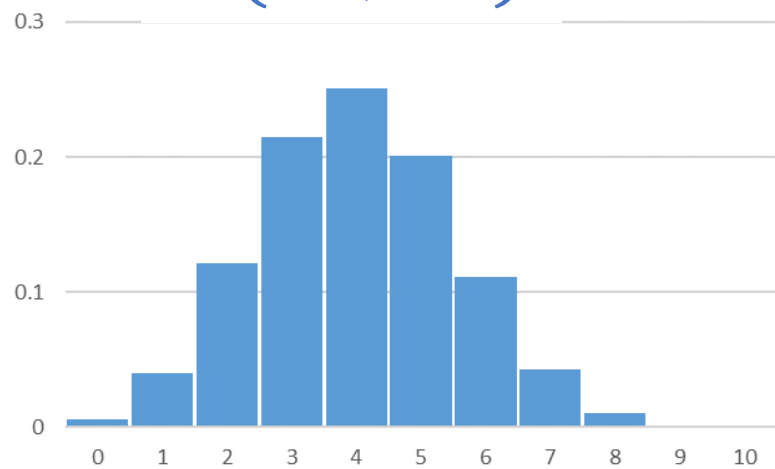
したがって,

二項係数

$$P(X = k) = \frac{n!}{k! (n - k)!} p^k (1 - p)^{n-k} = \binom{n}{k} p^k (1 - p)^{n-k}$$

