統計の分析と利用

堀田　敬介

確率変数 と 確率分布
期待値, 分散

2008/5/2,Fri.

試行とは？

- 試行
  - 何かの行為により「偶然による」ひとつの結果を導き出す

〔例〕

さいころ投げ

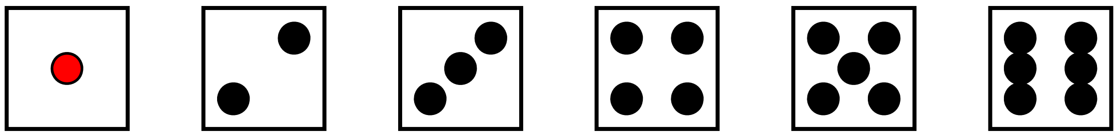
コイン投げ

〔例〕身長の測定，じゃんけん，宝くじを買う，
アンケート調査，製品品質検査，etc.

確率変数とは？

試行してみないと何が出るか
はわからない！
とりうる値はわかっている.

- 確率変数　random variable
  - それがとる各値に対し確率が与えられている変数
    - 例：さいころ投げ



試行結果

$X =$	1	2	3	4	5	6	確率変数の値
$P(X = x_k) =$	1/6	1/6	1/6	1/6	1/6	1/6	確率

確率変数とは？

- 例：コイン投げ

試行結果

$X =$ 表　　裏

確率変数の値

$P(X = x_k) = 1/2 \qquad 1/2$

確率

- 一般に，確率変数の確率は以下のように表現される

$$P(X = x_k) = p_k \quad (k = 1, 2, \cdots)$$

ただし， $p_k \geq 0\,(k = 1, 2, \cdots)$， $\sum_{k=1}^{\infty} p_k = 1$ である．

確率はすべて0以上

全ての確率を足すと1

演習1

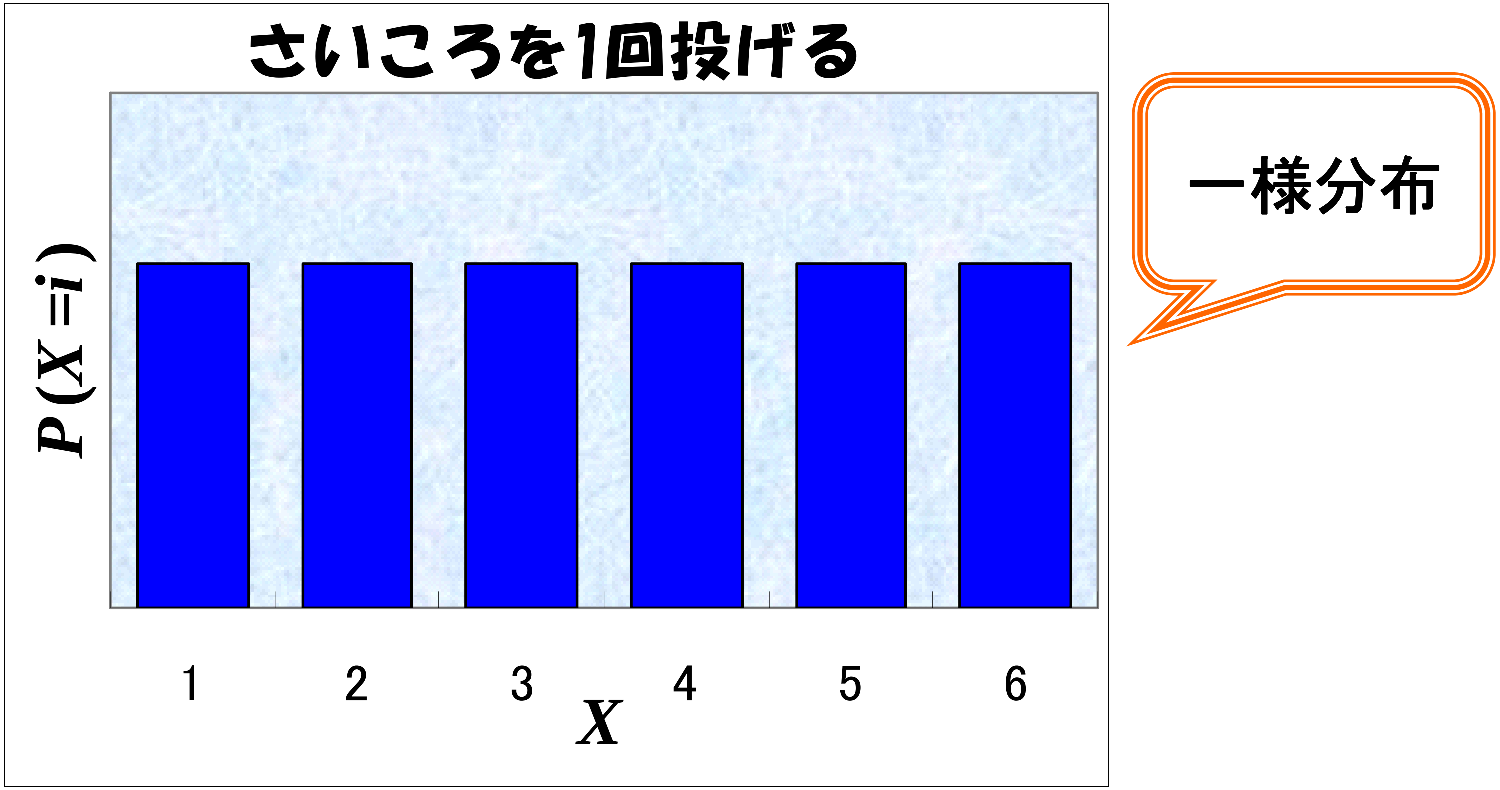
- 確率変数
  - 2個のさいころA, Bを振り出た目の差（Aの目ーBの目）を考える. この確率変数 X のとる値と, その値が出る確率を求めよ.
    - 例）Aが1で, Bが3の時, 1-3 = -2

X	-5	-4	-3	-2	-1	0	1	2	3	4	5
$P(X)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

