確率と統計(0)
「確率分布の例(第6章)」

- 担当教員： 杉山　将 (計算工学専攻)
- 居室： W8E-505
- 電子メール: sugi@cs.titech.ac.jp
- 授業のウェブサイト：
  http://sugiyama-www.cs.titech.ac.jp/~sugi/

主な離散型の確率分布

- 一様分布
- 二項分布
- 超幾何分布
- ポアソン分布
- 負の二項分布
- 幾何分布

一様分布

- **離散一様分布**(uniform distribution of discrete type)：
  N 個の事象が等確率で起こる

$$f(x) = \frac{1}{N} \text{ for } x = 1, 2, \ldots, N$$

- $f(x)$ は

$$f(x) \geq 0 \qquad \sum_{x=1}^{N} f(x) = 1$$

を明らかに満たすので，確率関数である

一様分布の性質

- **期待値**： $$E(X) = \frac{N+1}{2}$$
- **分散**： $$V(X) = \frac{N^2-1}{12}$$
- **証明**： $$E[X] = \sum_{x=1}^{N} xf(x) = \sum_{x=1}^{N} x/N = (N+1)/2$$

$$\sum_{x=1}^{N} x = N(N+1)/2$$

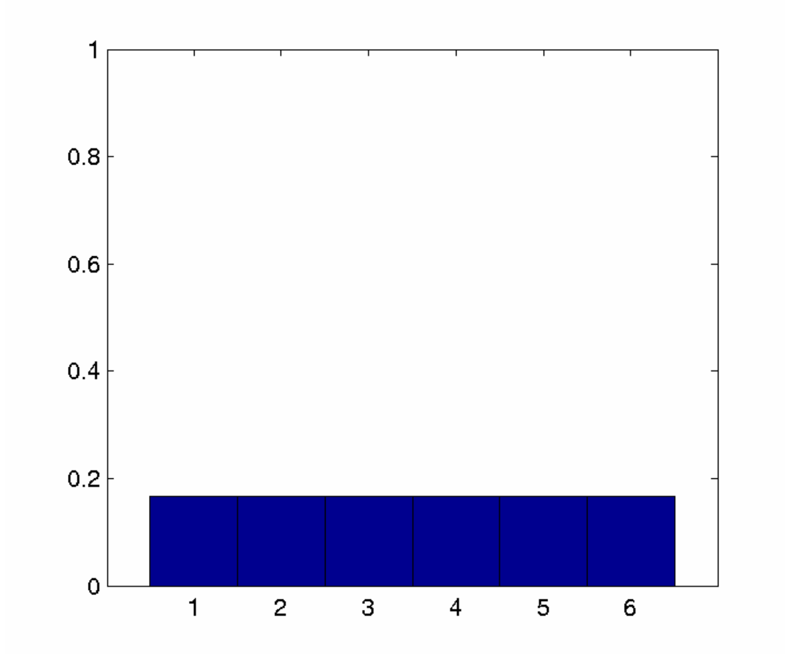
$$E[X^2] = \sum_{x=1}^{N} x^2 f(x) = \sum_{x=1}^{N} x^2/N = (N+1)(2N+1)/6$$

$$\sum_{x=1}^{N} x^2 = N(N+1)(2N+1)/6$$

$$V(X) = E[X^2] - \{E[X]\}^2 = (N^2-1)/12$$

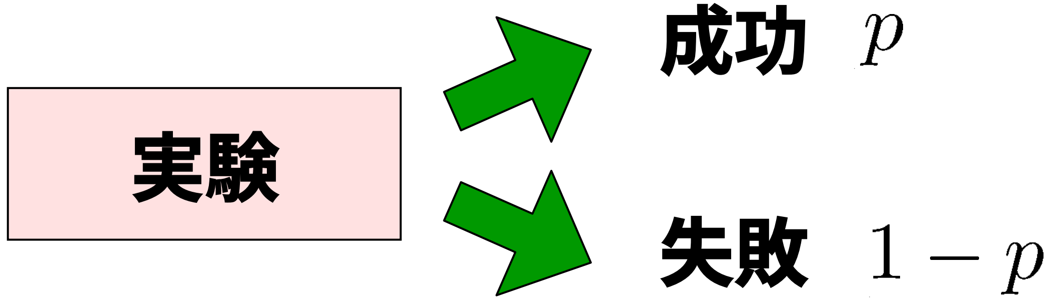
一様分布の例

- 例：さいころの目 $(N = 6)$



ベルヌーイ試行

- **ベルヌーイ試行**(Bernoulli trials)：成功する確率と失敗する確率がそれぞれ p と $1-p$ の実験を同じ条件で独立に繰り返す



- 例：コインを投げて表が出るか裏が出るか

$$p = 0.5$$

二項分布

- **二項分布**(Binomial distribution)：
  n 回のベルヌーイ試行に対して，実験が x 回成功する確率
  - x 回成功，$n-x$ 回失敗：　$p^x(1-p)^{n-x}$
  - 順番を入れ替えたときの組み合わせ数：　${}_nC_x$

