

意味がわかる統計学

p.128-202

60216114 作島大智

目次

- 確率変数とは
- 連続型と離散型
- 二項分布
- 二項分布と正規分布の関係

確率変数とは

X の値を定めるとそれに対応する
確率の値が決まるもの

EX) 2枚コインを投げ,表が X 枚の時の
確率立数を $P(X)$ とするとき

$$P(X=0) = 1/4$$

$$P(X=1) = 1/2$$

$$P(X=2) = 1/4$$

全確率

起こりうるすべての場合のうち
どれかが起こる可能性は「1」であること

確率変数 X とは、

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

ただし、 p_1 、 p_2 、 $\cdots$ 、 p_n はすべて 0 以上で、

$$p_1 + p_2 + \cdots + p_n = 1$$

という表で与えられる「 X と確率の対応の決まり」のことです。

確率変数の種類

離散型確率変数

- コイン

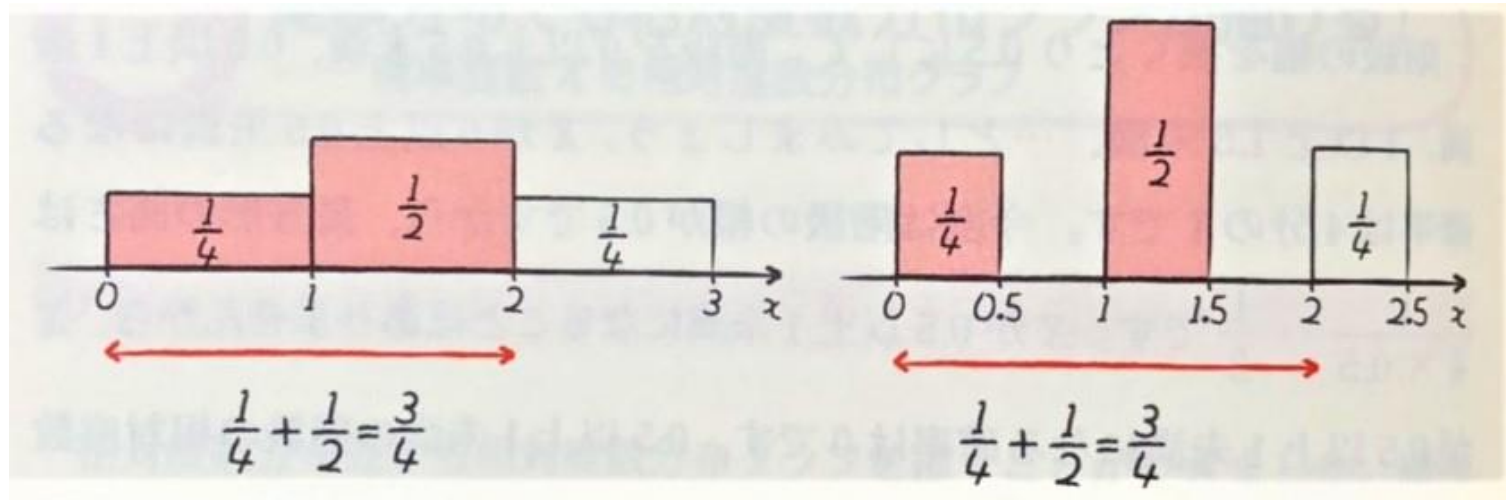
連続型確率変数

- ルーレット

離散型

階級の幅を小さくすると棒が間引かれて隙間ができる.これはもともと X の取りうる値がとびとびだからである.

