

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

	500	試行結果	
$X =$	表	裏	確率変数の値
$P(X=x_k) =$	$1/2$	$1/2$	確率

一般に、確率変数の確率は以下のように表現される
 $P(X=x_k) = p_k \quad (k=1,2,\dots)$
ただし、 $p_k \geq 0 \quad (k=1,2,\dots)$, $\sum_{k=1}^{\infty} p_k = 1$ 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-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、 $1-3 = -2$

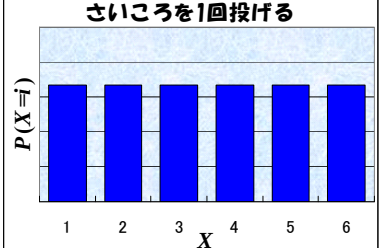
X	-5	-4	-3	-2	-1	0	1	2	3	4	5
$P(X)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

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

さいころを1回投げる



一様分布

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

確率分布 probability distribution

- 離散(型)確率分布 discrete distribution
 - 可算集合 $\{x_1, x_2, \dots\}$ 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
$$f(x_k) := P(X = x_k) \quad (k = 1, 2, \dots)$$
を X の **確率分布 probability distribution** という。ただし、
$$\begin{cases} f(x_k) \geq 0 & (k = 1, 2, \dots), \\ \sum_{k=1}^{\infty} f(x_k) = 1 \end{cases}$$

一般的な定義

確率分布 probability distribution

確率変数の期待値・分散

- 期待値 expectation, expected value
 - 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

コインを3回投げると、平均して1.5回表が出るのが期待される

期待値

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

確率変数 X の期待値

$$E(X) = \sum_x x \cdot f(x)$$

期待値は算術平均を計算しているのと同じ

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年オース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

「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: $E(aX) = \sum a x \cdot f(x) = a \sum x f(x) = aE(X)$

例:さいころを振って出目の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

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

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

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

一致する

確率変数の期待値・分散

- 分散 variance
 - 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)

$$V(X) = E(\{X - E(X)\}^2) \quad (V(X) = E(X^2) - E(X)^2)$$

例:コインを3回投げて表が出る回数の分布の分散は?

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

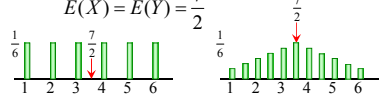
通常の分散を計算しているのと同じ

確率変数 X の分散

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

平均的にどの程度散らばっているか?

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
- 例題のそれぞれの分散の値を求めよ。
$$V(X) \approx 2.917, V(Y) \approx 1.458$$

期待値と分散を
知れば分布の
目安になる

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

証明：
$$V(aX) = E((aX)^2) - E(aX)^2 = a^2 E(X^2) - \{aE(X)\}^2 = a^2 \{E(X^2) - E(X)^2\} = a^2 V(X)$$
 - 例：さいころ1個を振り、出目の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

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

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

$$aV(X) = 1000^2 \left\{ \frac{1}{6} \times (1 - 3.5)^2 + \frac{1}{6} \times (2 - 3.5)^2 + \dots + \frac{1}{6} \times (6 - 3.5)^2 \right\} = 1000^2 \times 2.9166\dots = 2,916,666.66\dots$$

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,
$$\alpha_3 = E(X - \mu)^3 / \sigma^3$$
$$(\mu := E(X), \sigma^2 := V(X))$$

歪度の数値の意味

$\alpha_3 > 0$	... 右の裾が長い
$\alpha_3 < 0$	... 左の裾が長い
$ \alpha_3 $	... 歪みの程度
- 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,
$$\alpha_4 = E(X - \mu)^4 / \sigma^4$$

尖度の数値の意味

$\alpha_4 - 3 > 0$	正規分布より尖っている
$\alpha_4 - 3 < 0$	正規分布より丸く鈍い形

正規分布が $\alpha_4 = 3$ なので、これと比較

演習2

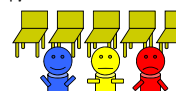
- 確率分布を求めよう
 - コインを5回投げて裏が出る回数の確率分布を求めよ。
 - 求めた確率分布をグラフに描画せよ。
 - 期待値を計算しよう。
 - 分散・標準偏差を計算しよう。
 - 歪度・尖度を計算しよう。

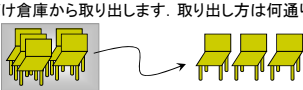
Coffee Break!