確率分布 probability distribution

- 確率分布 probability distribution
  - 例:さいころを1回投げる

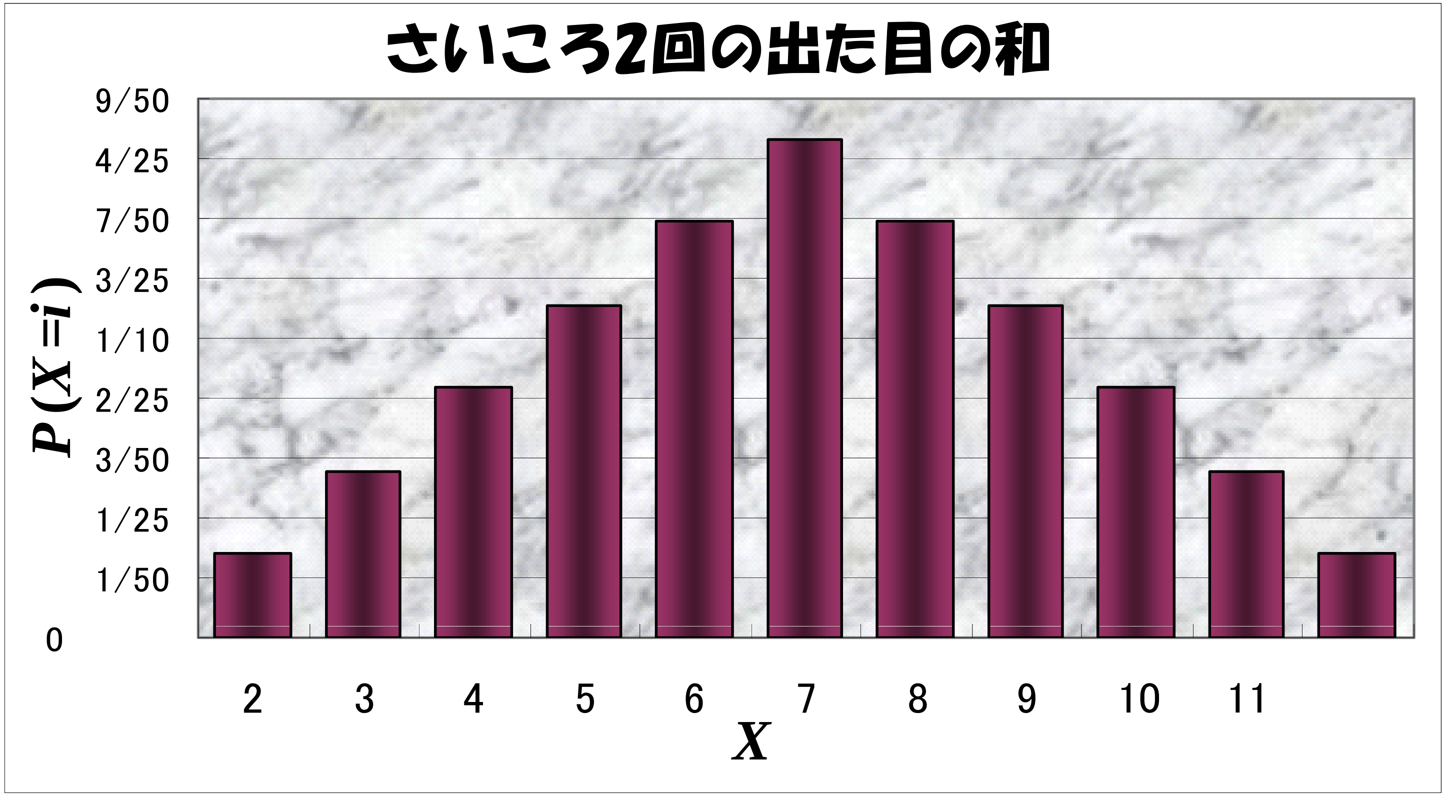
X	1	2	3	4	5	6
$P(X)$	1/6	1/6	1/6	1/6	1/6	1/6



確率分布 probability distribution

- 確率分布
  - 例:さいころを2回投げたときの出た目の和

X	2	3	4	5	6	7	8	9	10	11	12
$P(X)$	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36



三角分布

実は
二項分布

確率分布 probability distribution

- 離散(型)確率分布 discrete distribution
  - 可算集合 $\{x_1, x_2, \ldots\}$の中の値を取る確率変数 X は離散型 discrete type といわれる．このとき，それぞれの値の確率

$$f(x_k) := P(X = x_k) \quad (k = 1, 2, \cdots)$$

を X の確率分布 probability distribution という．

ただし，
$$\begin{cases} f(x_k) \geq 0 \quad (k = 1, 2, \cdots), \\ \displaystyle\sum_{k=1}^{\infty} f(x_k) = 1 \end{cases}$$

確率分布
probability distribution

一般的な定義

確率変数の期待値･分散

- <u>期待値　expectation, expected value</u>
  - 確率変数 X の期待値
    - 例：コインを3回投げて表が出る回数の確率分布

X	0	1	2	3
$P(X)$	1/8	3/8	3/8	1/8

    - 期待値

$$E(X)=(0\times\frac{1}{8})+(1\times\frac{3}{8})+(2\times\frac{3}{8})+(3\times\frac{1}{8})=\frac{3}{2}$$

コインを3回投げると，平均して1.5回表が出ることが期待される

$$\left(\begin{aligned}E(X)&=(\frac{0\times1}{8})+(\frac{1\times3}{8})+(\frac{2\times3}{8})+(\frac{3\times1}{8})\\&=\frac{0}{8}+\frac{1+1+1}{8}+\frac{2+2+2}{8}+\frac{3}{8}\\&=\frac{0+1+1+1+2+2+2+3}{8}\end{aligned}\right)$$

期待値は<u>算術平均を計算</u>しているのと同じ

  - 確率変数Xの期待値

$$E(X)=\sum_{x}x\cdot f(x)$$

演習2

宝くじに関する洒落
LOTTERY: a tax on people who are bad at math

- 期待値を求めよう
  - 宝くじの期待値

H18年オータムジャンボ宝くじ

（新市町村振興 第511回全国自治宝くじ）1億3千万枚限定販売

〔1千万枚あたりの当たり本数〕

1等	1億5000万円	×2本
前後賞	2500万円	×4本
組違賞	10万円	×198本
2等	1000万円	×2本
3等	100万円	×20本
4等	5万円	×3000本
5等	1万円	×20,000本
6等	3000円	×100,000本
7等	300円	×1,000,000本

確率変数の期待値・分散

- 補足：期待値の基本法則
  - スカラー倍の期待値

$$E(aX)=aE(X)$$

「確率変数のスカラー倍」の期待値は，「元の確率変数の期待値のスカラー倍」に等しい

証明：$E(aX)=\sum_x ax\cdot f(x)$
$=a\sum_x xf(x)=aE(X)$

- 例：さいころを振って出た目の1000倍円貰える賭

aX	1000	2000	3000	4000	5000	6000
$P(X)$	1/36	1/18	1/12	1/9	5/36	1/6

$$E(aX)=\frac{1}{6}\times 1000+\frac{1}{6}\times 2000+\cdots+\frac{1}{6}\times 6000=3500$$

X	1	2	3	4	5	6
$P(X)$	1/6	1/6	1/6	1/6	1/6	1/6

$$aE(X)=1000\left(\frac{1}{6}\times 1+\frac{1}{6}\times 2+\cdots+\frac{1}{6}\times 6\right)=3500$$