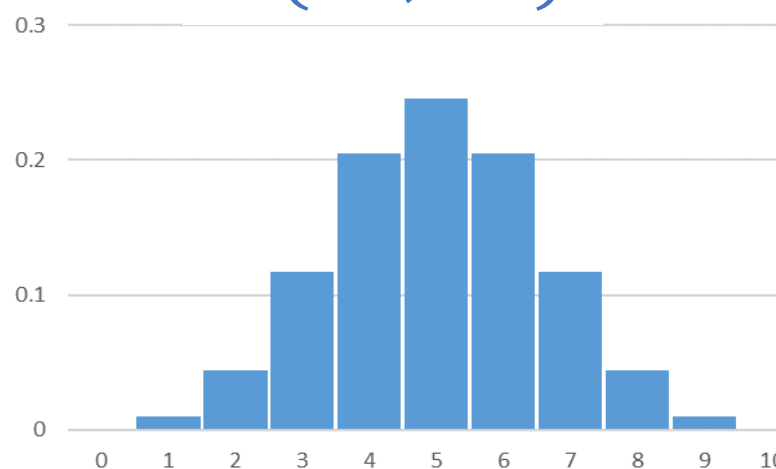
二項分布の具体例

$B(10, 0.4)$



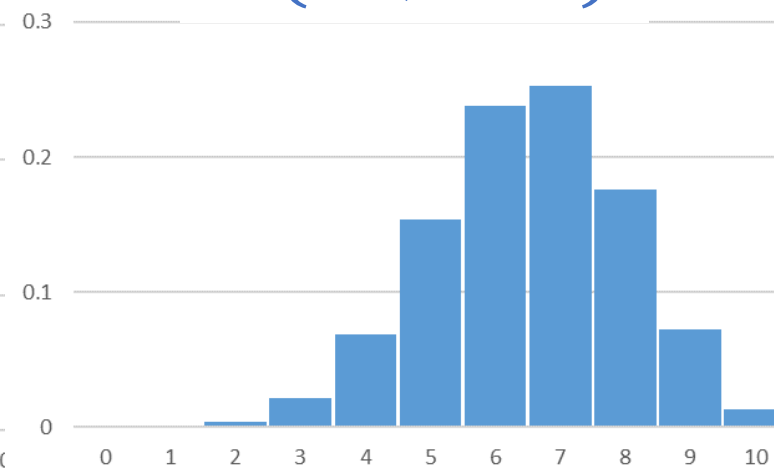
$$m = 4$$
$$\sigma^2 = 2.4$$

$B(10, 0.5)$



$$m = 5$$
$$\sigma^2 = 2.5$$

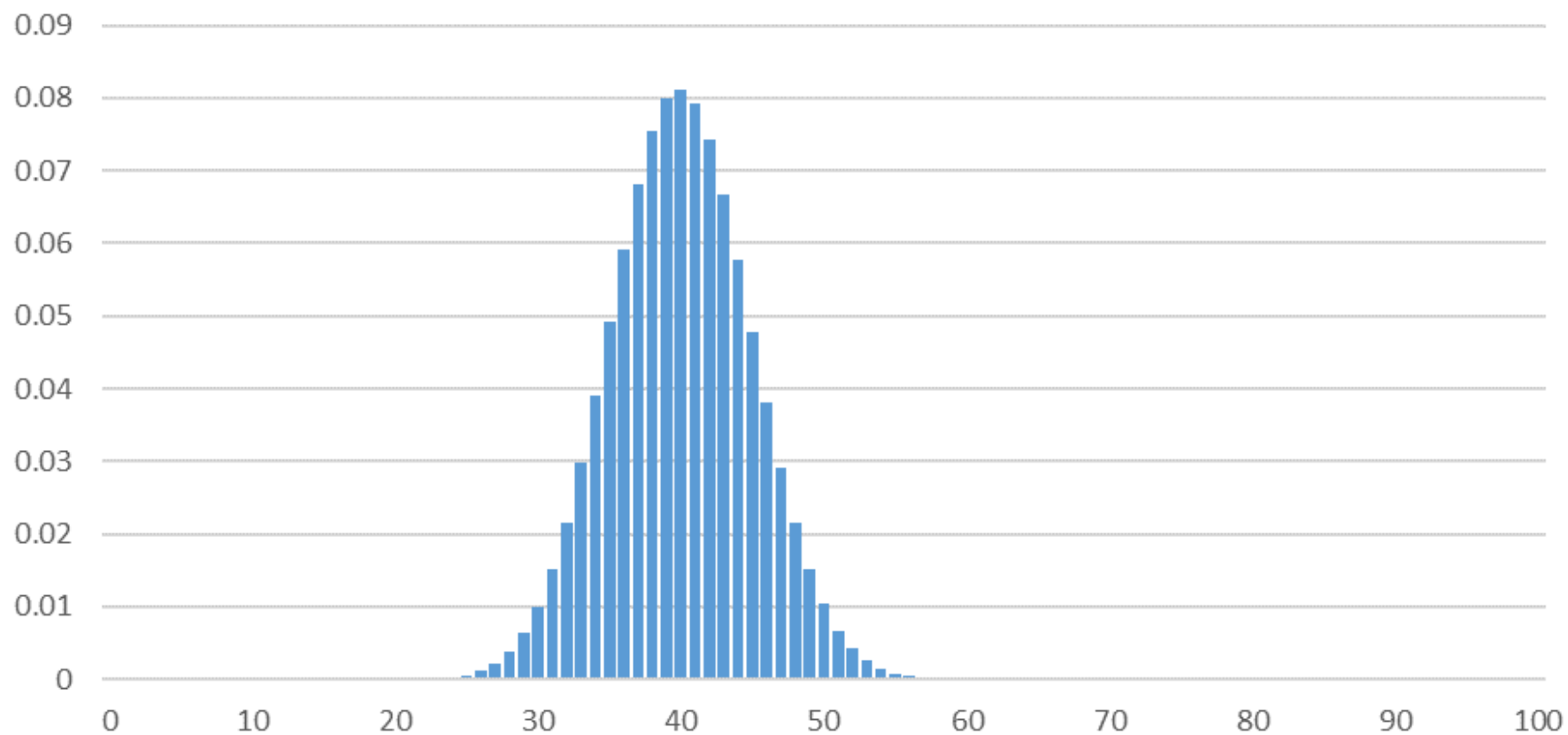
$B(10, 0.65)$



$$m = 6.5$$
$$\sigma^2 = 2.275$$

二項分布の具体例

$$B(100, 0.4) \quad m = 40, \quad \sigma^2 = 24$$



二項分布 $B(n, p)$ 平均値 $m = np$ 分散 $\sigma^2 = np(1 - p)$

$$\mathbf{E}[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathbf{E}[X^2] = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

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

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k}$$

※ $\mathbf{E}[X^2]$ のかわりに

$$\mathbf{E}[X(X-1)] = \sum_{k=0}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k}$$

を計算する方が簡単

和の計算

$$\begin{aligned}
& \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^n k \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^n \frac{n!}{(k-1)! (n-k)!} p^k (1-p)^{n-k} \\
&= n \sum_{k=1}^n \frac{(n-1)!}{(k-1)! (n-k)!} p^k (1-p)^{n-k}
\end{aligned}$$

$$\begin{aligned}
&= n \sum_{k=0}^{n-1} \frac{(n-1)!}{k! (n-(k+1))!} p^{k+1} (1-p)^{n-(k+1)} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k! (n-1-k)!} p^k (1-p)^{n-1-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} \\
&= np (p + (1-p))^{n-1} \\
&= np
\end{aligned}$$

和の計算

$$\begin{aligned} \sum_{k=0}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} &= \sum_{\substack{k=2 \\ \text{red}}}^n k(k-1) \frac{n!}{k! (n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=2}^n \frac{n!}{(k-2)! (n-k)!} p^k (1-p)^{n-k} = n(n-1) \sum_{k=2}^n \frac{(n-2)!}{(k-2)! (n-k)!} p^k (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} \frac{(n-2)!}{k! (n-2-k)!} p^k (1-p)^{n-2-k} \\ &= n(n-1)p^2 \sum_{k=0}^{n-2} \binom{n-2}{k} p^k (1-p)^{n-2-k} \\ &= n(n-1)p^2 (p + (1-p))^{n-2} = n(n-1)p^2 \end{aligned}$$

二項分布 $B(n, p)$

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

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\mathbf{E}[X] = \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k} = np$$

$$\begin{aligned} \mathbf{E}[X(X - 1)] &= \sum_{k=0}^n k(k - 1) \binom{n}{k} p^k (1 - p)^{n-k} \\ &= n(n - 1)p^2 \end{aligned}$$

$$\mathbf{E}[X^2] = \mathbf{E}[X(X - 1)] + \mathbf{E}[X]$$

$$= n(n - 1)p^2 + np$$

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

$$= n(n - 1)p^2 + np - (np)^2$$

$$= -np^2 + np$$

$$= np(1 - p)$$

$$\text{平均值} \quad m = np$$

$$\text{分散} \quad \sigma^2 = np(1 - p)$$

幾何分布

成功確率 p の試行を独立に繰り返す.
初めて成功するまでの失敗の回数を X とする.

$$P(X = k) = p(1 - p)^k \quad k = 0, 1, 2, \dots$$

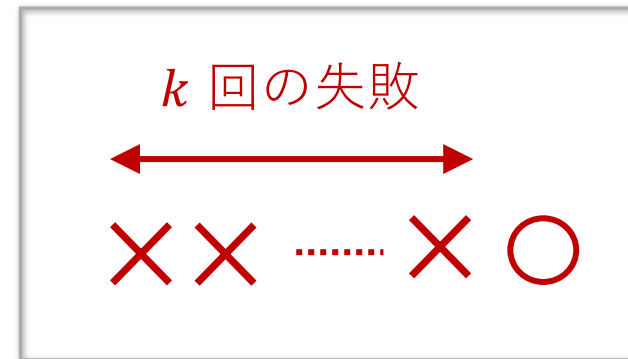
この分布をパラメータ p の幾何分布という.

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} kp(1-p)^k = \frac{1-p}{p} = \frac{1}{p} - 1$$

$$\mathbf{E}[X^2] = \sum_{k=0}^{\infty} k^2 p(1-p)^k = \frac{(2-p)(1-p)}{p^2}$$

$$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{1-p}{p^2}$$

事象 $\{X = k\}$



(注意) $X + 1$ は成功までの待ち時間

平均値	$m = \frac{1-p}{p}$
分散	$\sigma^2 = \frac{1-p}{p^2}$

確率母関数 (Generating function)

$n = 0, 1, 2, \dots$ に値をとる確率変数 X に対して $p_n = P(X = n)$ とおく

確率母関数 $f(x) = \sum_{n=0}^{\infty} p_n x^n$

$$f'(x) = \sum_{n=0}^{\infty} n p_n x^{n-1}$$

$$f''(x) = \sum_{n=0}^{\infty} n(n-1) p_n x^{n-2}$$

$$f'(1) = \sum_{n=0}^{\infty} n p_n = E[X] \quad \text{平均値}$$

$$f''(1) = \sum_{n=0}^{\infty} n(n-1) p_n = \sum_{n=0}^{\infty} n^2 p_n - \sum_{n=0}^{\infty} n p_n$$

$$= E[X^2] - f'(1)$$

$$E[X^2] = f''(1) + f'(1)$$

分散 $V[X] = E[X^2] - E[X]^2$

$$= f''(1) + f'(1) - f'(1)^2$$

二項分布の確率母関数

確率母関数

$$\begin{aligned} f(x) &= \sum_{k=0}^{\infty} p_k x^k \\ &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} x^k \\ &= \sum_{k=0}^n \binom{n}{k} (px)^k (1-p)^{n-k} \\ &= (px + 1 - p)^n \end{aligned}$$