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

※) Excel関数: FACT(5)
- 5脚の椅子に3人の学生が座る座り方は何通り？
$${}_5P_3 = 5 \cdot 4 \cdot 3 = 60$$

※) Excel関数: PERMUT(5,3)
- 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？
$${}_5C_3 = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} = 10$$

※) Excel関数: COMBIN(5,3)

それぞれどんな計算になるのかしら？

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)一様分布 uniform distribution (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

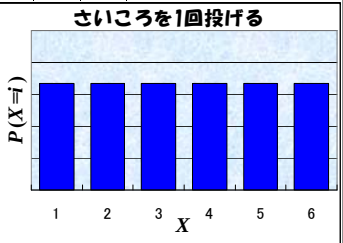
X	1	2	3	4	5	6
$P(X)$	1/6	1/6	1/6	1/6	1/6	1/6

● 確率分布

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

● 期待値・分散

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



離散型分布 discrete distribution

● ベルヌーイ分布 Bernoulli distribution

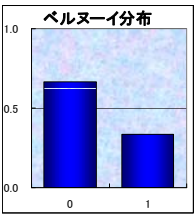
- 試行の結果が2通りしかない確率分布

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

● 二項分布 binomial distribution

- 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

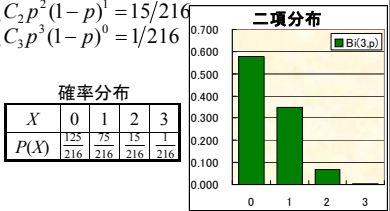
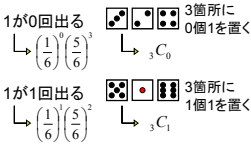
- 例:サイコロを3回投げて1の目がx回出る確率は...

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

$$1\text{回出る確率} \dots {}_3C_1 p^1 (1-p)^2 = 75/216 \quad (p=1/6)$$

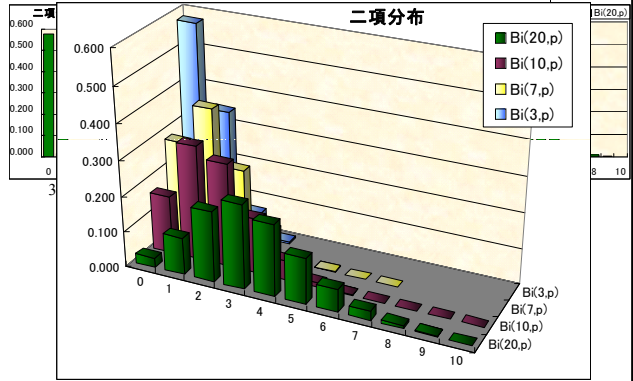
$$2\text{回出る確率} \dots {}_3C_2 p^2 (1-p)^1 = 15/216$$

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



離散型分布 discrete distribution

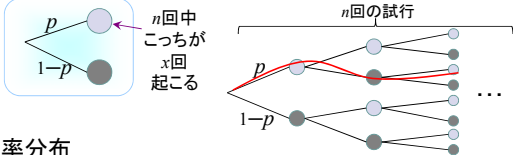
■サイコロを投げる回数を増やすと1の目がx回出る確率は...



離散型分布 discrete distribution

● 二項分布 $Bi(n, p)$

- 確率pをもつ事象がn回の施行中x回起こる確率



● 確率分布

$$f(x) = {}_n C_x p^x (1-p)^{n-x} \quad (x=0, \dots, n)$$

● 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!