確率変数の期待値･分散

- <u>分散 variance</u>
  - 確率変数 X の分散

分散（ばらつき）
＝平均（期待値）からのずれ
（の2乗）の平均（期待値）

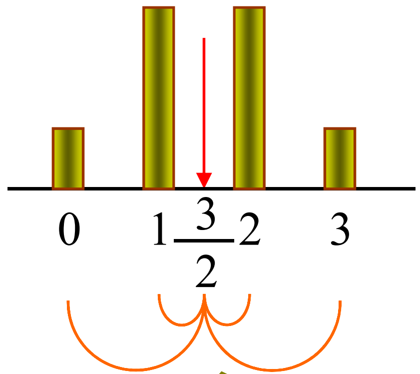
$$V(X) = E(\{X - E(X)\}^2) \qquad \left(V(X) = E(X^2) - E(X)^2\right)$$

- 例：コインを3回投げて表が出る回数の分布の分散は?

$$V(X) = \frac{1}{8}\cdot(0-1.5)^2 + \frac{3}{8}\cdot(1-1.5)^2 + \frac{3}{8}\cdot(2-1.5)^2 + \frac{1}{8}\cdot(3-1.5)^2 = \frac{3}{4}$$

$$\left(\begin{aligned} V(X) &= \frac{1}{8}\cdot\left\{1\times(0-1.5)^2 + 3\times(1-1.5)^2 + 3\times(2-1.5)^2 + 1\times(3-1.5)^2\right\} \\ &= \frac{1}{8}\cdot\left\{(0-1.5)^2 + (1-1.5)^2 + (1-1.5)^2 + (1-1.5)^2 + (2-1.5)^2 + (2-1.5)^2 + (2-1.5)^2 + (3-1.5)^2\right\} \end{aligned}\right)$$

<u>通常の分散を計算</u>しているのと同じ



平均的にどの程度
散らばっているか？

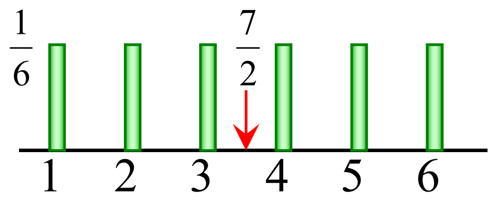
- 確率変数Xの分散

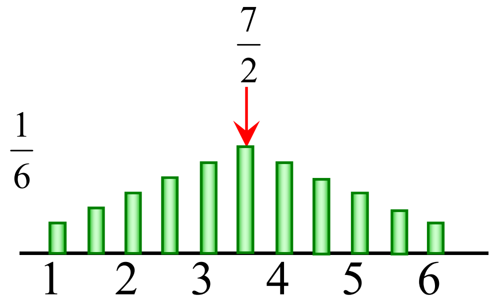
$$V(X) = \sum_x \left(x - E(X)\right)^2 f(x)$$

確率変数の期待値・分散

- 分散は何故必要か？
  - 例：確率変数Xを，さいころを1回振ったときの目，
    確率変数Yを，さいころを2回振ったときの目の平均
    としたとき，それぞれの期待値を求めよ．

$$E(X) = E(Y) = \frac{7}{2}$$





期待値は平均を，分散は散らばりを表す

期待値と分散を知れば分布の目安になる

  - 例題のそれぞれの分散の値を求めよ．

$$V(X) \approx 2.917,$$
$$V(Y) \approx 1.458$$

確率変数の期待値・分散

- 補足：分散の基本法則
  - スカラー倍の分散

$$V(aX)=a^2V(X)$$

「確率変数のスカラー倍」の分散は，「元の確率変数の分散のスカラー2乗倍」に等しい

証明：

$$\begin{aligned}V(aX)&=E((aX)^2)-E(aX)^2\\&=a^2E(X^2)-\{aE(X)\}^2\\&=a^2\{E(X^2)-E(X)^2\}=a^2V(X)\end{aligned}$$

- 例：さいころ1個を振り，出目の1000倍円貰える賭

$a\,X$	1000	2000	3000	4000	5000	6000
$P(X)$	1/36	1/18	1/12	1/9	5/36	1/6

$$\begin{aligned}V(aX)&=\frac{1}{6}\times(1000-3500)^2+\frac{1}{6}\times(2000-3500)^2+\cdots+\frac{1}{6}\times(6000-3500)^2\\&=2{,}916{,}666.66\cdots\end{aligned}$$

X	1	2	3	4	5	6
$P(X)$	1/6	1/6	1/6	1/6	1/6	1/6

$$\begin{aligned}aV(X)&=1000^2\left\{\frac{1}{6}\times(1-3.5)^2+\frac{1}{6}\times(2-3.5)^2+\cdots+\frac{1}{6}\times(6-3.5)^2\right\}\\&=1000^2\times2.9166\cdots=2{,}916{,}666.66\cdots\end{aligned}$$

もし「576倍円貰える」だったらどちらが計算が楽?

一致する

確率変数の期待値・分散

- 標準偏差 standard deviation
  - 確率変数Xの標準偏差

標準偏差
＝分散の平方根

$$D(X)=\sqrt{V(X)}$$

- 例:コインを3回投げて表が出る回数の分布の標準偏差は?

X	0	1	2	3
$P(X)$	1/8	3/8	3/8	1/8

$$D(X)=\sqrt{V(X)}=\sqrt{\frac{3}{4}}=\frac{\sqrt{3}}{2}$$

補足：確率変数の歪度・尖度

- 歪度 skewness
  - 確率変数Xの確率分布の非対称性の指標
  - 歪度＝α_3. ただし,

$$\alpha_3 = E(X-\mu)^3 / \sigma^3$$

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

歪度の数値の意味

$$\begin{cases} \alpha_3 > 0 \ldots \text{右の裾が長い} \\ \alpha_3 < 0 \ldots \text{左の裾が長い} \\ |\alpha_3| \quad \ldots \text{歪みの程度} \end{cases}$$

- 尖度 kurtosis（超過係数 coefficient of excess）
  - 確率変数Xの確率分布の尖り具合を表す指標
  - 尖度＝$\alpha_4 - 3$. ただし,

$$\alpha_4 = E(X-\mu)^4 / \sigma^4$$

正規分布が$\alpha_4 = 3$
なので，これと比較

尖度の数値の意味

$$\begin{cases} \alpha_4 - 3 > 0 \\ \quad \text{正規分布より尖っている} \\ \alpha_4 - 3 < 0 \end{cases}$$

正規分布より丸く鈍い形

演習2

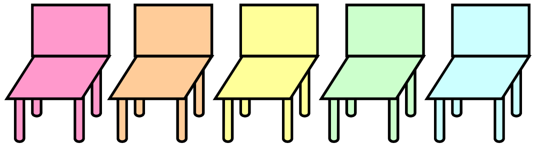
- 確率分布を求めよう
  - コインを5回投げて裏が出る回数の確率分布を求めよ.
  - 求めた確率分布をグラフに描画せよ.
  - 期待値を計算しよう.
  - 分散・標準偏差を計算しよう.
  - 歪度・尖度を計算しよう.