$$f'(x) = np(px + 1 - p)^{n-1}$$

$$f'(1) = np$$

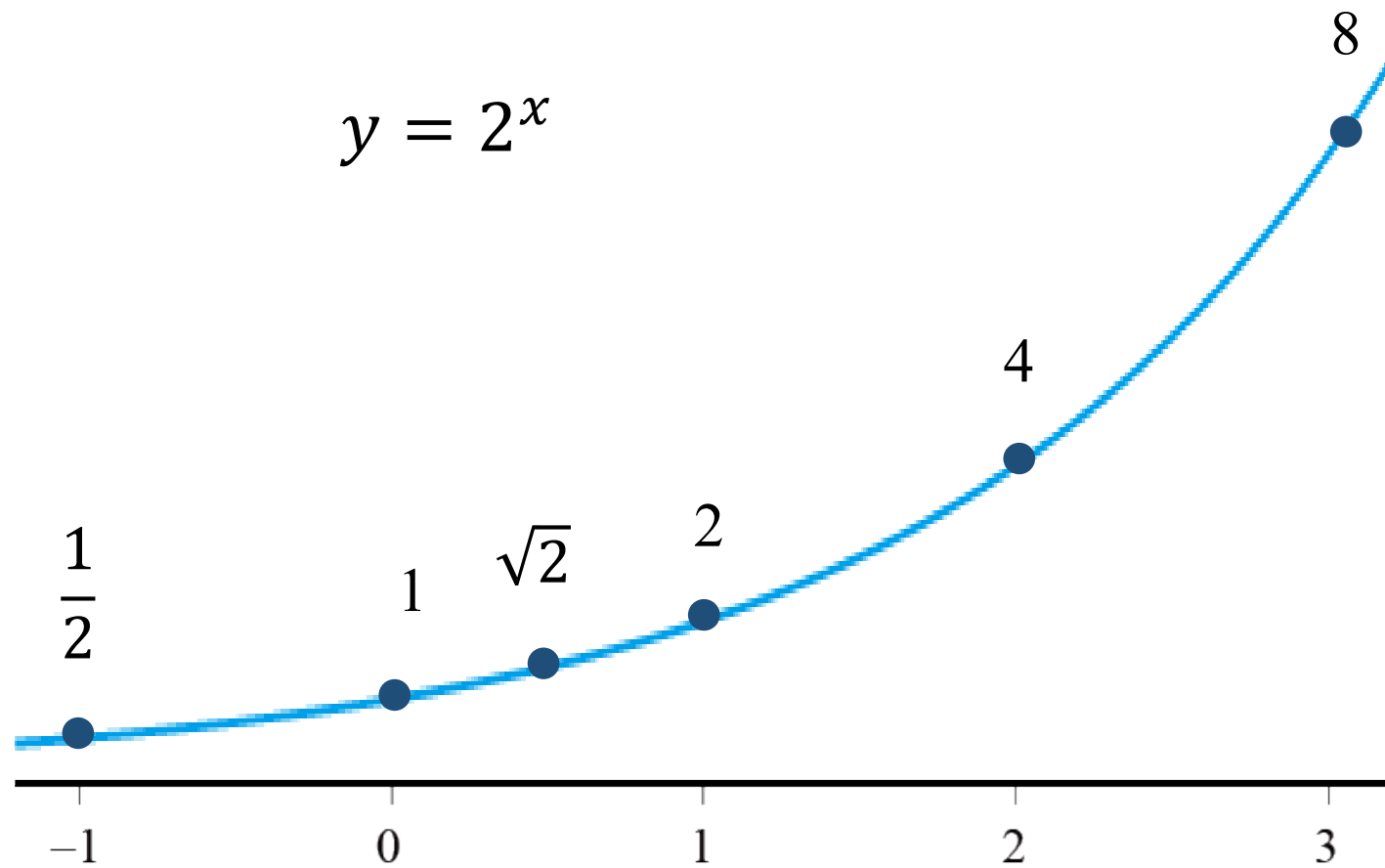
$$f''(x) = n(n-1)p^2(px + 1 - p)^{n-2}$$