$$\begin{matrix} & & 1 & & \\ & 1 & 2 & 1 & \\ 1 & 3 & 3 & 1 & \\ 1 & 4 & 6 & 4 & 1 \end{matrix}$$

$$\begin{matrix} & & {}_1C_0 & & {}_1C_1 \\ & {}_2C_0 & {}_2C_1 & {}_2C_2 \\ {}_3C_0 & {}_3C_1 & {}_3C_2 & {}_3C_3 \\ {}_4C_0 & {}_4C_1 & {}_4C_2 & {}_4C_3 & {}_4C_4 \end{matrix}$$

$$\begin{matrix} & & 1 & & \\ & 1 & 5 & 10 & 10 & 5 & 1 \end{matrix}$$

$$\begin{matrix} & & {}_5C_0 & {}_5C_1 & {}_5C_2 & {}_5C_3 & {}_5C_4 & {}_5C_5 \end{matrix}$$

★ 二項定理, 二項係数
 $(a+b)^n$ の各項の係数

★ 組合せ数の和法則
 ${}_nC_k = {}_{n-1}C_{k-1} + {}_{n-1}C_k$
(for $1 \leq k \leq n-1$)

★ ${}_nC_k = \sum_{m=k-1}^{n-k} {}_mC_{m+k-1-n}$

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

$$\begin{matrix} 10 = 1 + 3 + 6 \\ 10 = 1 + 2 + 3 + 4 \end{matrix}$$

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個ずつランダム抜き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

● ポアソン分布 Poisson distribution

● 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」

$$\begin{matrix} 0 \\ 1/2 \text{に分割} \\ 1/4 \text{に分割} \\ 1/8 \text{に分割} \\ 1/16 \text{に分割} \\ \vdots \\ 1/n \text{に分割} \end{matrix}$$

n が十分大きければ2回以上 S が起きる区間が無くなる

$\Rightarrow P(\text{ある区間で} S \text{が2回以上起きる確率}) = 0$ とする

離散型分布 discrete distribution

● ポアソン分布 Poisson distribution

$$\begin{matrix} 0 \\ 1/n \text{に分割} \end{matrix}$$

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

$\Rightarrow S$ が起きる回数は二項分布 $Bi(n, p)$ に従う

ベルヌーイ試行とみなす

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 とすると...

$\lambda = np$ (二項分布の期待値より)

よって, S が k 回起きる確率は,

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

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

$$\lim_{n \rightarrow \infty} \frac{\lambda^k}{k!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Confidential

5

離散型分布 discrete distribution

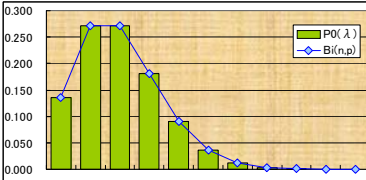
- **ポアソン分布 Poisson distribution**
 - 2項分布においてある事象が起こる確率が**非常に小さい**場合に適用できる分布
 - 例: 工場の生産ラインでの不良率が1/500のとき, 1000個の製品を作ったときx個不良品だった
$$f(x) = {}_{1000}C_x \left(\frac{1}{500}\right)^x \left(1 - \frac{1}{500}\right)^{1000-x}$$

二項分布

ポアソン分布

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

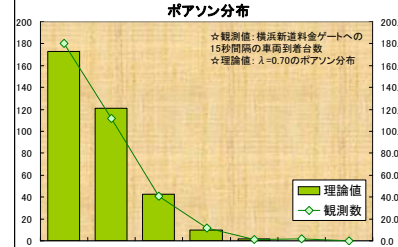
1000個のうち, 平均的に2個不良品



離散型分布 discrete distribution

- **ポアソン分布 $Po(\lambda)$**
 - 確率分布
$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad (x = 0, 1, 2, \dots)$$
 - 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

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

二項分布

ポアソン分布

$np = \lambda$

 - 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

$f(0) = 0.135065,$	$f(0) = 0.135335,$
$f(1) = 0.270670,$	$f(1) = 0.270671,$
$f(2) = 0.270942,$	$f(2) = 0.270671,$
$f(3) = 0.180628,$	$f(3) = 0.180447,$
$f(4) = 0.090223,$	$f(4) = 0.090224,$
...	...