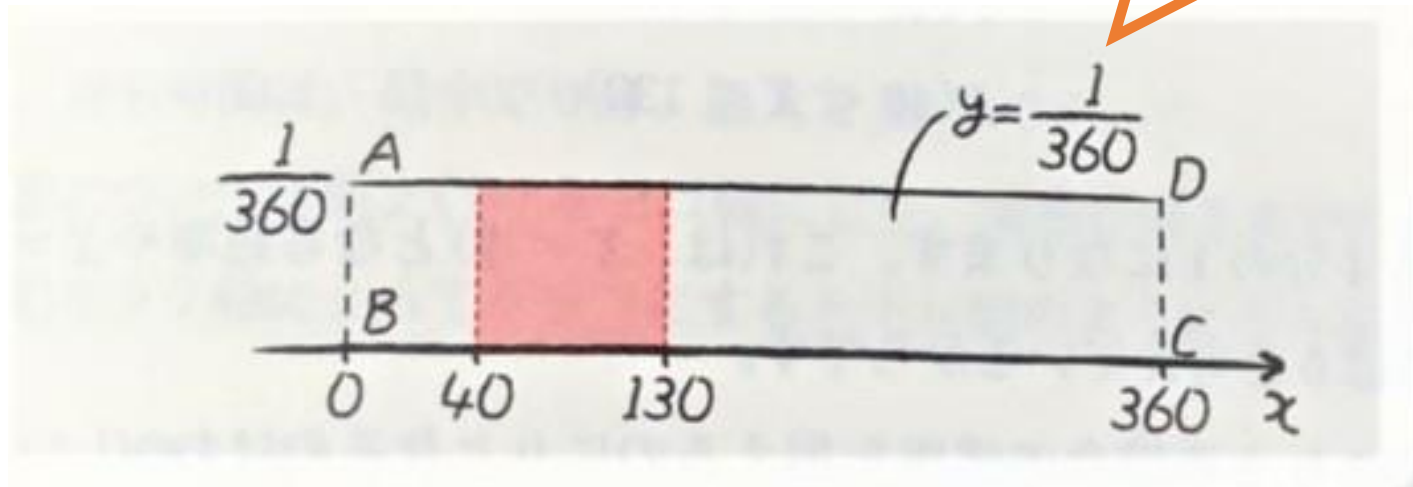
→ 離散型という.



連続型

連続型の確率変数の分布を表すグラフを示す
 曲線を**確率密度関数**をいう。
 一様分布ともいう。

これが確率密度関数
 直線は曲線の一種

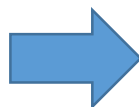


連続型確率分布の特徴

- $P(40 \leq X < 130)$

- $P(40 < X < 130)$

- $P(40 < X \leq 130)$



すべて同じ確率

1点での確率は0になる

$$\rightarrow 1/\text{全事象} \Leftrightarrow 1/\infty \rightarrow 0$$

例題

a 駅の時刻表では、何時台であっても00分と15分に電車が到着する.

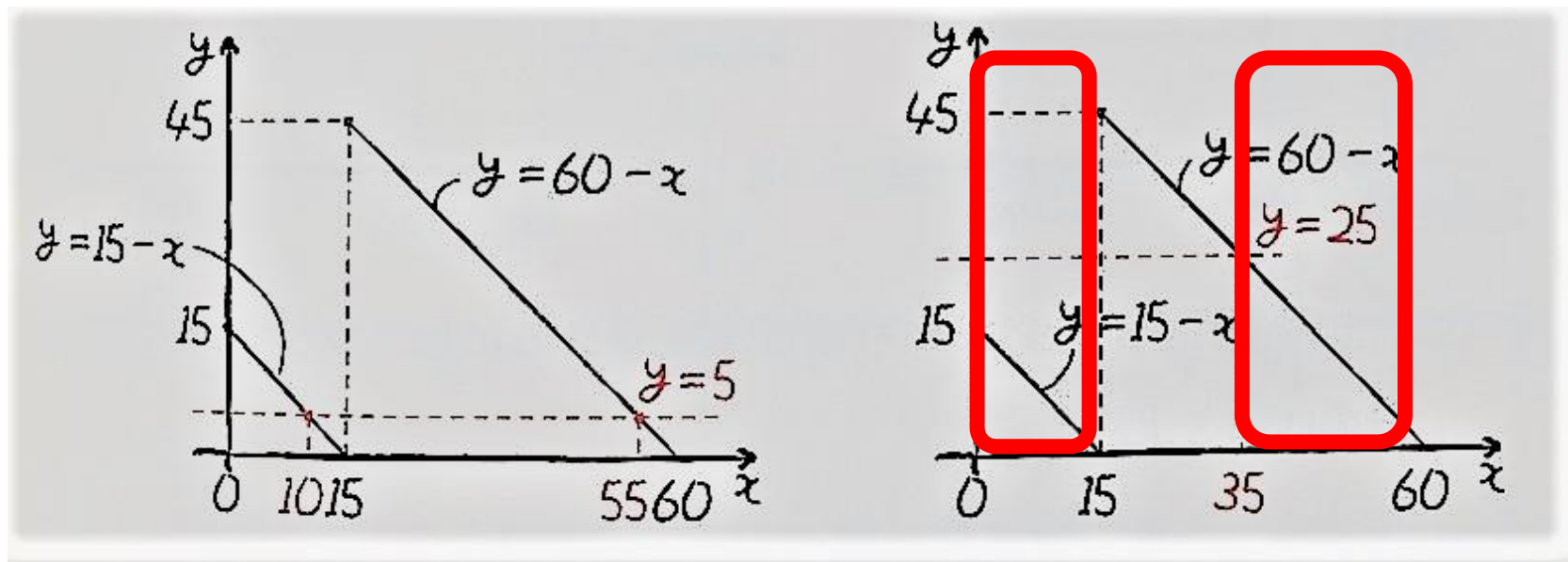
ある人が a 駅に着いたとき電車が来るまでの待ち時間を X とする.