$$f(x) = {}_nC_x p^x (1-p)^{n-x}$$

$$\text{for } x = 0, 1, \ldots, n$$

- 二項分布を $Bi(n, p)$ で表す．
- $Bi(1, p)$ を特に**ベルヌーイ分布**(Bernoulli distribution)と呼ぶ．

$${}_nC_r = \frac{n!}{r!(n-r)!}$$

二項分布の性質

■ **積率母関数:**

$$M_X(t) = (pe^t + q)^n$$

証明:二項定理

$$\sum_x {}_nC_x p^x q^{n-x} = (p+q)^n$$

より

$$M_X(t) = \sum_x e^{tx} {}_nC_x p^x q^{n-x}$$
$$= \sum_x {}_nC_x (pe^t)^x q^{n-x} = (pe^t + q)^n$$

(Q. E. D.)

二項分布の性質 (続き)

- **期待値**: $E(X) = np$

  成功確率が p で n 回実験を行ったとき, 平均成功回数が np であることは直感的にもわかる

- **分散**: $V(X) = np(1-p)$

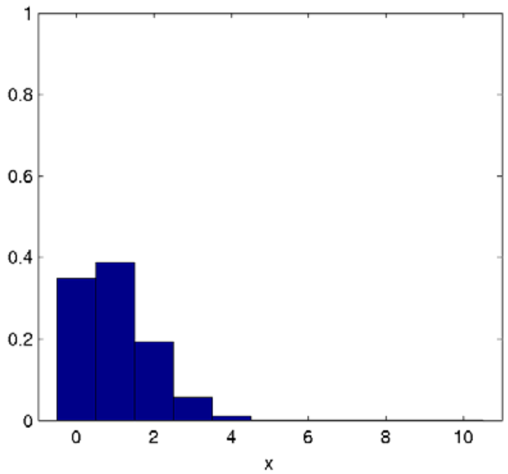
  分散は $p = 0.5$ のとき最大になる. これは, 成功と失敗の確率が五分五分のときに予想が難しいという直感と合っている