Coffee Break!

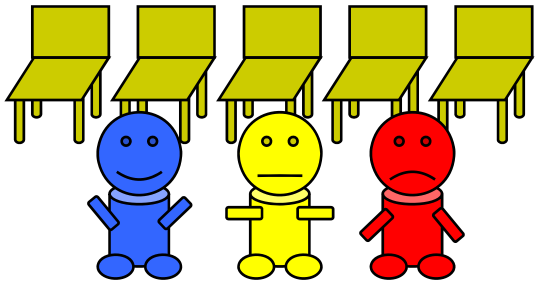
階乗・順列・組合せ

factorial・permutation・combination

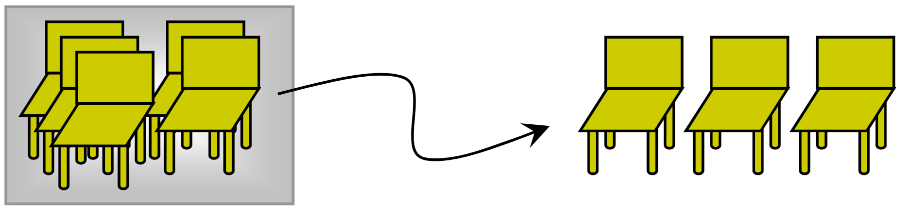
- 異なる5脚の椅子を一列に並べる並べ方は何通り?



- 5脚の椅子に3人の学生が座る座り方は何通り?



- 倉庫に5脚の椅子があります．使うのに必要な分3脚だけ倉庫から取り出します．取り出し方は何通り?



factorial

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

※）Excel関数：FACT(5)

permutation

$$_5P_3 = 5 \cdot 4 \cdot 3 = 60$$

※）Excel関数：PERMUT(5,3)

combination

$$_5C_3 = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$

※）Excel関数：COMBIN(5,3)

確率分布
probability distribution

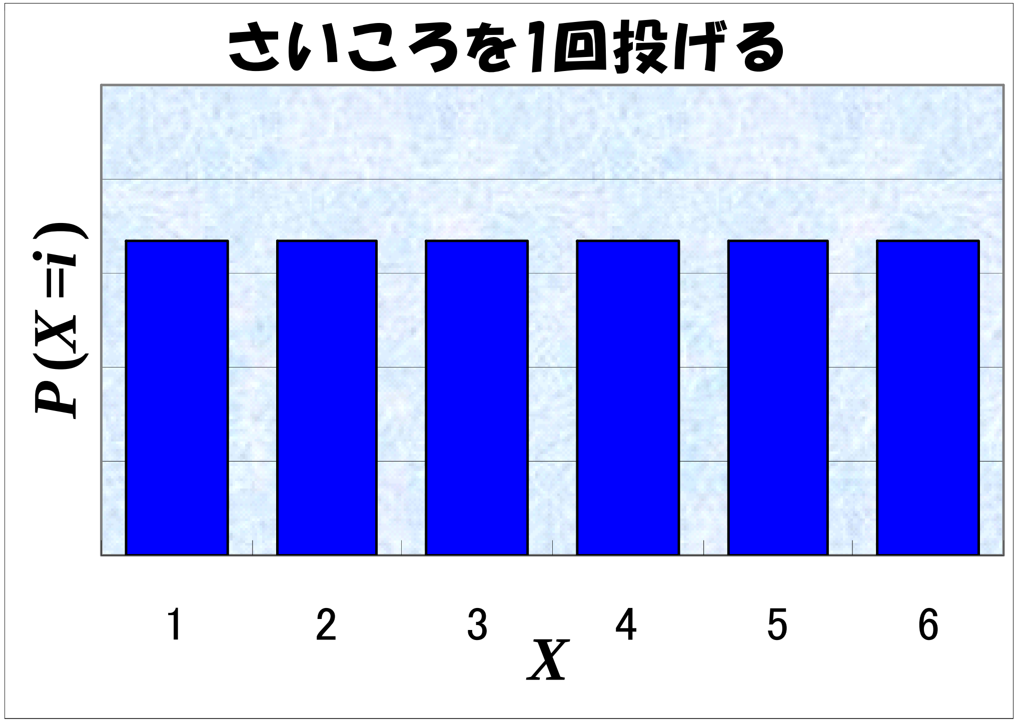
離散(型)分布 discrete distribution

★(離散)一様分布 uniform distribution
★ベルヌーイ分布 Bernoulli distribution
★二項分布 binomial distribution
★ポアソン分布 Poisson distribution
★幾何分布 geometric distribution
★負の二項分布 negative binomial distribution
★超幾何分布 hypergeometric distribution

離散型分布 discrete distribution

- (離散)<u>一様分布 uniform distribution</u> (of discrete type)
  - すべての確率が等しい分布
    - 例：さいころを1回投げる

X	1	2	3	4	5	6
$P(X)$	1/6	1/6	1/6	1/6	1/6	1/6



  - 確率分布

$$f(x)=\frac{1}{n},\ (x=1,\cdots,n)$$

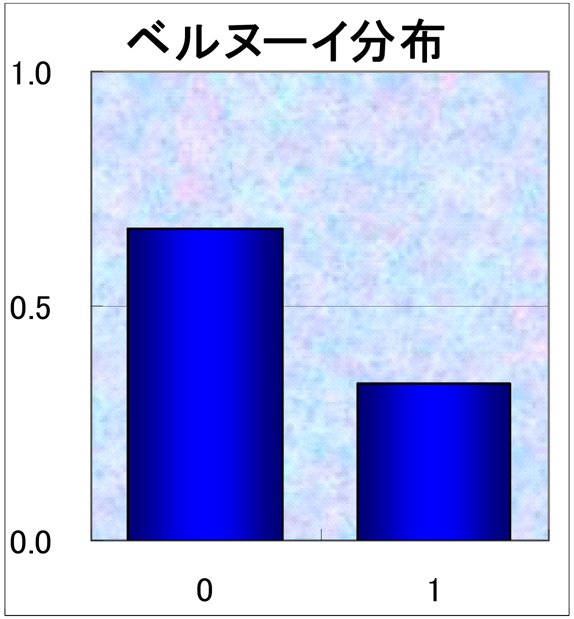
  - 期待値・分散

$$\begin{cases} E(X)=\dfrac{n+1}{2}, \\ V(X)=\dfrac{(n^2-1)}{12} \end{cases}$$

離散型分布 discrete distribution

- ベルヌーイ分布 Bernoulli distribution
  - 試行の結果が2通りしかない確率分布

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



  - 例:コインを1回投げる
    - 表が確率2/3，裏が確率1/3で出る硬貨

X	0	1
$P(X)$	2/3	1/3

- ベルヌーイ試行
  - 2通りの結果しかない観測があり，
    これを同条件で独立にn回行うこと.