(1) X の確率分布を求めグラフで表せ.

(2) 待ち時間が10分以上20分未満である確率を求めよ.

(1) X の確率分布

待ち時間の最小値は0,最大値は45

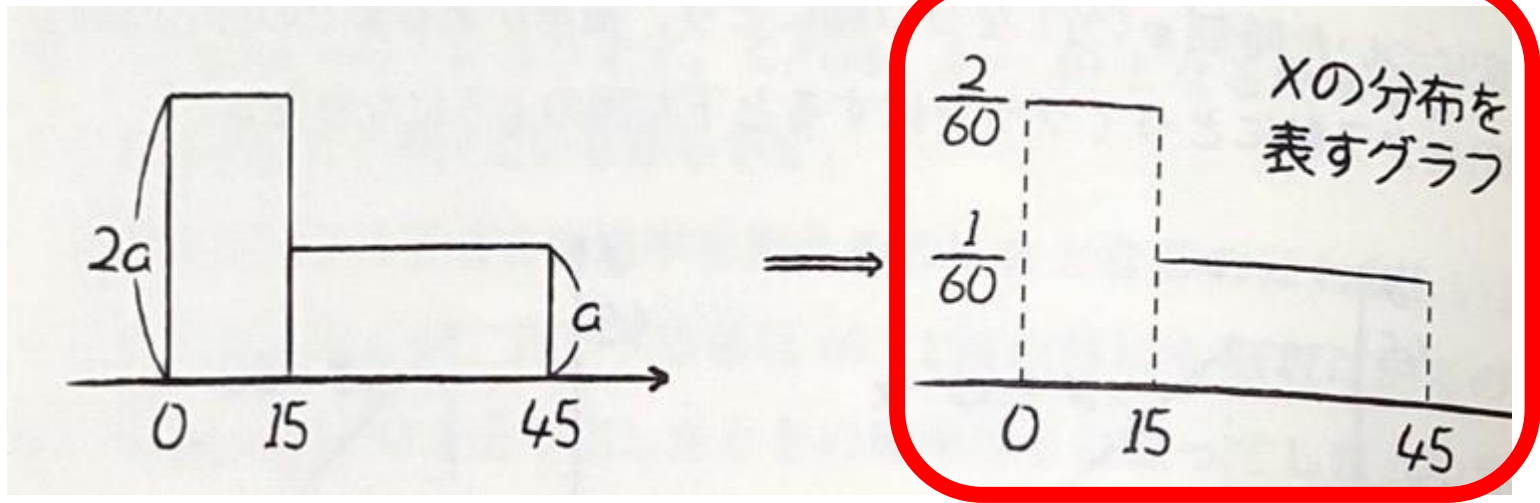


$0 < X < 45$ において $P(0 < X < 15)$ は $P(15 < X < 45)$ の2倍

(1) X の確率分布

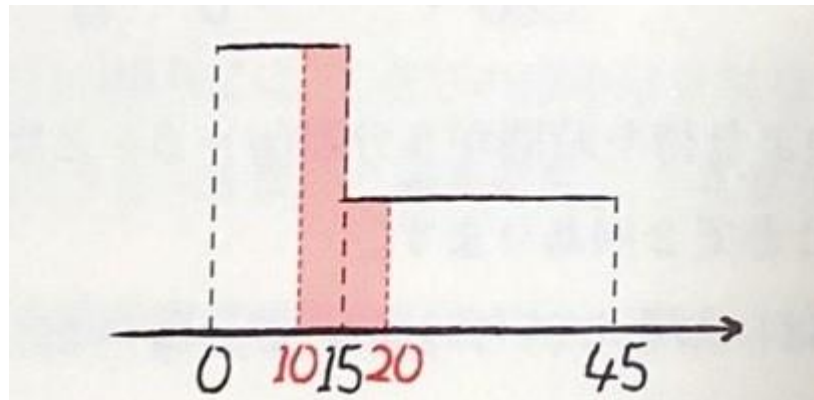
$$2a \cdot 15 + a(45 - 15) = 60a$$

$$60a = 1 \Leftrightarrow a = 1/60$$