ポアソン分布

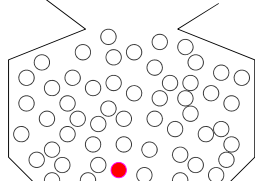
離散型分布 discrete distribution

- **二項分布とポアソン分布**
 - 例: 一時間に 生ずる事故件数は?
 - 一 m件の事故が 生じたとする。これを1時間, 1分, 1秒 と めていき, 1 みに1件の事故が 生ずるようにし(同時に2件 生ずることはないとする), その みをnとする。
 - すると, この話は n個の みの中で1件事故が 生ずるかしないかとみなすことができる。 ち, n個の みの中から事件が 生じた x個の みの個数を考えることになる
 - 事故 生率は $p = m / n = \text{一定!}$
(ポアソン分布の期待値)
 - このときの $m = np$ は二項分布の期待値!

離散型分布 discrete distribution

- **ポアソン分布の例**
 - 例1: 行 事故(事故はめったに起きない)
 - 行 事故の確率1/10万, 行 乗回数1万回としたとき, 一度も事故にあわない確率は?
 - 例2: 大 生産品の不良品数(めったにない)
 - 不良率が1/10000の生産ラインで1万個生産したとき不良品が3個以上出る確率は?
 - 例3: 中数(めったに当たらない)
 - 第二 大 中の イ の 中 度は $\lambda=0.93$ のポアソン分布に従うという。1000 けて1 当たる確率は?
 - 例4: の 作用
 - 作用の確率が1/200の を5000人が 用したとき, 30人以上に 作用が出る確率は?
 - 例5: 生 ・ の生 ・ を ず分布
 - あたりの アの個数

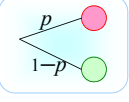
演習

- **ポアソン分布を求めよう**
 - 赤玉1個, 白玉99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(ポアソン分布)を求めてみよう!

離散型分布 discrete distribution

- 幾何分布 geometric distribution
 - ベルヌーイ試行において、試行回数を x とするときの $X=x$ の確率分布
$$f(x) = p(1-p)^{x-1} \quad (x=1,2,\dots)$$

幾何数列(等比数列)の形なので、幾何分布とよばれる

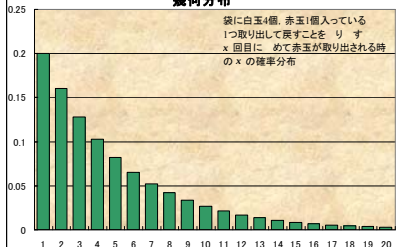


$x-1$ 回 x 回目
 - 幾何分布は、時間を離散的に(1,2,3,...)考えるとき、初めて何かが起こるまで待つ時間の長さの確率分布〔(離散的な)待ち時間分布〕

離散型分布 discrete distribution

- 幾何分布
 - 確率分布
$$f(x) = p(1-p)^{x-1}, \quad (x=1,2,\dots)$$
 - 期待値・分散
$$\begin{cases} E(X) = \frac{1}{p}, \\ V(X) = \frac{1-p}{p^2} \end{cases}$$

例では...
$$\begin{cases} E(X) = \frac{1}{1/5} = 5.00, \\ V(X) = \frac{(1-1/5)}{(1/5)^2} = 20.00 \end{cases}$$



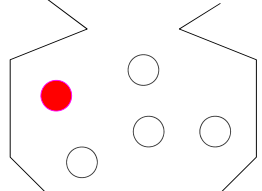
幾何分布
袋に白玉4個、赤玉1個入っている
1つ取り出して戻すことを x 回目まで繰り返して赤玉が取り出される確率分布

離散型分布 discrete distribution

- 幾何分布の例
 - 例1: どの年に起こる確率が1/25であるとする。平均何年に1回か?
 - 例2: 袋から...
 - 白玉4つ、赤玉1つが入っている袋がある。1つ取り出して元に戻すという試行を繰り返したとき、10回目に初めて赤玉が取り出される確率は?
 - 例3: アを n 回試行するとき、何回目でアが初めて起こるか? ただし、試した回数は1回を超えないとする

演習

- 幾何分布を求めよう
 - 白玉4つ、赤玉1つが入っている袋がある。1つ取り出して元に戻すという試行を繰り返したとき、初めて赤玉が取り出される回の確率分布(幾何分布)を求めてみよう! 10回目に初めて赤玉が取り出される確率はどれくらいか?