$$f''(1) = n(n-1)p^2$$

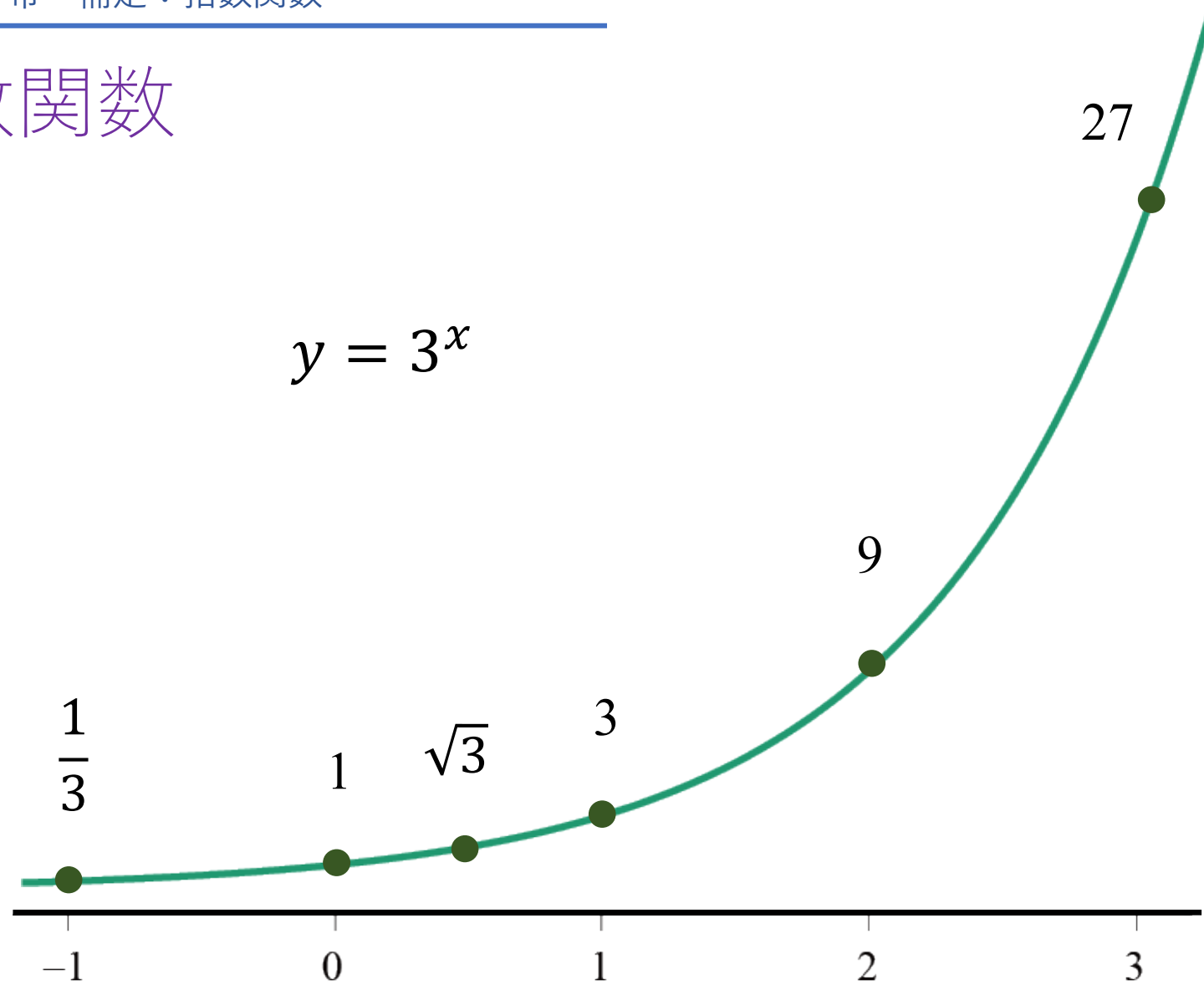
平均値 $E[X] = f'(1) = np$

分散
$$\begin{aligned} V[X] &= f''(1) + f'(1) - f'(1)^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= np \end{aligned}$$