(2) 待ち時間

X が10以上20以下になる確率を求めるには
グラフの赤い部分を求める.



$$\text{面積} = 2/60 * (15-10) + 1/60 * (20-15) = 1/4 \text{ より}$$

$$P(10 \leq X \leq 20) = 1/4$$

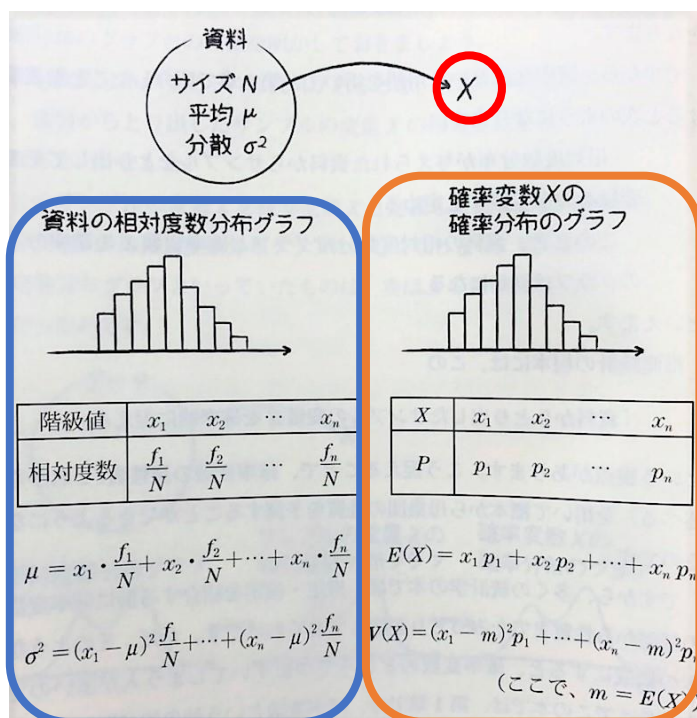
$E(X)$ と $V(X)$

- 確率変数 X
平均 X を $E(X)$,分散を $V(X)$ とする.
- 相対度数分布から計算した
平均した $E(X)$,分散を $V(X)$ と同じ記号を使う.

同じ記号を用いるのはなぜか.