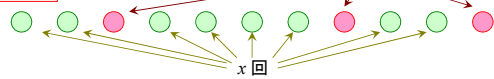


離散型分布 discrete distribution

- 負の二項分布 negative binomial distribution
 - ある事象が k 回起こるまでのもうひとつの事象の回数を x としたときの $X=x$ ($x=0,1,2,\dots$) の従う分布
$$f(x) = C_{k+x-1}^{k-1} p^k (1-p)^x \quad (x=0,1,2,\dots)$$

二項分布で二項係数に負もめた場合にこの分布になるので「負」の二項分布とよばれる

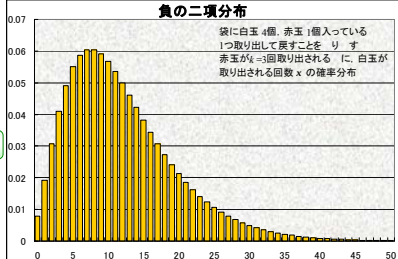
k 回 x 回
 - $k=1$ のときは幾何分布に等しいため幾何分布の一般となっている



離散型分布 discrete distribution

- 負の二項分布
 - 確率分布
$$f(x) = C_{k+x-1}^{k-1} p^k (1-p)^x, \quad (x=0,1,2,\dots)$$
 - 期待値・分散
$$\begin{cases} E(X) = \frac{k(1-p)}{p}, \\ V(X) = \frac{k(1-p)}{p^2} \end{cases}$$

幾何分布の k 倍



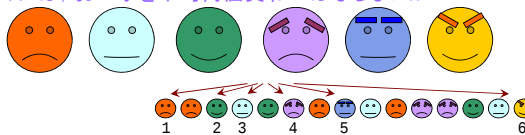
負の二項分布
袋に白玉4個、赤玉1個入っている
1つ取り出して戻すことを繰り返して赤玉が k 回取り出されるまでに、白玉が取り出される回数 x の確率分布

離散型分布 discrete distribution

- 負の二項分布の例
 - 例1: の到
 - ある1年に が起こる確率が1/25であるとする。が起こるのは平均何年に1回か?
 - 上と同じ が20年以内に起こる確率は?
 - 例2: 袋から...
 - 白玉4つ、赤玉1つが入っている袋がある。1つ取り出して元に戻すという試行を したとき、赤玉が3回取り出されるまでに白玉が40回取り出される確率は?
 - 例3: — ものコ ター
 - 12 の ヤラ ターが売られている。ただし、 を けるまで中にどれが入っているかはわからない。あるコ ターが全ての ヤラ ターを集めるためには何個買わ ばならないか?

演習

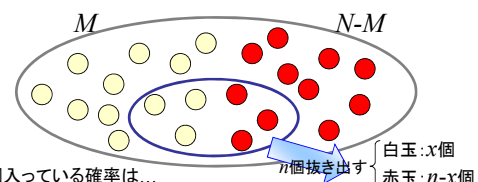
- 負の二項分布を求めよう
 - お 子 の に6 の 負けがある。この 負けはそれ ぞれの に全 がランダムに等確率に入っているとす る。 を けるまで中身はわからない。全ての 負けを集める ためには、お 子 を平均何個買わ ばならないか?



一ボン 集 題と ばれる 題であり、正しくは、負の二項分布ではない。

離散型分布 discrete distribution

- 超幾何分布 hypergeometric distribution
 - 例: 白玉がM個、赤玉がN-M個(全 でN個)ある。ここ からn個抜き出したとき、白玉がx個入っている確率は?