離散型分布 discrete distribution

- <u>二項分布 binomial distribution</u>
  - ベルヌーイ試行で，一つの結果が起こる回数の確率
  - 確率pをもつ事象がn回の施行中x回起こる確率
    - 例：サイコロを3回投げて1の目がx回出る確率は…

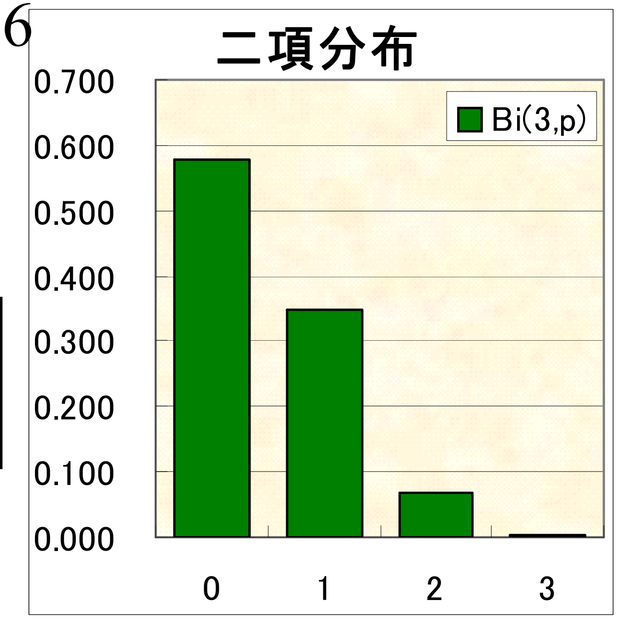
0回出る確率 … ${}_3C_0\,p^0(1-p)^3 = 125/216$