サンプル X と確率変数 X

資料から取り出したサンプルの変量を
確率変数 X とするとき



x_i に対する相対度数 $\frac{f_i}{N}$ と
 $X=x_i$ が起こりうる確率 p_i
が等しい

X の平均・分散はサンプルを取り
出す資料の平均・分散に等しい

$$\mu = E(X), \sigma^2 = V(X)$$

確率

確率変数 X の平均 $E(X)$ を期待値とも呼ぶ。

資料のデータ $\rightarrow$ X の相対度数分布 $\rightarrow$ X の確率分布
過去のデータ 未来のデータ 確率

資料の平均 μ はただ平均と呼ぶ。

連続確率変数の 平均・分散の定義

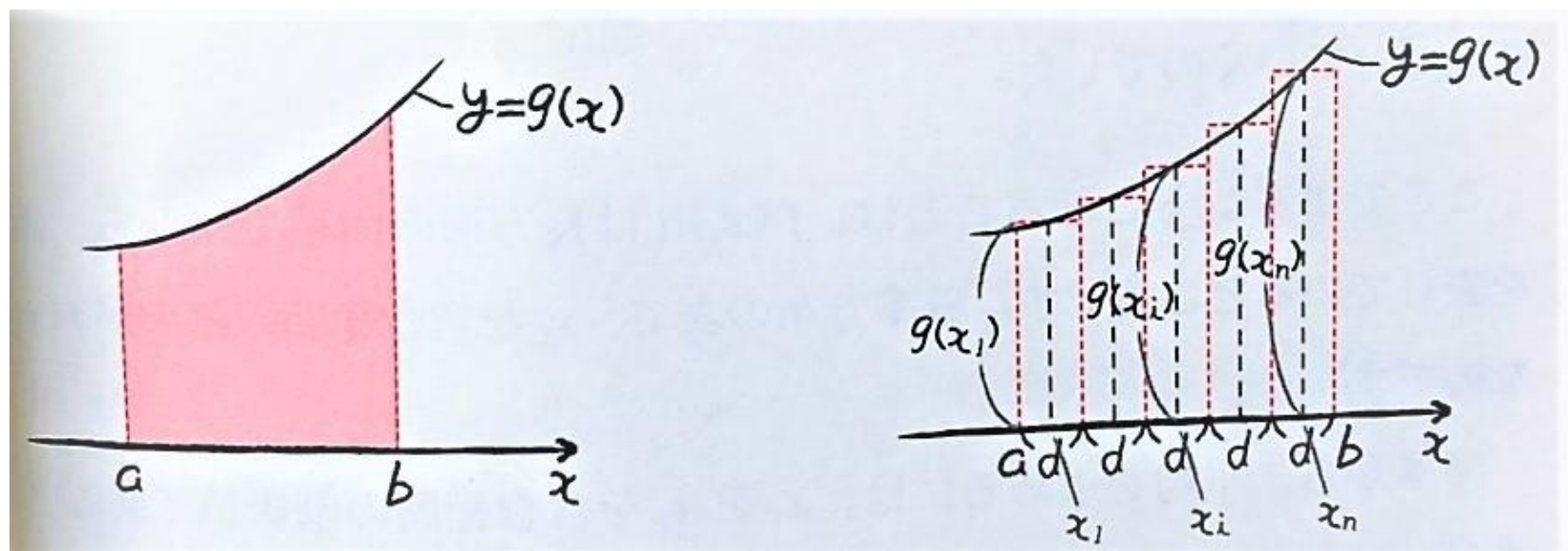
確率変数 X が確率密度関数 $f(x)$ に従うとき、平均は、

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

分散は、 $m = E(X)$ において、

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

連続確率変数の 平均・分散の定義



$E(V), V(X)$ の公式

確率変数 X, Y に対して

- $E(aX+b) = aE(X) + b$
- $E(X+Y) = E(X) + E(Y)$
- $V(X) = E(X^2) - \{E(X)\}^2$
- $V(aX+b) = a^2V(X)$

X の目の出方との Y 目の出方が
影響しあわない, 無関係の時

特に X, Y が独立な確率変数の場合

- $E(XY) = E(X)E(Y)$
- $V(X+Y) = V(X) + V(Y)$

二項分布

- 組み合わせの個数を示す **C** (Combination)

$${}_nCk = \frac{n(n-1) \dots n(k+1)}{k(k-1) \dots 1}$$

${}_nCk$:異なる n 個のものから k 個のものを選ぶ方法

例題

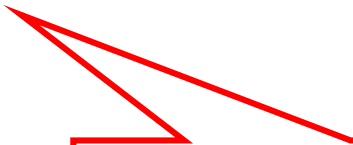
サイコロを5回投げたとき
5回中2回だけ3の倍数が出る確率を求めよ.

$$\begin{aligned} {}^5C_2 \left(\frac{2}{6}\right)^2 \left(\frac{4}{6}\right)^3 &= \frac{5 * 4}{2 * 1} * \left(\frac{1}{9}\right) \left(\frac{8}{27}\right) \\ &= \frac{80}{243} \end{aligned}$$

