確率と統計（O）
「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
- 超幾何分布
- 二項分布
- 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**: $E(X) = \dfrac{N+1}{2}$
- **分散**: $V(X) = \dfrac{N^2-1}{12}$
- **証明**: $E[X] = \sum_{x=1}^{N} x f(x) = \sum_{x=1}^{N} x/N = (N+1)/2$

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

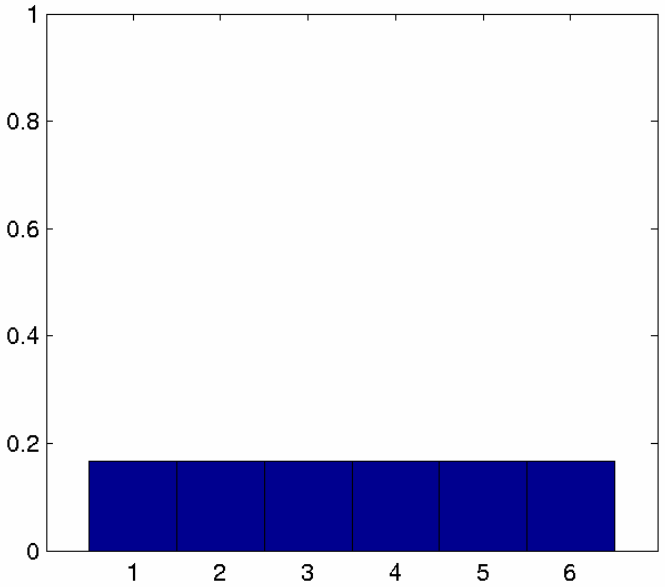
$$E[X^2] = \sum_{x=1}^{N} x^2 f(x) = \sum_{x=1}^{N} x^2/N = (N+1)(2N+1)/6$$

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

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

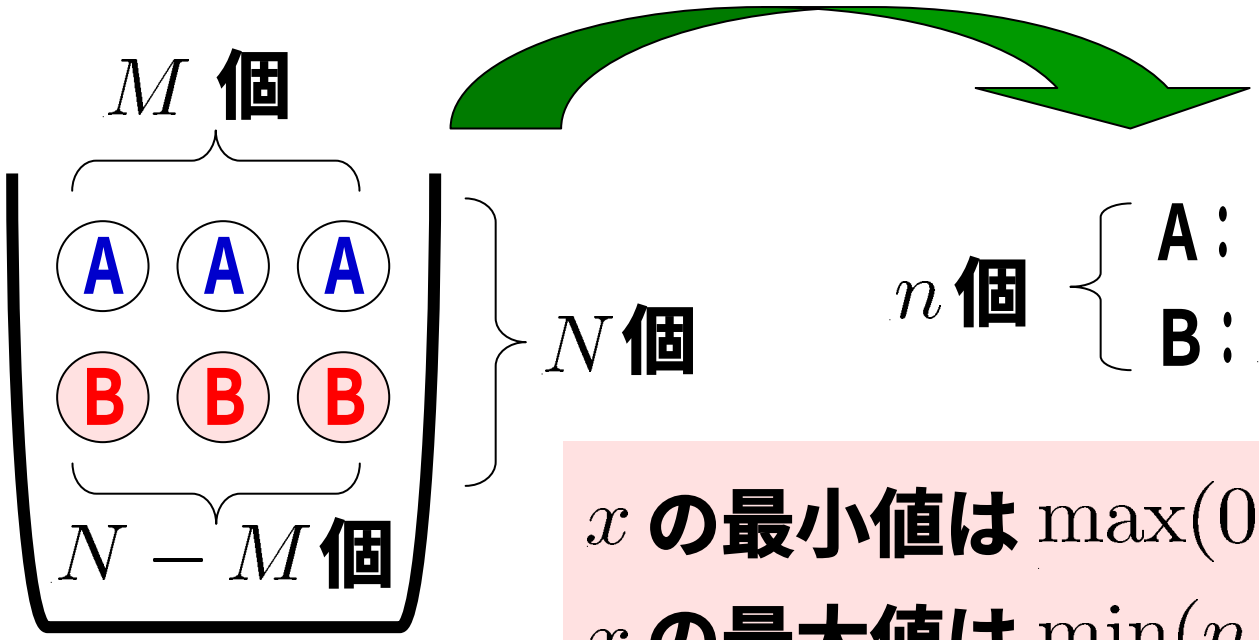
一様分布の例

- 例：さいころの目 $(N = 6)$



超幾何分布

- 袋の中にAがM個，Bが $N-M$個，合計 N個の玉が入っている
- この袋から無作為に玉を n 個取り出したときのAの数を x 個，Bの数を $n-x$ 個とする



x の最小値は $\max(0, n-(N-M))$

x の最大値は $\min(n, M)$

超幾何分布 (続き)