白玉がx個入っている確率は...

$$f(x) = \frac{M \cdot C_x \cdot N-M \cdot C_{n-x}}{N \cdot C_n}$$

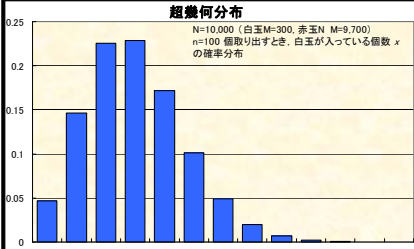
ただし、xの取り える は

$(x = \max\{0, n - (N - M)\}, \dots, \min\{n, M\})$

n個取り出す組合せのうち、白玉x個、赤 玉n-x個取り出す組合せの確率

離散型分布 discrete distribution

- 超幾何分布
 - 確率分布



- 期待値・分散

$$\begin{cases} E(X) = \frac{Mn}{N}, \\ V(X) = \frac{Mn}{N} \cdot \frac{N-M}{N} \cdot \frac{N-M}{N-1} \end{cases}$$

例では...

$$\begin{aligned} E(X) &= \frac{300 \cdot 100}{10000} = 3.00, \\ V(X) &= \frac{300 \cdot 100}{10000} \cdot \frac{9700}{10000} \cdot \frac{9700}{9999} = 2.82 \end{aligned}$$

離散型分布 discrete distribution

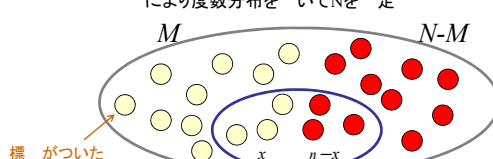
- 超幾何分布の性質
 - 非 元 出(とったものを戻さない)の時に現れる分布
 - 元 出の場合、 $M/N = p$ とした二項分布となる
 - $N \rightarrow \infty$ の場合、条件 $M/N \rightarrow p$ の元で二項分布となる

$$\begin{cases} E(X) = \frac{Mn}{N}, & \longrightarrow np \\ V(X) = \frac{Mn}{N} \cdot \frac{N-M}{N} \cdot \frac{N-M}{N-1} & \longrightarrow np(1-p) \end{cases}$$

二項分布の平均と分散

離散型分布 discrete distribution

- 超幾何分布の例
 - 例: 調査「法 capture-recapture method」
ともいう
 - の中の の個 数 定など
に何 の (N)がいるのか知りたいが、くので しい
により度数分布を いてNを 定



- 例: ある の中に生 している対象 について、200 を し標 をつけた。さてしばらく後、 から を10 ったとき、標 がついて いる が2 いた。この にはこの は何 いると 定されるか?

Confidential

8

Coffee Break!

Q1) 1から100までの和は?
 $1 + 2 + 3 + \dots + 100 = ?$

Q2) 1から100までの2乗和は?
 $1^2 + 2^2 + 3^2 + \dots + 100^2 = ?$

1からnまでの和と2乗和は以下

$$\sum_{k=1}^n k = 1 + 2 + \dots + n = \frac{n(n+1)}{2}$$
$$\sum_{k=1}^n k^2 = 1^2 + 2^2 + \dots + n^2 = \frac{1}{6}n(n+1)(2n+1)$$

$\therefore$ by $(k+1)^3 - k^3 = 3k^2 + 3k + 1$

$$\sum_{k=1}^n \{(k+1)^3 - k^3\} = \sum_{k=1}^n \{3k^2 + 3k + 1\}$$

326から579までの和は?

42から283までの2乗和は?

確率 度関数

p. d. f. (probability density function)

(型)分布 continuous distribution

- ★ () 一様分布 uniform distribution
- ★ 正規分布 normal distribution
- ★ 標準正規分布 standard normal distribution
- ★ 指数分布 exponential distribution
- ★ γ 分布 Gamma distribution (χ^2 分布, 指数分布)
- ★ ベータ分布 Beta distribution

確率 度関数 p. d. f.

- (型)分布 continuous distribution
 - 確率変数 X の取る値が関数 $f(x)$ により, 以下で与えられている場合, X は 型の確率分布を つという

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

ただし,

$$\begin{cases} f(x) \geq 0 (-\infty \leq x \leq \infty), \\ \int_{-\infty}^{\infty} f(x) dx = 1 \end{cases} \iff \begin{cases} f(x_k) := P(X=x_k) \ (k=1,2,\dots) \\ f(x_k) \geq 0 \ (k=1,2,\dots), \\ \sum_{k=1}^{\infty} f(x_k) = 1 \end{cases}$$

確率 度関数
probability density function

- 分布関数 c.d.f., cumulative distribution function

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \iff F(x) = \sum_{u \leq x} f(u)$$

確率 度関数 p. d. f.