二項分布の平均・分散

X が二項分布に $Bin(n, p)$ に従うとき

- $E(X) = np$
- $V(X) = np(i - p)$



二項分布orベルヌーイ分布
 Bin は $Binary$ が由来

二項分布

結果が成功か失敗のいずれかである n 回の独立な
試行を行ったときの成功数で表される離散確率分布

一回目の試行

X_1	0	1
P	$1 - p$	p

$$E(X_1) = 0 * (1 - p) + 1 * p = p$$

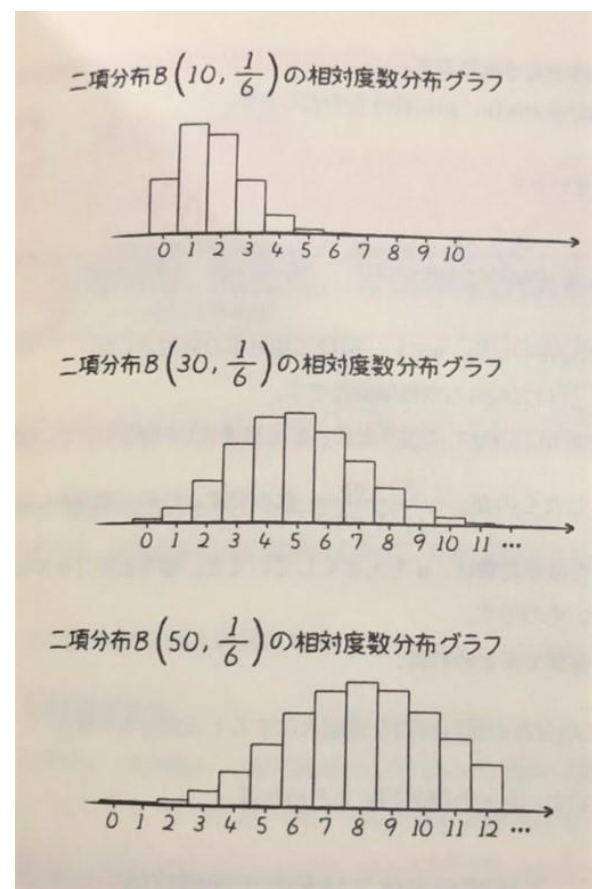
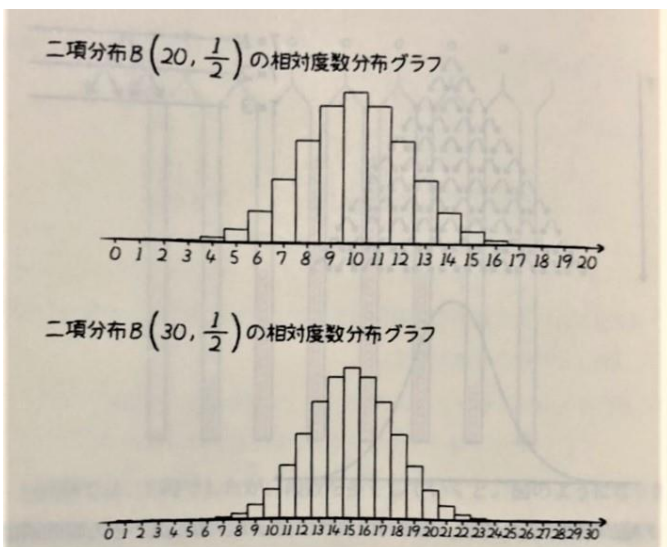
$$V(X_1) = (0 - p)^2 * (1 - p) + (1 - p)^2 * p = p(1 - p)$$

4回目の試行

- $E(X) = E(X_1 + X_2 + X_3 + X_4)$
 $= E(X_1) + E(X_2) + E(X_3) + E(X_4)$
 $= p + p + p + p$
 $= 4p$
- $V(X) = V(X_1 + X_2 + X_3 + X_4)$
 $= V(X_1) + V(X_2) + V(X_3) + V(X_4)$
 $= 4(1 - p)$

二項分布と正規分布の関係

試行回数が大きくなるにつれて
正規分布に近づいている



二項分布の極限

n が大きいとき

二項分布の平均

$\frac{X-np}{\sqrt{np(1-p)}}$ は $N(0, 1^2)$ に従うとみなせる

二項分布の標準偏差

正規分布は二項分布の極限状態

→ラプラス定理

母比率の推定検定
適合度検定などに応用