1回出る確率 … ${}_3C_1\,p^1(1-p)^2 = 75/216$

2回出る確率 … ${}_3C_2\,p^2(1-p)^1 = 15/216$

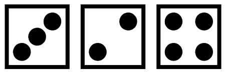
3回出る確率 … ${}_3C_3\,p^3(1-p)^0 = 1/216$

$(p = 1/6)$



1が0回出る

$\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3$

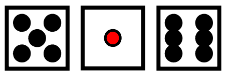


3箇所に0個1を置く

${}_3C_0$

1が1回出る

$\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2$



3箇所に1個1を置く

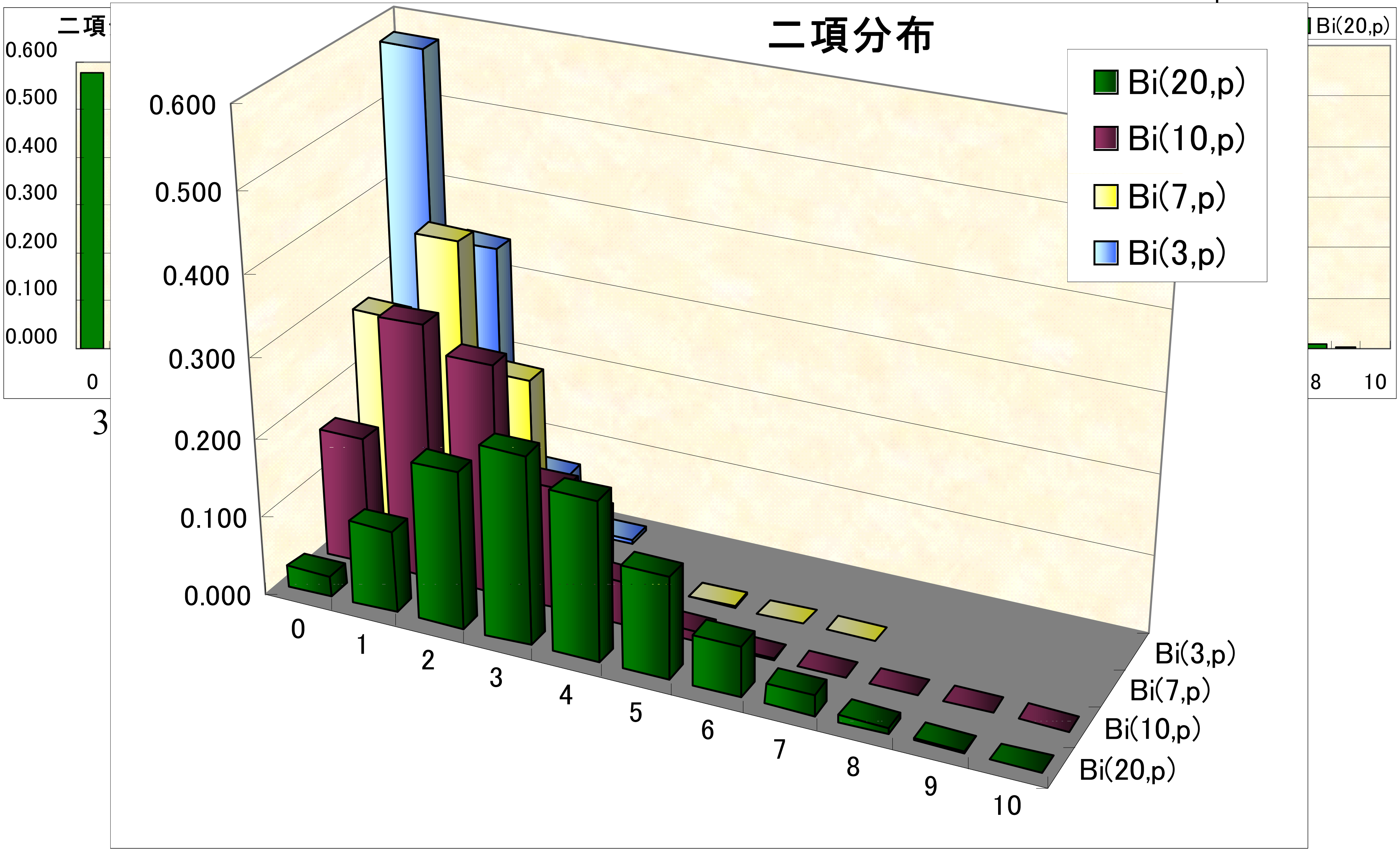
${}_3C_1$

確率分布

X	0	1	2	3
$P(X)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

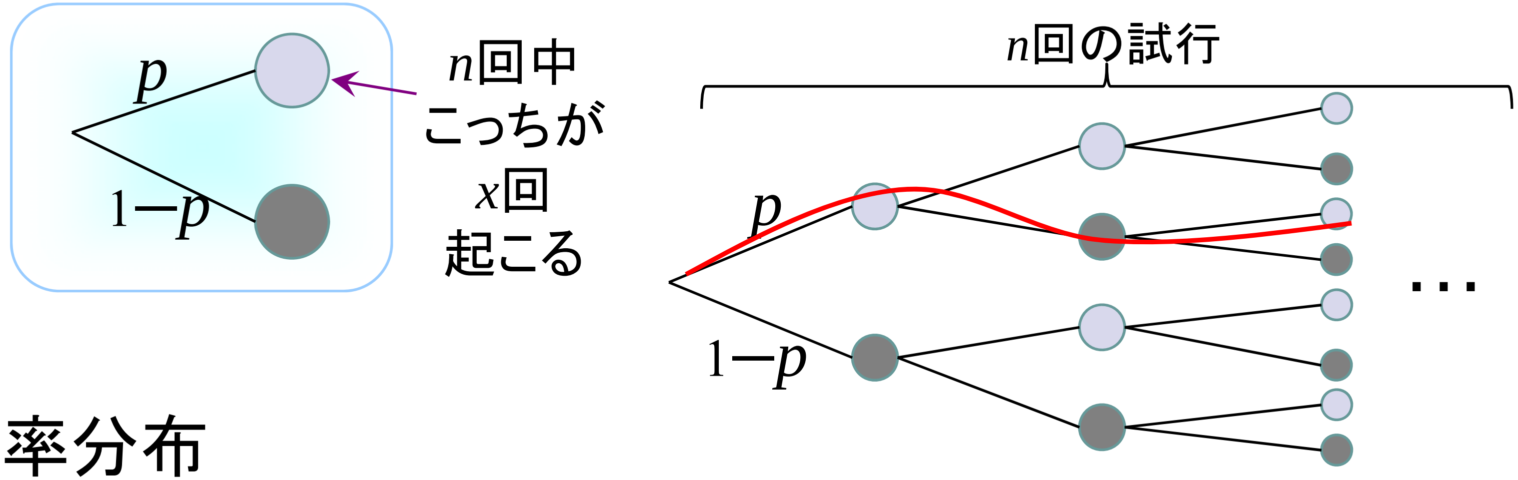
離散型分布 discrete distribution

- サイコロを投げる回数を増やすと1の目が x 回出る確率は…



離散型分布 discrete distribution

- 二項分布 $\boxed{Bi(n, p)}$
  - 確率pをもつ事象が<u>n回</u>の施行中<u>x回</u>起こる確率



  - 確率分布

$$f(x) = {}_nC_x\, p^x (1-p)^{n-x} \quad (x = 0, \cdots, n)$$

  - 期待値・分散

$$\begin{cases} E(X) = np, \\ V(X) = np(1-p) \end{cases}$$

Coffee Break!

パスカルの三角形と二項係数

$${}_3C_0 = 1$$

$${}_3C_1 = \frac{3}{1} = 3$$

$${}_3C_2 = \frac{3 \cdot 2}{2 \cdot 1} = 3$$

$${}_3C_3 = \frac{3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1} = 1$$

$${}_4C_0 = 1$$

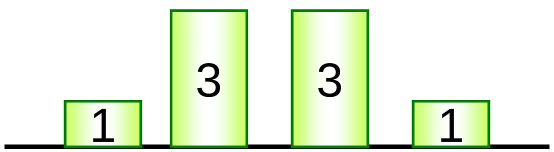
$${}_4C_1 = \frac{4}{1} = 4$$

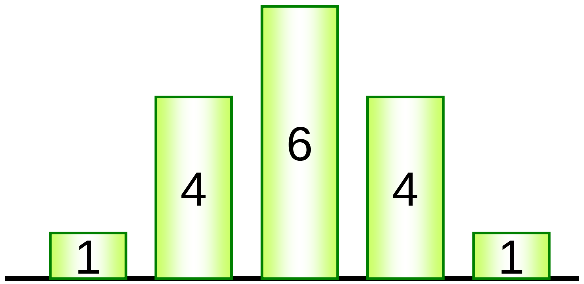
$${}_4C_2 = \frac{4 \cdot 3}{2 \cdot 1} = 6$$

$${}_4C_3 = \frac{4 \cdot 3 \cdot 2}{3 \cdot 2 \cdot 1} = 4$$

$${}_4C_4 = \frac{4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1} = 1$$

…





…

Coffee Break!

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & & & & \vdots & & & & &
\end{array}
\qquad
\begin{array}{c}
\longleftrightarrow \\
\longleftrightarrow \\
\longleftrightarrow \\
\longleftrightarrow \\
\longleftrightarrow
\end{array}
\qquad
\begin{array}{cccccc}
{}_1C_0 & {}_1C_1 & & & & \\
{}_2C_0 & {}_2C_1 & {}_2C_2 & & & \\
{}_3C_0 & {}_3C_1 & {}_3C_2 & {}_3C_3 & & \\
{}_4C_0 & {}_4C_1 & {}_4C_2 & {}_4C_3 & {}_4C_4 & \\
{}_5C_0 & {}_5C_1 & {}_5C_2 & {}_5C_3 & {}_5C_4 & {}_5C_5 \\
 & & \vdots & & &
\end{array}
$$

☆ 二項定理，二項係数

$(a+b)^n$ の各項の係数

$$
\begin{cases}
(a+b)^2 = a^2 + 2ab + b^2 \\
(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3 \\
(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4 \\
(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5 \\
\qquad\qquad \vdots
\end{cases}
$$

☆ 組合せ数の和法則

$$
{}_nC_k = {}_{n-1}C_{k-1} + {}_{n-1}C_k \quad (\text{for } 1 \le k \le n-1)
$$