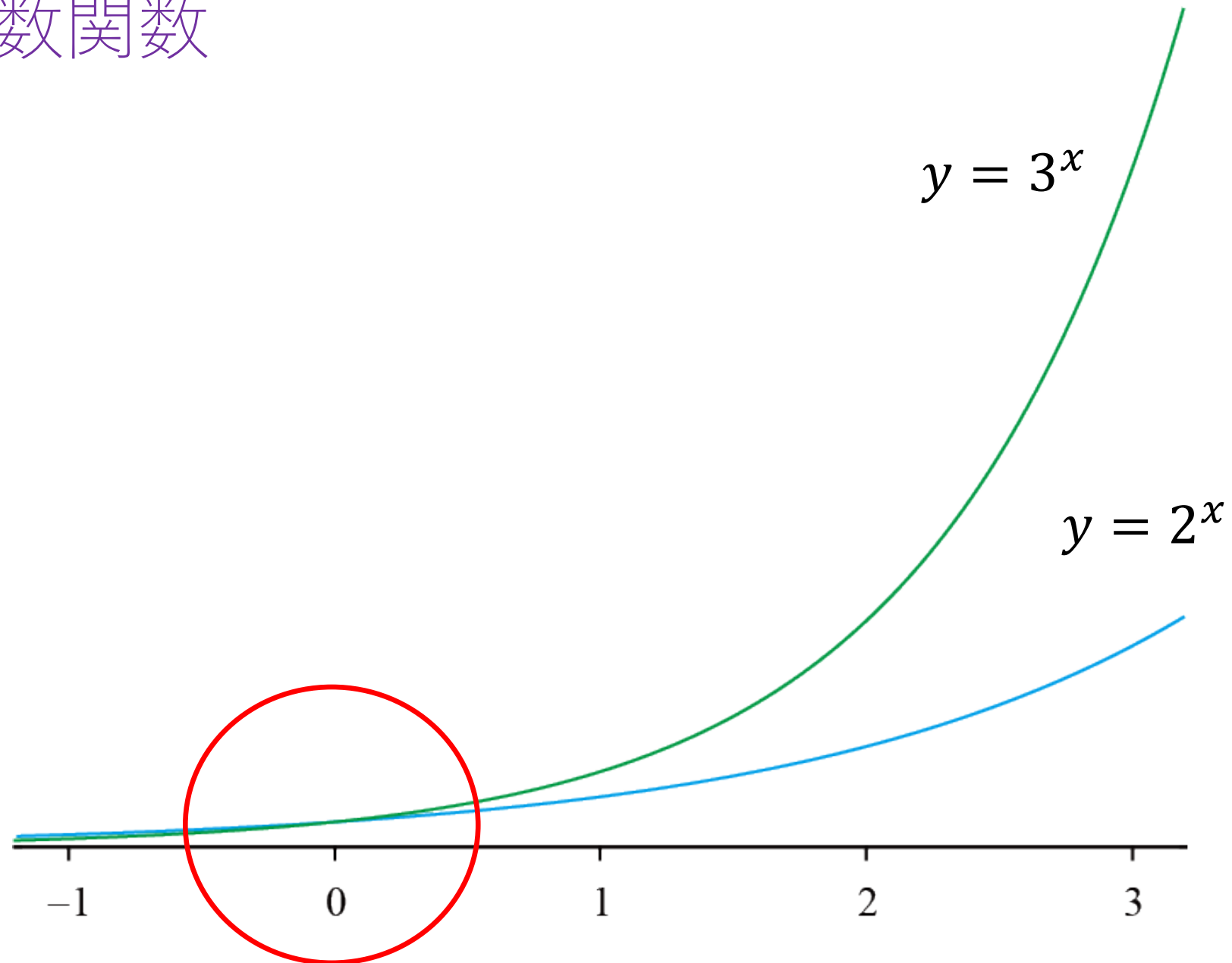
補足：指数関数



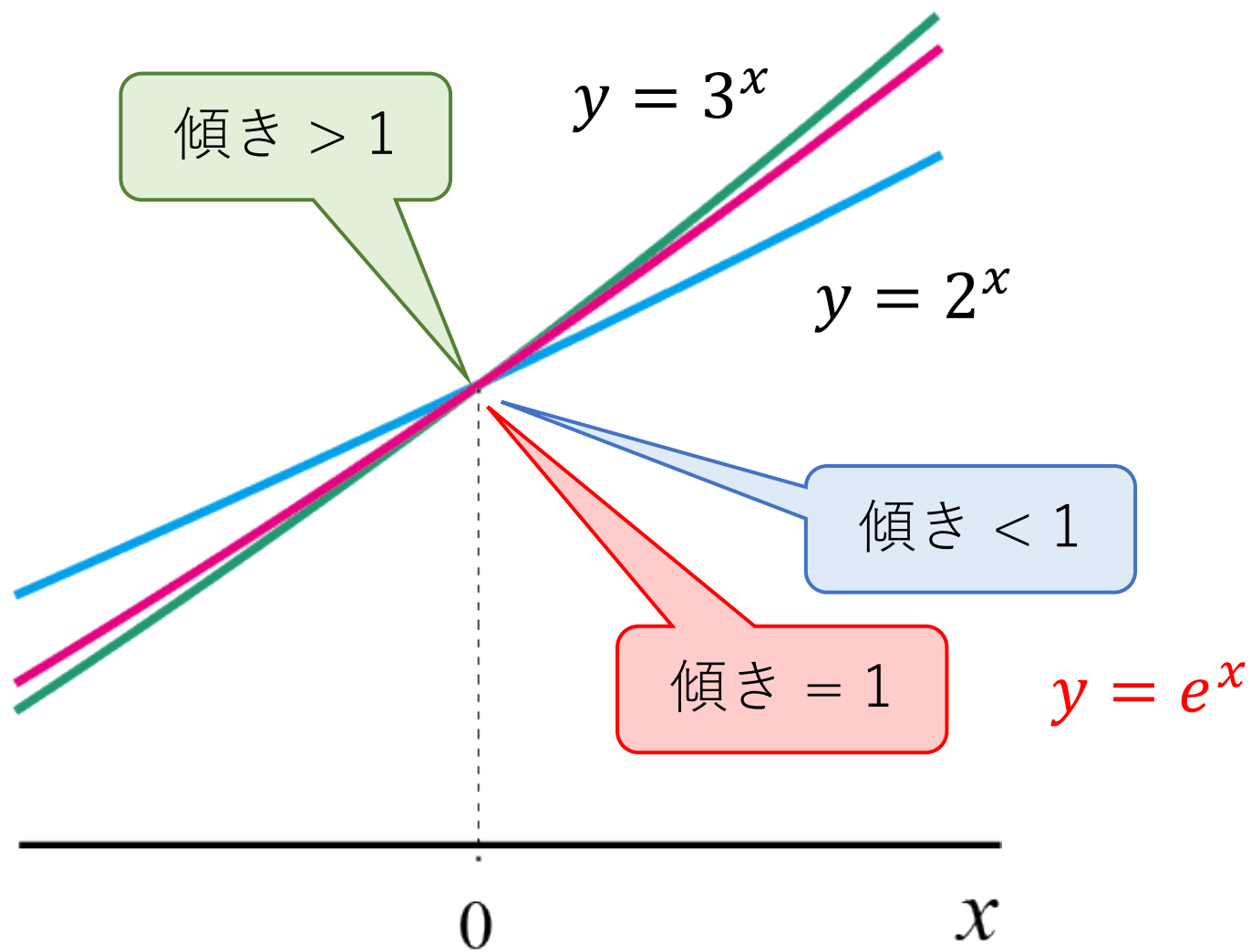
補足：指数関数



補足：指数関数

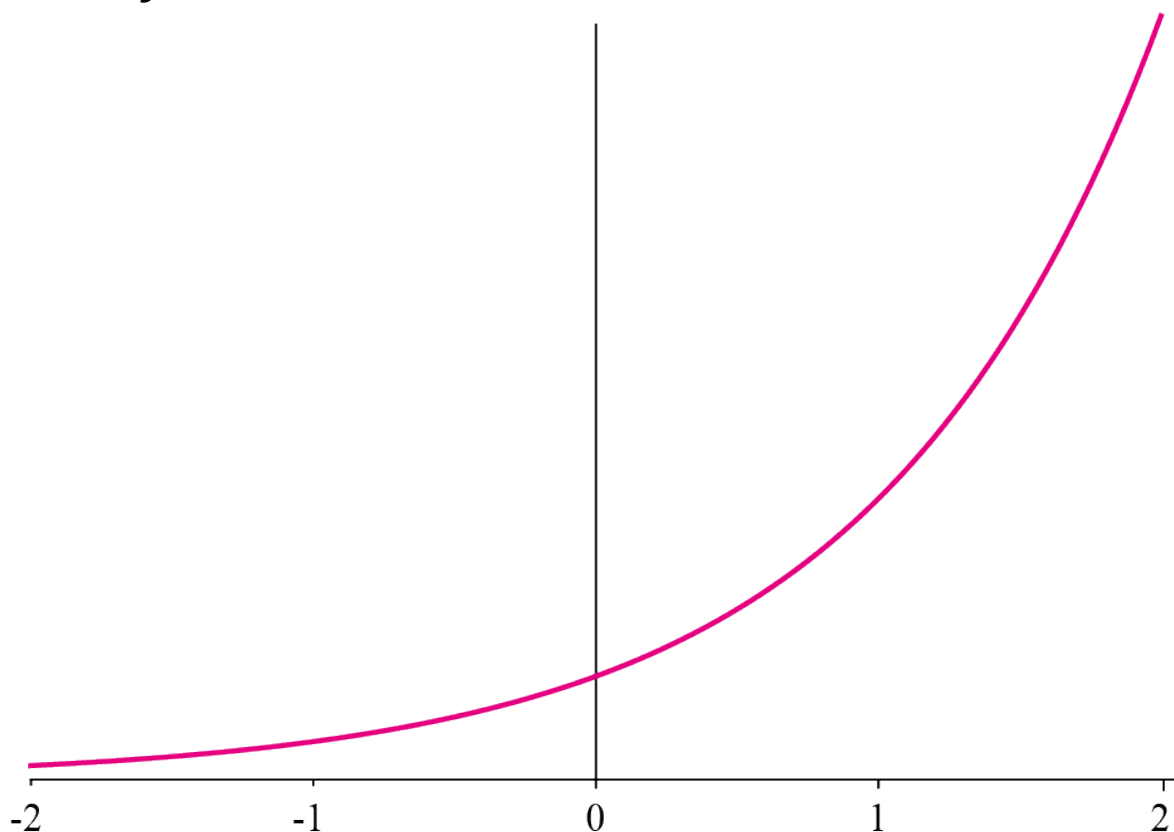


補足：指数関数



補足：指数関数

$$y = e^x$$



$$e = 2.718 \dots \text{ (無理数)}$$

e の呼び名

- 自然対数の底
- ネピアの数

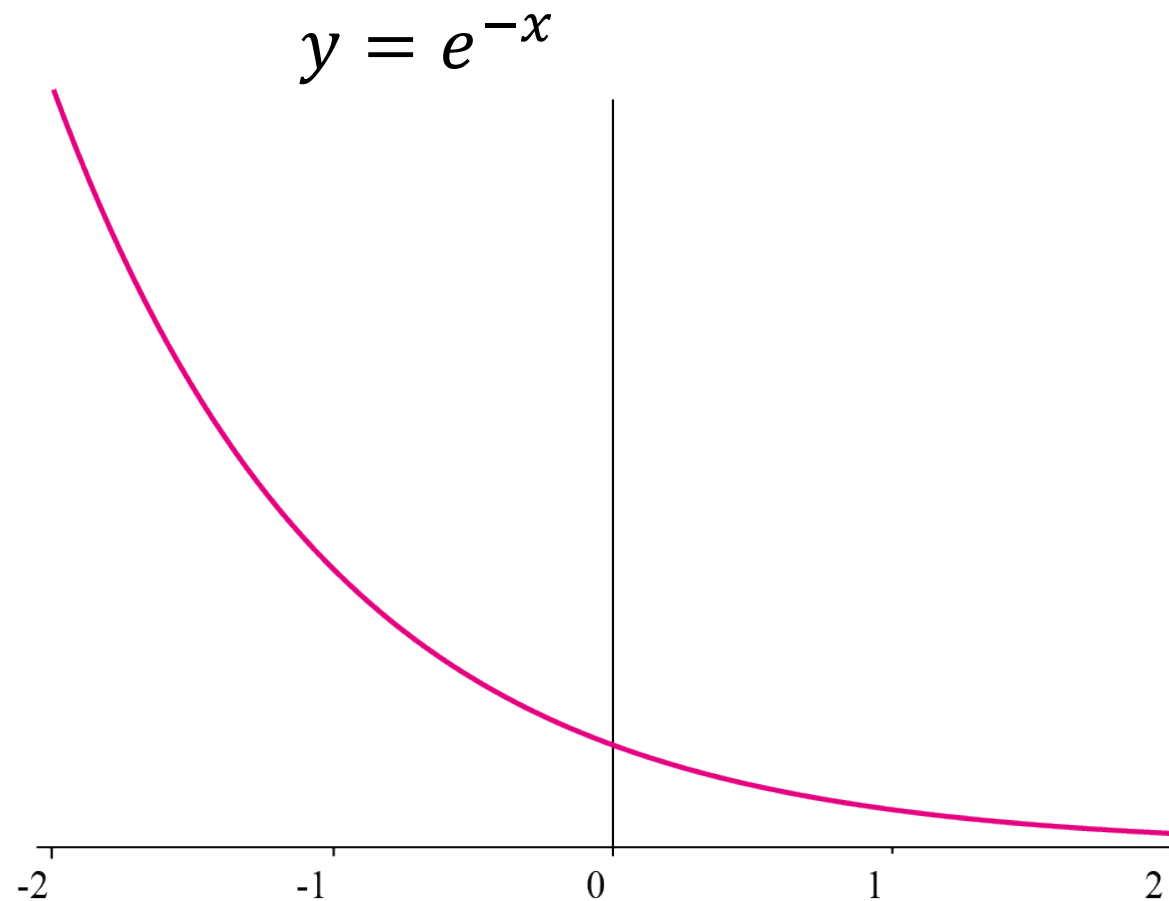
テーラー展開

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

指数法則

$$e^{x+y} = e^x e^y$$

補足：指数関数



$e = 2.718 \dots$ (無理数)

e の呼び名