- 型確率変数の期待値と分散
 - 型確率変数 X の期待値

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

- 型確率変数 X の分散

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

確率 度関数 p. d. f.

型分布 continuous distribution

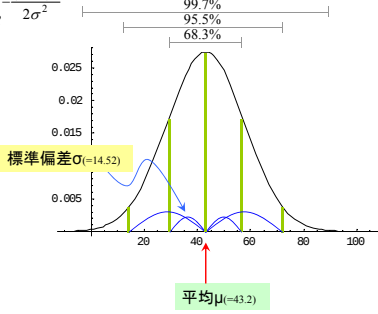
- () 一様分布 uniform distribution
 - 確率 度関数

$$f(x) = \begin{cases} 1 & [0,1] \\ 0 & o.w. \end{cases}$$

型分布 continuous distribution

- 正規分布 normal distribution
 - 確率 度関数
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

平均 μ , 分散 σ^2
 $N(\mu, \sigma^2)$



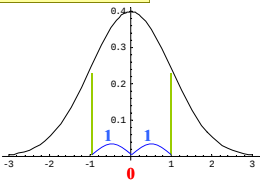
型分布 continuous distribution

- 標準正規分布 standard normal distribution
 - 確率変数の標準
 - 平均 μ , 分散 σ^2 の正規分布に従う確率変数 X について
$$X \rightarrow Z = \frac{X - \mu}{\sigma}$$

標準正規分布

確率変数 Z は、平均0, 分散1の正規分布に従う
 - 確率 度関数
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

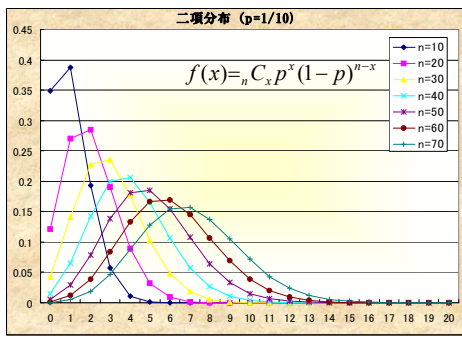
$N(0,1)$



型分布 discrete distribution

- 二項分布から正規分布へ...
 - 試行回数 n を大きくすると、二項分布は正規分布に
く
$$Bi(n, p) \xrightarrow{n \rightarrow \infty} N(\mu, \sigma^2) \begin{cases} \mu = np \\ \sigma^2 = np(1-p) \end{cases}$$
 - 試行回数 n が一定の時に、確率 p を0.5に
二項分布は正規分布に
く
$$Bi(n, p) \xrightarrow[n \rightarrow 0.5]{p} N(\mu, \sigma^2)$$

型分布 discrete distribution

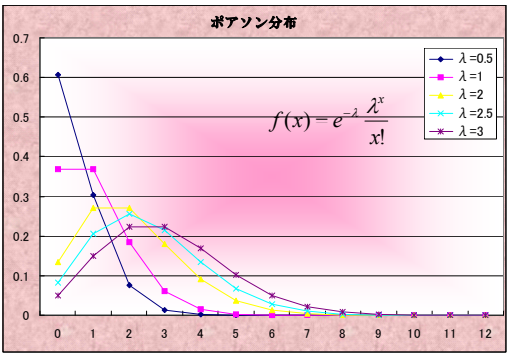
- 二項分布から正規分布へ...
 - 試行回数 n を大きくすると、二項分布は正規分布に
く

型分布 discrete distribution

- 正規分布による二項分布の
 - 例: 内 率
 - 500人の人に内 を するかどうか いたところ、275人
が指 すると えた。

内 率: $p = \frac{275}{500} = 0.55$
 - 内 率を p (不 率 $q = 1-p$) とすると、これは二項
分布となる。
 - 定では内 率は55 である。正規分布 を考
えると、 $\bar{x} = np = 500 \times 0.55 = 275$
 $\sigma = \sqrt{npq} = \sqrt{500 \times 0.55 \times 0.45} \approx \sqrt{124} \approx 11$
 - より、95% 区間における区間 定では、内 率
は $275 \pm 1.96 \times 11 \Leftrightarrow 253 \leq x \leq 297$ より50.6% 59.4%

型分布 discrete distribution

- ポアソン分布から正規分布へ...