☆ $$
{}_nC_k = \sum_{n-k-1 \le m \le n-1} {}_mC_{m+k+1-n} = \sum_{k-1 \le m \le n-1} {}_mC_{k-1}
$$

$$10 = 1 + 3 + 6$$

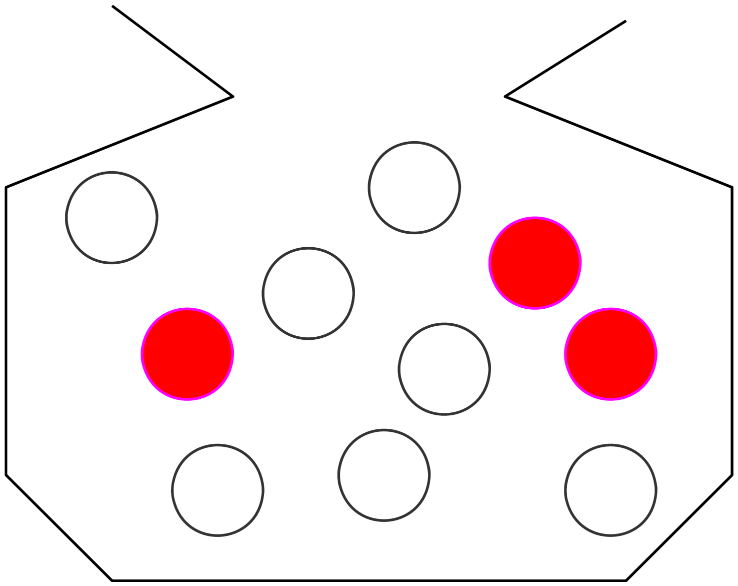
$$10 = 1 + 2 + 3 + 4$$

離散型分布 discrete distribution

- 二項分布の例
  - 例1：製品ラインの不良品抜き取り検査
    - 不良率p=0.5%のロットから独立に1個ずつランダム抜取り検査をした時に検出される不良品数x の従う分布
    - 参考）不良率の期待値 $E(x/n) = np/n = p$
  - 例2：袋から球を取り出す
    - 赤玉3，白玉7入っている袋から1つ取り出しては戻すという行為をn回行ったとき，赤玉が5回出る確率は?
  - 例3：サイコロをn回投げて偶数の目が出る回数の従う分布
    - サイコロを5回投げて偶数が出る回数の確率分布を求めよ

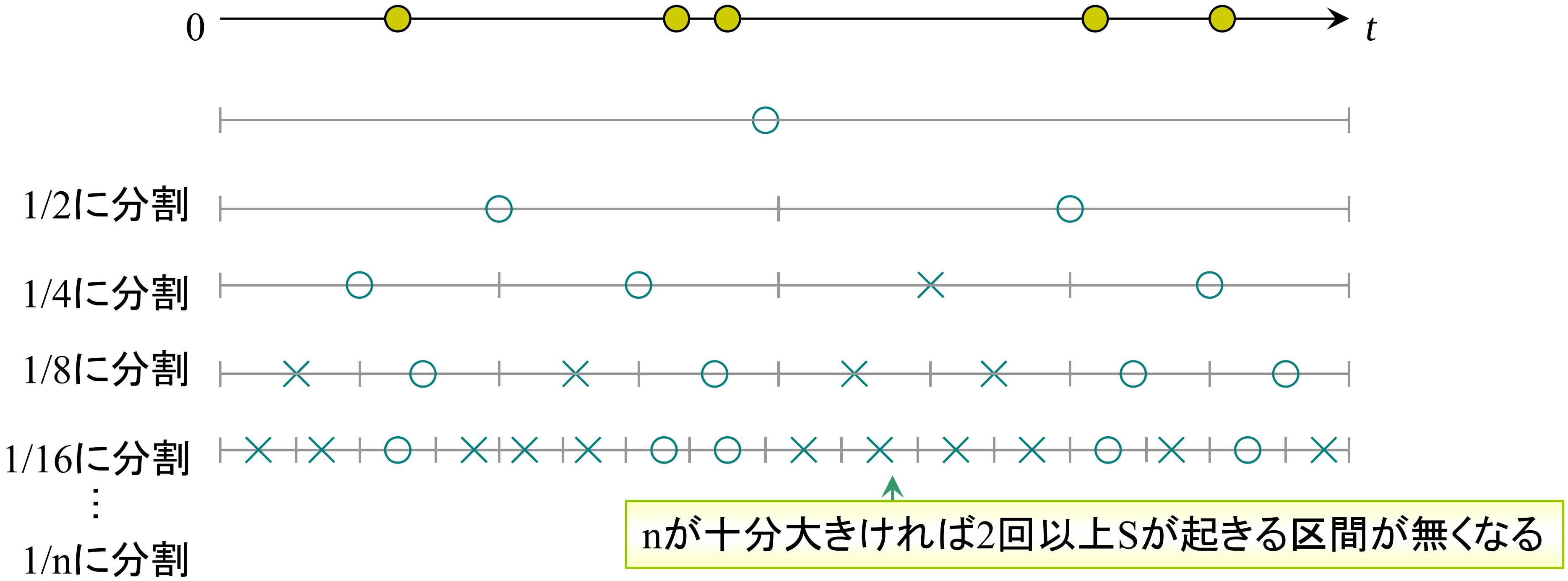
演習3

- 二項分布を求めよう
  - 赤玉3個，白玉7個入っている袋から1つ取り出しては戻すという行為を5回行ったとき，赤球が出る回数の確率分布（二項分布）を求めてみよう！



離散型分布 discrete distribution

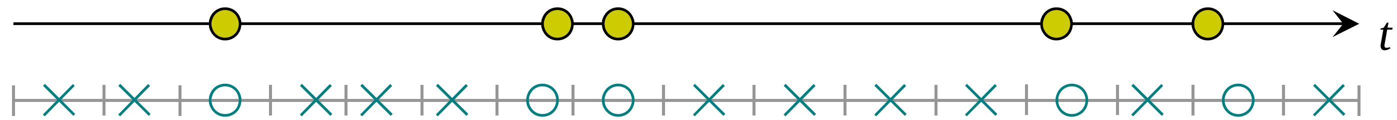
- <u>ポアソン分布 Poisson distribution</u>
  - ある時間帯の中で，ある事象が何回起きるか？
    - 例：電話番号案内
      - 事象S＝「通話がある」



P(ある区間でSが2回以上起きる確率)=0とする

離散型分布 discrete distribution

- ポアソン分布 Poisson distribution



$$\left\{\begin{array}{ll}\text{各区間でSが起きる確率} & : p \\ \text{各区間でSが起きない確率} & : q=1-p\end{array}\right\}\text{とする}$$

ベルヌーイ試行とみなす

⇒ Sが起きる回数は二項分布 $Bi\,(n, p)$ に従う

確率pの事象がn回の試行の中でS回起こる

ところで，この時間内にSが起きる回数の期待値をλとおくと…

$$\lambda = np \qquad \text{（二項分布の期待値より）}$$

よって，Sがk回起きる確率は，

$$
\begin{aligned}
&{}_nC_k\, p^k (1-p)^{n-k} \\
&= {}_nC_k \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!}\left((1-\frac{1}{n})\cdots\cdots(1-\frac{k-1}{n})\right)\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n} \underset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}
\end{aligned}
$$

$$\left(\because \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}\right)$$

離散型分布 discrete distribution

- ポアソン分布 Poisson distribution
  - 2項分布においてある事象が起こる確率が非常に小さい場合に適用できる分布
    - 例：工場の生産ラインでの不良率が1/500のとき，1000個の製品を作ったときx個不良品だった