- **演習: 上記の期待値の分散を証明せよ**

二項分布の例

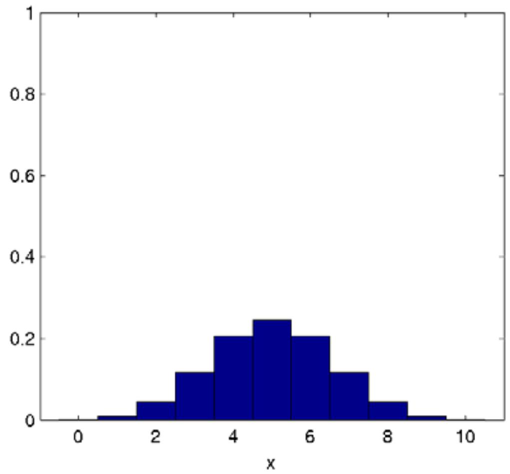
- 実験の成功率を変えたとき

$$n = 10$$



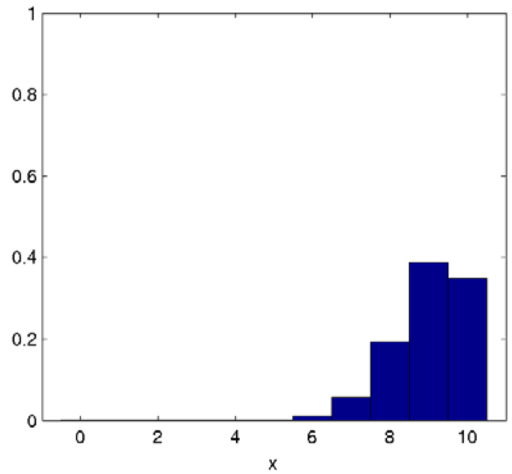
$$p = 0.1$$

成功率が低い



$$p = 0.5$$

五分五分

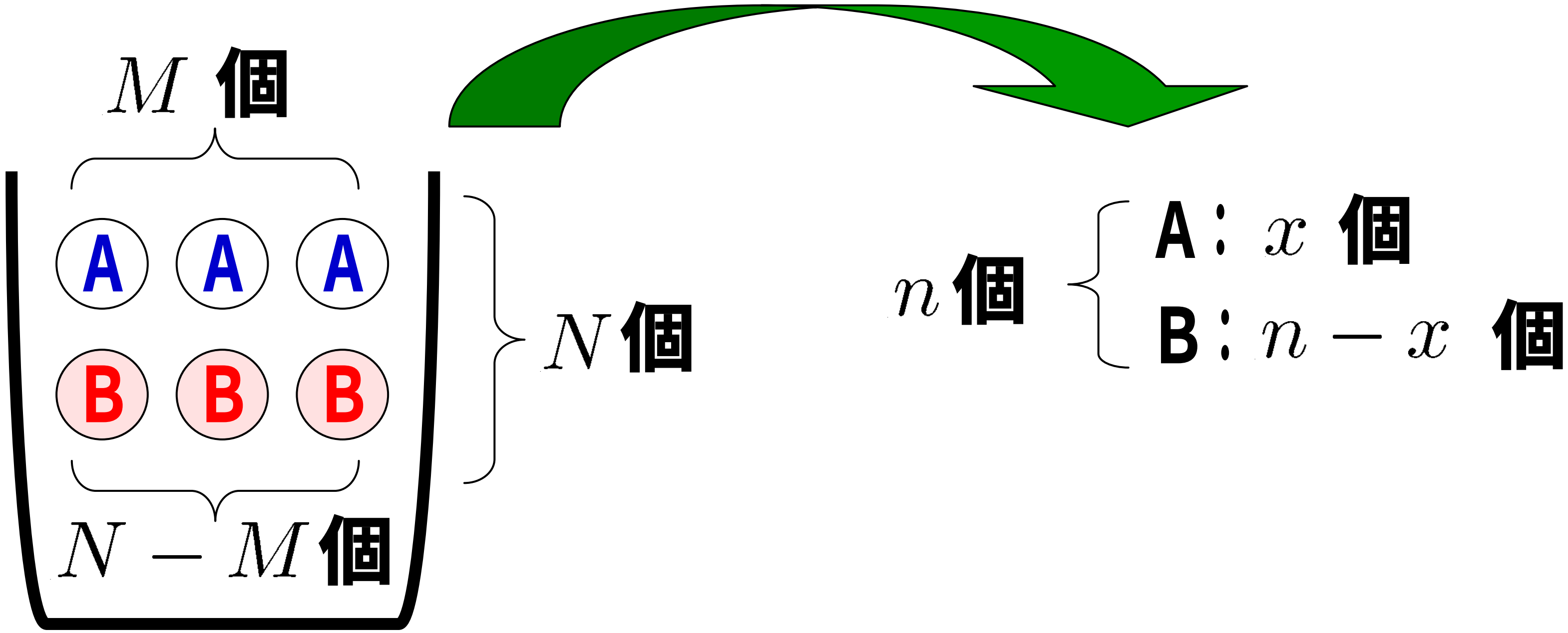


$$p = 0.9$$

成功率が高い

復元抽出と非復元抽出

- AがM 個, Bが $N-M$ 個, 合計 N 個の玉が入っている袋から無作為に玉を n 個取り出す.

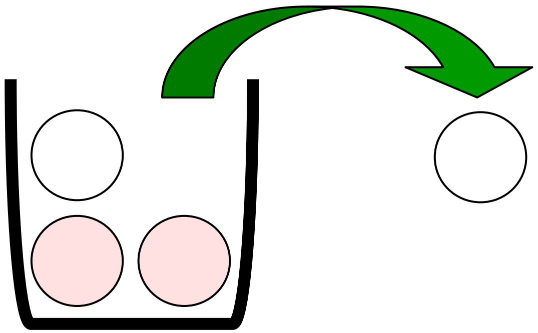


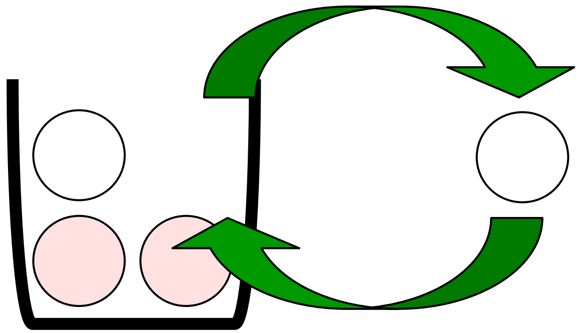
復元抽出と非復元抽出 (続き)