- 自然対数の底
- ネピアの数

テーラー展開

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

指数法則

$$e^{x+y} = e^x e^y$$

ポアソン分布

$\lambda > 0$ を定数とする.

確率変数 X がパラメータ λ のポアソン分布に従うとは,

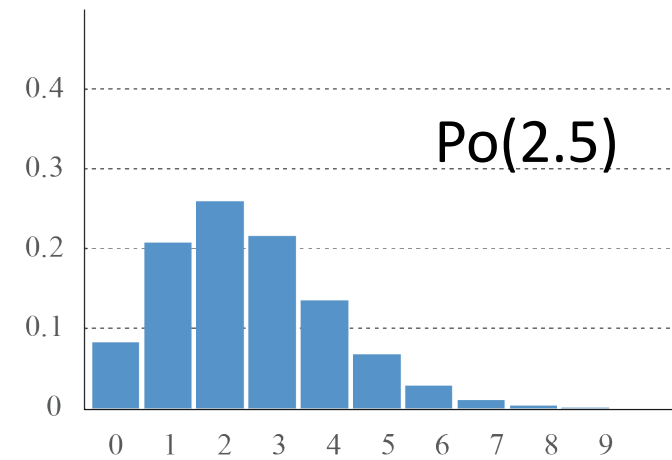
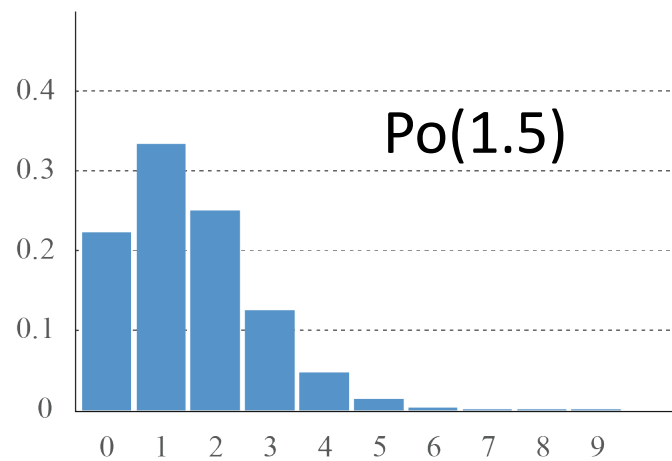
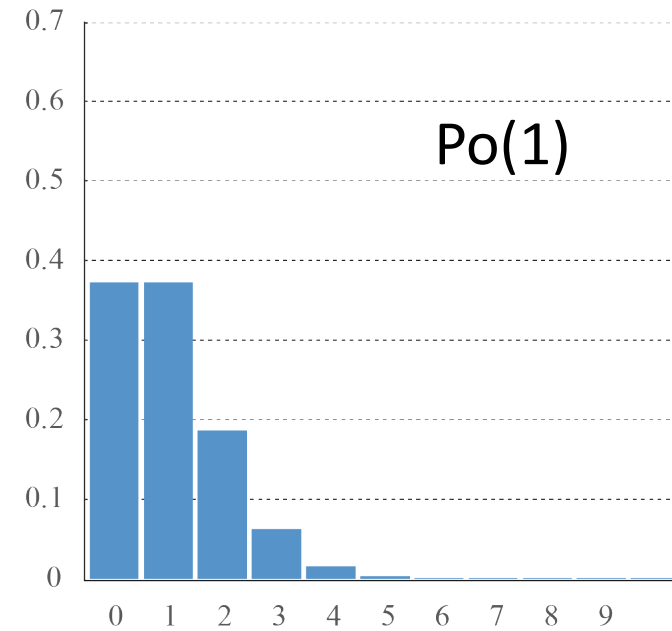
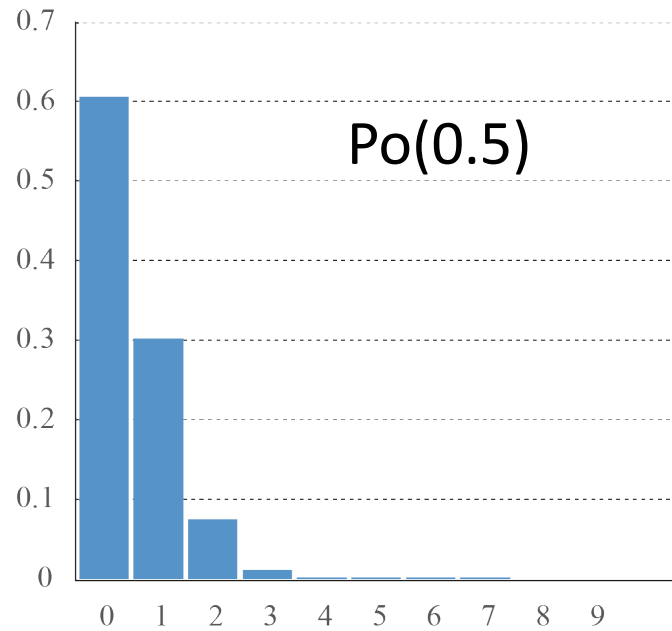
$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots$$

$X \sim \text{Po}(\lambda)$ と書く.