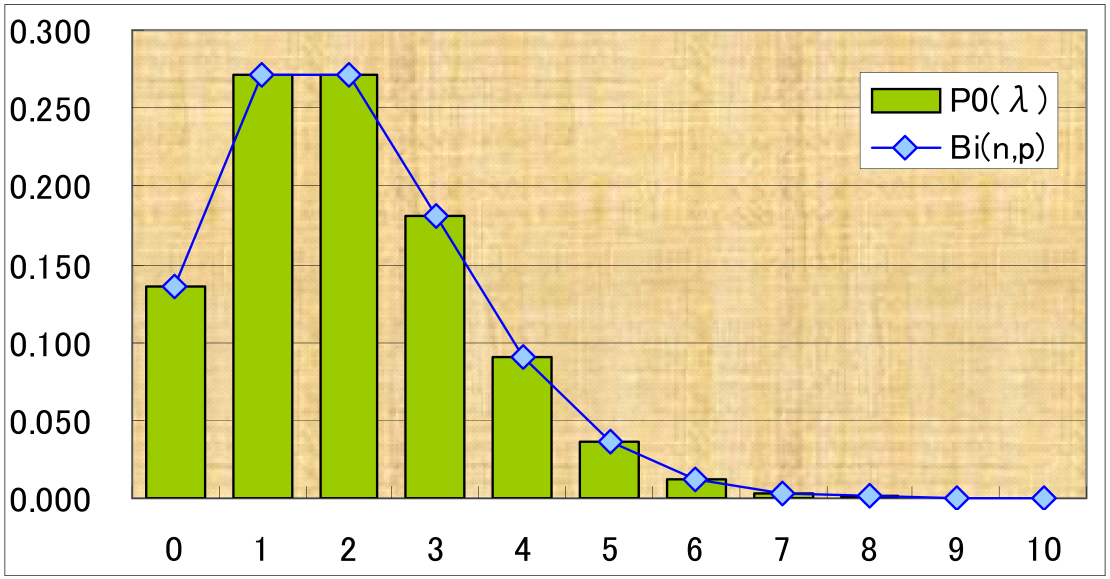
$$f(x) = {}_{1000}C_x \left(\frac{1}{500}\right)^x \left(1-\frac{1}{500}\right)^{1000-x}$$

二項分布

ポアソン分布

$$f(x) = e^{-2}\frac{2^x}{x!}$$

1000個のうち，平均的に2個不良品



離散型分布 discrete distribution

- ポアソン分布 $\boxed{Po(\lambda)}$
  - 確率分布

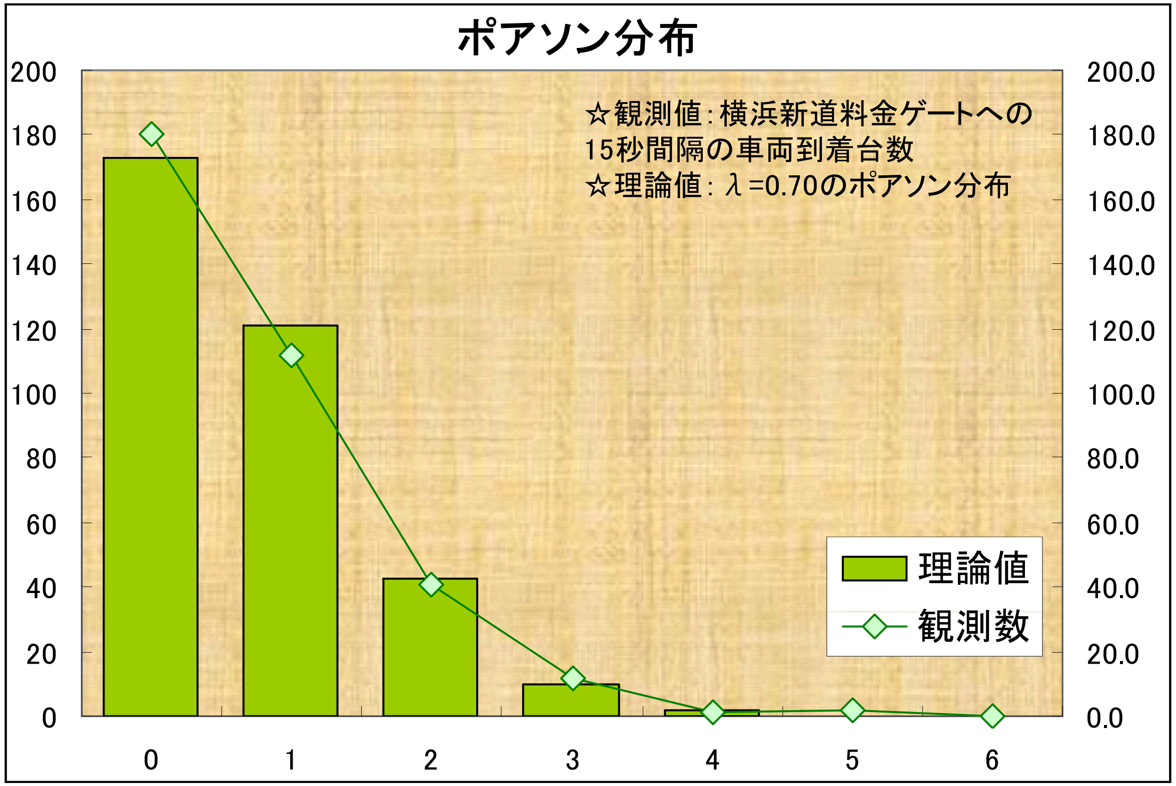
$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!},\ (x = 0,1,2,\cdots)$$

  - 期待値・分散

$$\begin{cases} E(X) = \lambda, \\ V(X) = \lambda \end{cases}$$

例では…

$$\begin{cases} E(X) = 0.70, \\ V(X) = 0.70 \end{cases}$$



離散型分布 discrete distribution

- 二項分布からポアソン分布へ

ポアソンの小数の法則
Poisson's law of small numbers

$$ {}_nC_x p^x (1-p)^{n-x} \to e^{-\lambda}\frac{\lambda^x}{x!} \quad \left(\begin{array}{l} n\to\infty, \\ p\to 0 \end{array}\right) $$

$np = \lambda$

二項分布　　ポアソン分布

- 例 : ある工場の生産ラインでの不良率が1/500である. 1000個の製品を作ったときx個不良品だった.

二項分布	ポアソン分布
$f(0)=0.135065,$	$f(0)=0.135335,$
$f(1)=0.270670,$	$f(1)=0.270671,$
$f(2)=0.270942$	$f(2)=0.270671,$
$f(3)=0.180628,$	$f(3)=0.180447,$
$f(4)=0.090223,$	$f(4)=0.090224,$
$\vdots$	$\vdots$