- **復元抽出**(sampling with replacement)：
  ひとつ玉を取り出したら，それを元に戻してから次の玉を取り出す

  ⇒ **Aが x 個出てくる確率は**
  二項分布 $Bi(n, M/N)$ に従う

- **非復元抽出**(sampling without replacement)：
  ひとつ玉を取り出したら，それを元に戻さずに次の玉を取り出す





超幾何分布

- **超幾何分布**(hypergeometric distribution)：
  非復元抽出したときに, Aが x 個出てくる確率
  - Aが x 個出てくる組み合わせ数： ${}_{M}C_{x}$
  - Bが $n-x$ 個出てくる組み合わせ数： ${}_{N-M}C_{n-x}$
  - 合計 n 個取り出す組み合わせ数： ${}_{N}C_{n}$

$$f(x) = \frac{{}_{M}C_{x} \times {}_{N-M}C_{n-x}}{{}_{N}C_{n}}$$

- x の最小値, 最大値は

$$x_{max} = \min(n, M)$$

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

- 名称は微分方程式論の超幾何級数に由来

便利な恒等式

- 確率関数の性質 $\sum_x f(x) = 1$ より,

$$ {}_N C_n = \sum_{x=0}^{n} {}_M C_x \; {}_{N-M} C_{n-x} $$

- 一応確認のため証明する:
  - 二項定理より

$$ (1+z)^N = \sum_{n=0}^{N} {}_N C_n \; z^n $$

  - 一方, 分解してから二項定理を使えば,

$$ (1+z)^N = (1+z)^M (1+z)^{N-M} = \sum_{x=0}^{M} \sum_{y=0}^{N-M} {}_M C_x \; {}_{N-M} C_y \; z^{x+y} $$

$$ = \sum_{n=0}^{N} z^n \sum_{x=0}^{n} {}_M C_x \; {}_{N-M} C_{n-x} \qquad n = x + y $$

z^n の係数が等しいことから上記の恒等式が得られる.

(Q. E. D.)

超幾何分布の性質

- 期待値:

$$E(X) = \frac{nM}{N}$$

- 分散:

$$V(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

- 以下, 期待値を証明する. 分散は宿題.

超幾何分布の期待値の証明

$$E[X] = \frac{1}{{}_N C_n} \sum_{x=0}^{n} x \; {}_M C_x \; {}_{N-M} C_{n-x} \quad = \frac{1}{{}_N C_n} \sum_{x=1}^{n} x \; {}_M C_x \; {}_{N-M} C_{n-x}$$

$$= \frac{M}{{}_N C_n} \sum_{x=1}^{n} {}_{M-1} C_{x-1} \; {}_{N-M} C_{n-x} \qquad {}_M C_x = \frac{M}{x} {}_{M-1} C_{x-1}$$

$$= \frac{M}{{}_N C_n} \sum_{x=0}^{n-1} {}_{M-1} C_x \; {}_{N-M} C_{n-x-1} \qquad x \leftarrow x-1$$

$$= \frac{nM}{N} \frac{1}{{}_{N-1} C_{n-1}} \sum_{x=0}^{n-1} {}_{M-1} C_x \; {}_{N-M} C_{n-x-1}$$

恒等式 $\boxed{{}_N C_n = \sum_{x=0}^{n} {}_M C_x \; {}_{N-M} C_{n-x}}$ **で**

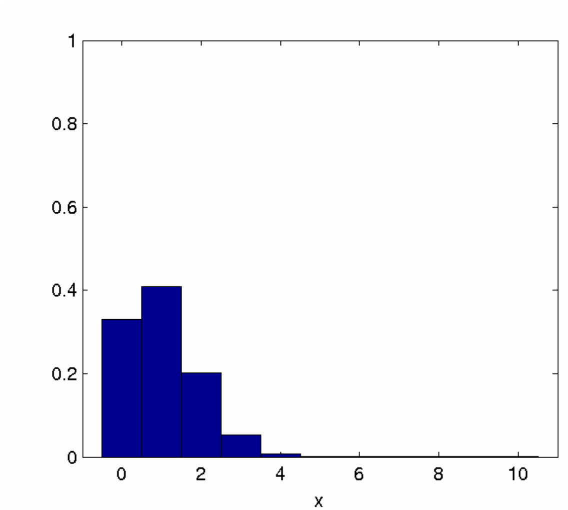
$M \leftarrow M-1, N \leftarrow N-1, n \leftarrow n-1$ **とおけば,**