➤ 確かに確率分布になっている

指数関数のテーラー展開

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \quad \longrightarrow \quad 1 = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$$



ポアソン分布の確率母関数

確率分布

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

確率母関数

$$\begin{aligned} f(x) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} x^k \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda x)^k}{k!} \\ &= e^{-\lambda} e^{\lambda x} \end{aligned}$$

$$f'(x) = \lambda e^{-\lambda} e^{\lambda x}$$

$$f'(1) = \lambda$$

$$f''(x) = \lambda^2 e^{-\lambda} e^{\lambda x}$$

$$f''(1) = \lambda^2$$

平均値 $m = f'(1) = \lambda$

分散 $\sigma^2 = f''(1) + f'(1) - f'(1)^2$
 $= \lambda^2 + \lambda - \lambda^2 = \lambda$

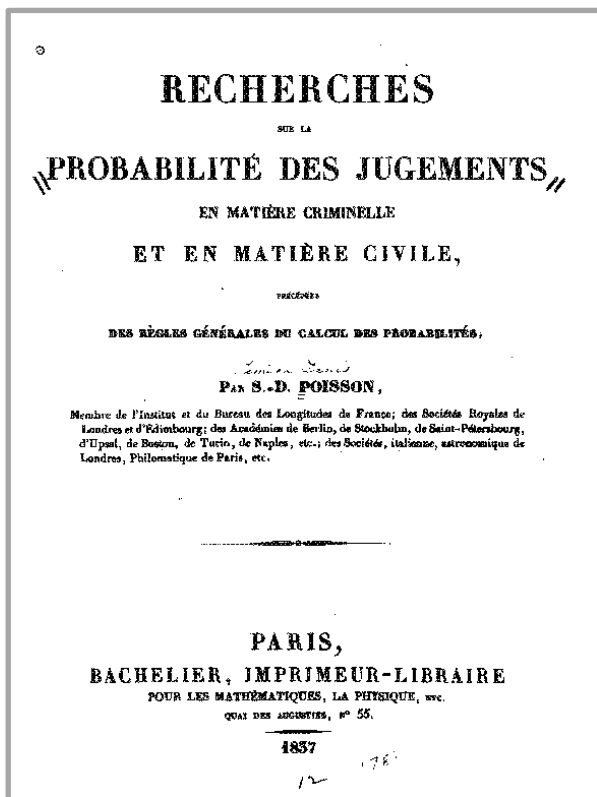
ポアソン分布の特徴

平均と分散が等しい

Siméon Denis Poisson (1781-1840)

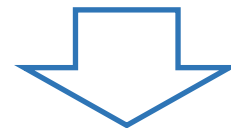


- エコール・ポリテクニクでラプラスらに学ぶ.
- 数学・物理学に多大な貢献
- ポアソン〇〇と名のついた概念多数



誤った有罪判決の回数について研究(1837)

誤審が, ある一定期間に起こる回数 X の確率分布
($\Rightarrow$ ポアソン分布) を導出

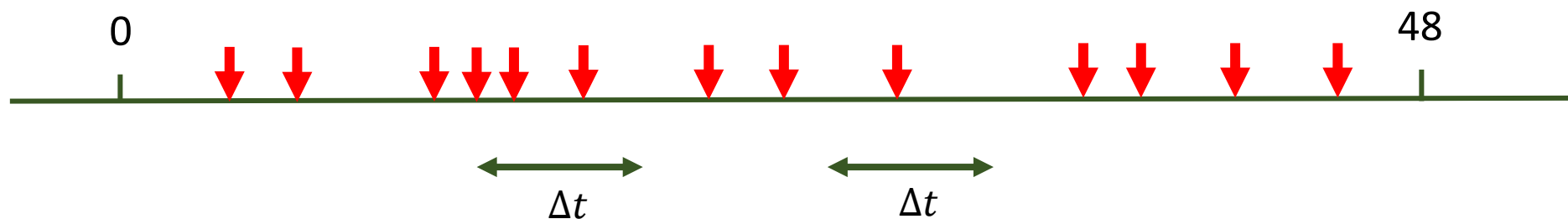


稀なイベントの発生回数の確率分布として広い応用

- 馬に蹴られて死亡した兵士数 (ボルトキーヴィッチ)
- 電話の呼び出し, メールの着信
- サッカーのゴール数, 野球のホームラン数

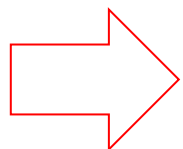
メールの着信

メールの着信を観測して, **48時間**に**324回**のメール着信があった. これをもとに, ある特定の1時間のメール着信回数を調べよう.

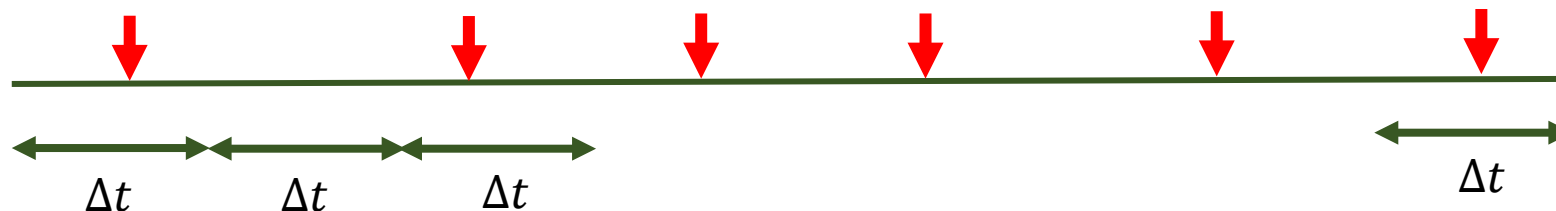


仮定

- 微小時間区間 Δt の中では着信は1回, または0回
- 微小時間区間 Δt がずれていれば着信の有無は独立



微小時間区間 Δt 毎のメール着信をコイン投げとみなす

1時間当たりの着信回数 X をモデル化する

- 1時間を微小時間区間 Δt に N 分割

$$\Delta t = \frac{1}{N}$$

$p = P(\text{時間区間 } \Delta t \text{ にメール着信がある})$

$1 - p = P(\text{時間区間 } \Delta t \text{ にメール着信がない})$

- 1時間の間に N 回のコイン投げ

表の回数 $X \sim B(N, p)$

平均値 $\mu = Np$

- 観測から1時間当たり

$$\frac{324}{48} = 6.75 \text{ 回の着信}$$

- p の推定

$$\mu = Np = 6.75 \quad \Rightarrow \quad p = \frac{6.75}{N}$$

- こうして

$$X \sim B\left(N, \frac{6.75}{N}\right)$$

ポアソン分布の導出 (ポアソンの少数の法則)

$B\left(N, \frac{\lambda}{N}\right)$ は $N \rightarrow \infty$ とすると $Po(\lambda)$ に収束する.