型分布 continuous distribution

指数分布 exponential distribution

$Ex(\lambda)$

ポアソン分布に従って起きる事象の生起間隔を表現

確率 度関数

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

分布関数

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & (x \geq 0) \\ 0 & o.w. \end{cases}$$

期待値・分散

$$E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}$$

型分布 continuous distribution

指数分布 exponential distribution

- 例: サー スの待ち時間 (ケット売場の列)

時間をn等分し、ある区間kで かが入を したら $X_k=1$, そうでないなら $X_k=0$ とする確率変数 $X_1, X_2, X_3, \dots$ を考える。
ケット販売 時 = 0, 1人目の 入 時 = T とする。
 $X_1, X_2, X_3, \dots$ はパラ ータ $p = \lambda/n$ のベルヌーイ試行に従うとする。
→ 以上の 定で、確率変数 T の確率分布を求める

区間N+1で 入 → $T \approx (1/n)N$ (nが十分大きい時)
意の $t>0$ に対し、 k を、 $k \leq nt$ $k+1$ を たす 数とする
 $N+1$ は幾何分布に従う
→ $P(0 \leq T \leq t) \approx P(0 \leq N \leq nt) = \sum_{k=0}^k (1 - \lambda/n)^k \lambda/n$
 $= 1 - (1 - \lambda/n)^{k+1} \approx 1 - (1 - \lambda/n)^{nt} \xrightarrow{n \rightarrow \infty} 1 - e^{-\lambda t} = \int_0^t \lambda e^{-\lambda x} dx$

分布関数

確率 度関数

型分布 continuous distribution

ン 分布 Gamma distribution

$Ga(\alpha, \lambda)$

確率 度関数

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad (x \geq 0)$$

ン 関数 Gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad (\alpha > 0)$$
$$\Gamma(n) = (n-1)! \quad (n \in \mathbb{N})$$

nが自然数の時

→ $\begin{cases} Ga(n/2, 1/2) : \text{自 度 } n \text{ の } \chi^2 \text{ 分布} \\ Ga(1, \lambda) : \text{指数分布} \end{cases}$

型分布 continuous distribution

ベータ分布 Beta distribution

$Be(\alpha, \beta)$

確率 度関数

$$f(x) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (0 < x < 1)$$

ベータ関数 Beta function

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx \quad (\alpha, \beta > 0)$$
$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Coffee Break!

Monty-Hole Dilemma

確率的

3枚の の こうに
・ 万 ル (当たり)
・ (はずれ)
・ (はずれ)

が されているよ、
あなたは を1つだ
け んでいいのよ。

ところで、あなた
が ばなかった2つ
の のうち、 の
を くから、それ
を た後で、 く
を 変えてもいいよ。
さ、どうする？

Coffee Break!

Monty-Hole Dilemma

どうしても
いかない人
のため、 の
数を増やして
みましょ！

に
が100万も
あったらどう
かしら？

100万の からあなたが1つを んだ後で、 り99万9999の のうち
(はずれ)の99万9998の を いて せます。
それでもあなたは、 の を 変えない？ あなたの の は
かり的な に まれているのかしら？

参考



- 大学 学 統計学 「統計学入」 大学出 (1991)
- 村上 人「なる ど統計学」 (2002)
- G.Blom, L.Holst, D.Sqndell / 「確率論へようこそ」 シ ー・フ アラー (1995,2005[新])
- 市「 学 統計 析」 (2003)
- 大学 学 統計学 「自然 学の統計学」 大学出 (1992)
- 白 ニ「例題で学 Excel統計入」 出 (2001)
- J.Matousek, J.Nesetril / 根上生 ・中本 「離散数学への 待 上」 シ ー・フ アラー (2002)
- 「工学基 離散数学とその 用」数理工学 (2003)
- B.Schechter / グラベルロー 「My Brain is Open」 立出 (2003)