${}_{N-1} C_{n-1} = \sum_{x=0}^{n-1} {}_{M-1} C_x \; {}_{N-M} C_{n-x-1}$ **となり,** $E(X) = \frac{nM}{N}$ **を得る.**

超幾何分布の例

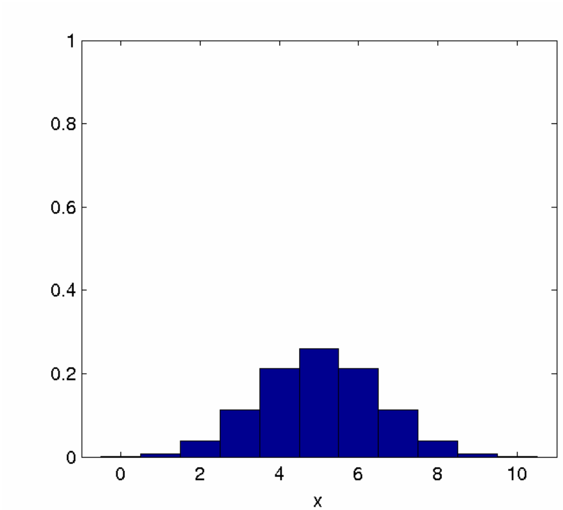
- 袋に入っている玉Aの数を変えたとき

$$N = 100, n = 10$$



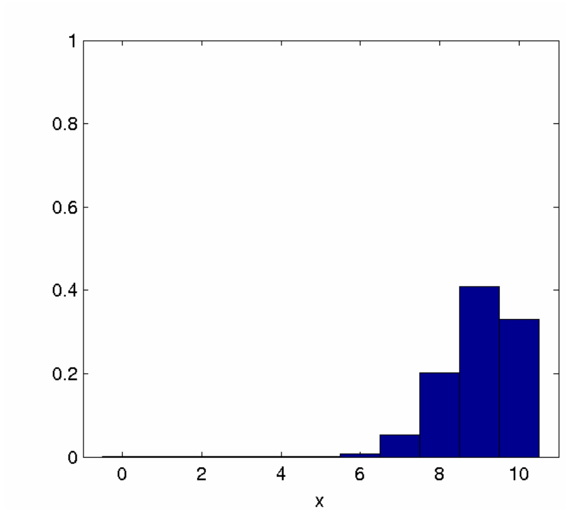
$$M = 10$$

Aが少ない



$$M = 50$$

A＝B



$$M = 90$$

Aが多い

小レポート

1. $\{a, a+1, \ldots, b\}$ $(a < b)$ 上の一様分布の期待値と分散を求めよ．
2. 超幾何分布の分散が次式で与えられることを示せ．

$$V(X) = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

ヒント：分散は

$$V[X] = E[X(X-1)] + E[X] - (E[X])^2$$

と表すことができる．期待値のときと同様な計算をすれば，次式が成り立つ．

$$E[X(X-1)] = \frac{n(n-1)M(M-1)}{N(N-1)}$$

恒等式　${}_N C_n = \sum_{x=0}^{n} {}_M C_x \; {}_{N-M} C_{n-x}$ も役立つ．

小レポート (続き)

3. (発展的内容) ポアソン分布と負の二項分布の積率母関数, 期待値, 分散が次のように表せることを証明せよ.

A) ポアソン分布 $Po(\lambda)$: $M_X(t) = \exp\left\{\lambda(e^t - 1)\right\}$

$$E(X) = \lambda$$

$$V(X) = \lambda$$

B) 負の二項分布 k, p : $M_X(t) = p^k\{1-(1-p)e^t\}^{-k}$

$$E(X) = k(1-p)/p$$

$$V(X) = k(1-p)/p^2$$

ヒント:

$$e^\lambda = \sum_x \frac{\lambda^i}{x!}$$

$$\sum_x {}_aC_x t^x = (1+t)^a$$

$$E[X] = M'_X(0)$$

$$V(X) = M''_X(0) - M'_X(0)^2$$