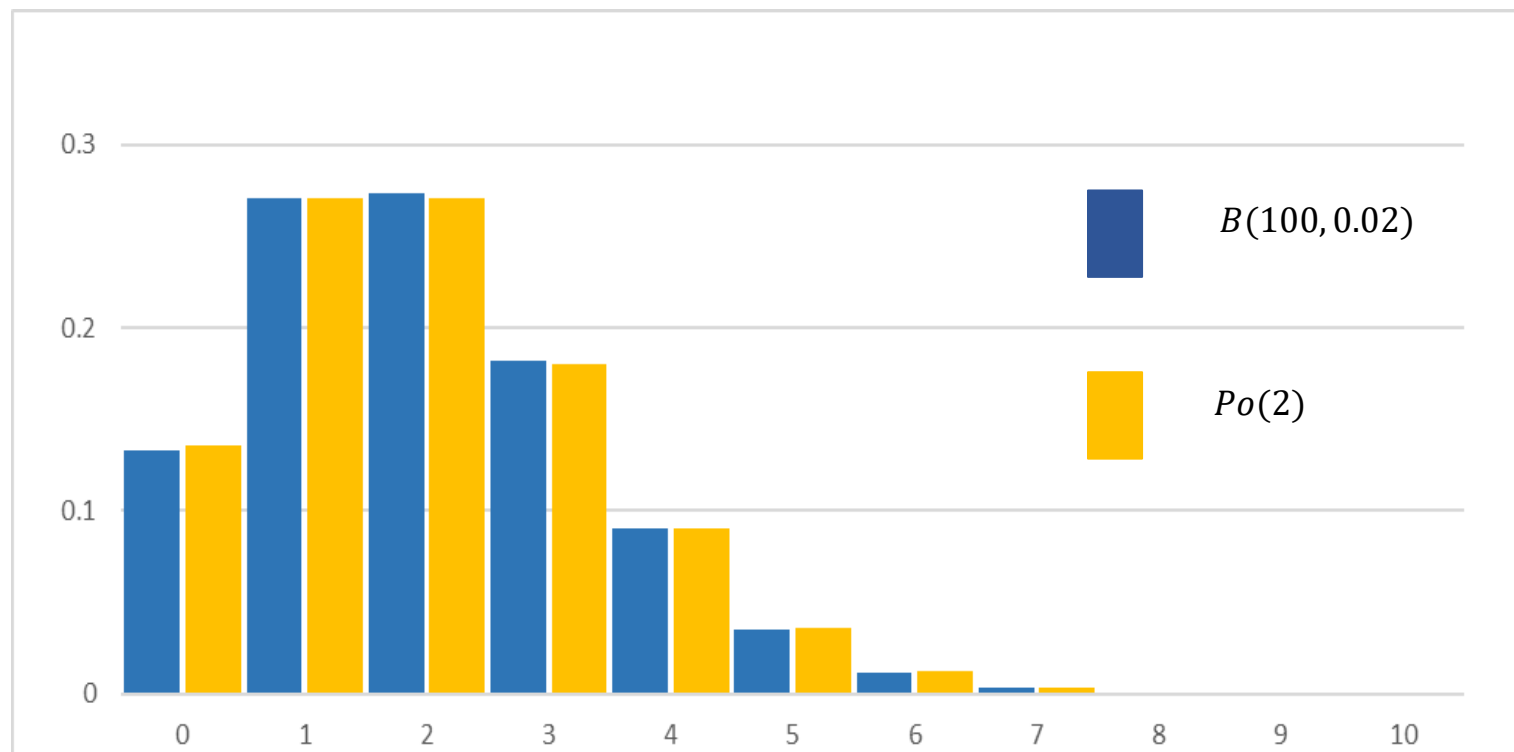
言い換え $B(N, p)$ は, $N \rightarrow \infty, p \rightarrow 0, Np \rightarrow \lambda$ とすると $Po(\lambda)$ に収束する.

証明 $X \sim B\left(N, \frac{\lambda}{N}\right)$ として

$$\begin{aligned} P(X = k) &= \binom{N}{k} p^k (1 - p)^{N-k} = \frac{N(N-1)\cdots(N-k+1)}{k!} \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k} \\ &\rightarrow \frac{\lambda^k}{k!} e^{-\lambda} \quad (N \rightarrow \infty) \end{aligned}$$

例 $B(100, 0.02) \approx Po(2)$

k	0	1	2	3	4	5	6	7	8	9
二項分布	0.13262	0.27065	0.27341	0.18228	0.09021	0.03535	0.01142	0.00313	0.00074	0.00015
ポアソン分布	0.13534	0.27067	0.27067	0.18045	0.09022	0.03609	0.01203	0.00344	0.00086	0.00019



メールの着信

メールの着信を観測して, 48時間に324回のメール着信があった. これをもとに, ある特定の1時間のメール着信回数を調べよう.

➤ X : 1時間当たりの着信回数

➤ 1時間を微小時間区間 Δt に N 分割してモデル化: $X \sim B\left(N, \frac{6.75}{N}\right)$

➤ $N \rightarrow \infty$ として $X \sim \text{Po}(6.75)$

➤ たとえば

$$\begin{aligned} P(X \leq 2) &= \frac{6.75^0}{0!} e^{-6.75} + \frac{6.75^1}{1!} e^{-6.75} + \frac{6.75^2}{2!} e^{-6.75} \\ &= (1 + 6.75 + 22.78) \times 0.00117 = 0.036 \end{aligned}$$

N は人為的なので
消去したい

Python を試してみる

二項分布

- 確率計算のコード
- 確率分布表
- 確率分布の図示

ポアソン分布

- 確率計算のコード
- 確率分布表
- 確率分布の図示

[Binomial and Poisson - Jupyter Notebook.pdf](#)

ポアソンの少数の法則

$$B(N, p) \approx Po(\lambda)$$

$$N \rightarrow \infty, p \rightarrow 0, Np \rightarrow \lambda$$

[PoissonLimit - Jupyter Notebook.pdf](#)