- Aが x 個, Bが $n-x$ 個である確率は

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

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

$${}_n C_r = \frac{n!}{r!(n-r)!} = \frac{n \times (n-1) \times \cdots \times (n-r+1)}{r \times (r-1) \times \cdots \times 1}$$

- これを**超幾何分布**(hypergeometric distribution)とよぶ
- 名称は, 微分方程式論の超幾何級数に由来

超幾何分布 (続き)

- $f(x)$ が確率関数であることの証明
  - $f(x) \geq 0$ は明らか
  - 二項定理より $(1+z)^N = \sum_{n=0}^{N} {}_N C_n \, z^n$ ．一方，

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

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

z^n の係数を比較すれば $\quad {}_N C_n = \sum_{x=0}^{n} {}_M C_x \; {}_{N-M} C_{n-x}$

これより，

$$\sum_x f(x) = 1$$

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

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

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

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

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

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

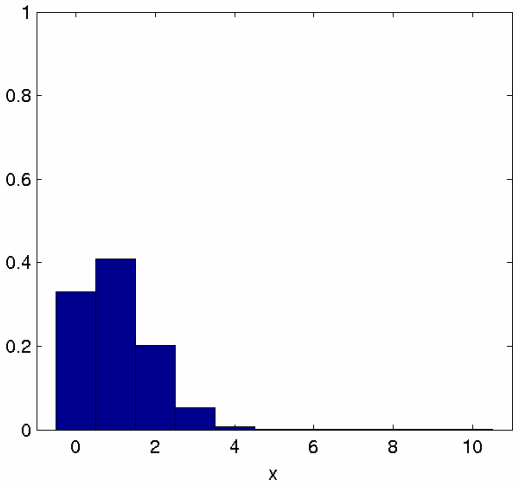
$M \leftarrow M-1, N \leftarrow N-1, n \leftarrow n-1$ **とおけば,**

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

超幾何分布の例

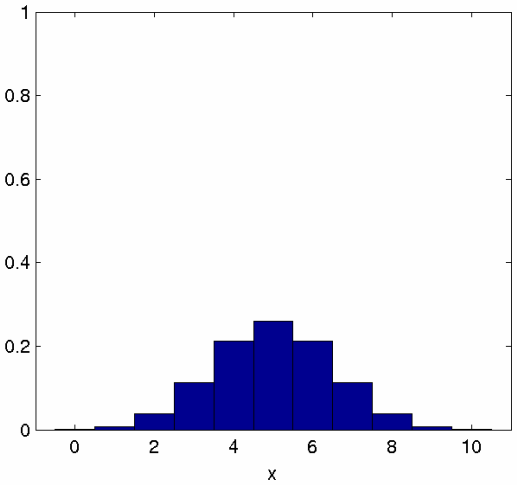
- 袋に入っている玉Aの数を変えたとき

$$N = 100, n = 10$$



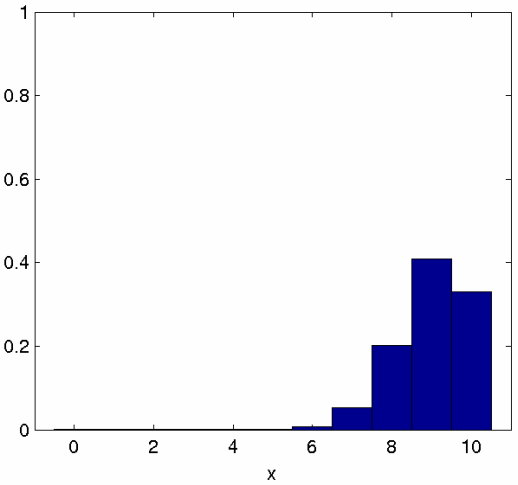
$M = 10$

Aが少ない



$M = 50$

A＝B

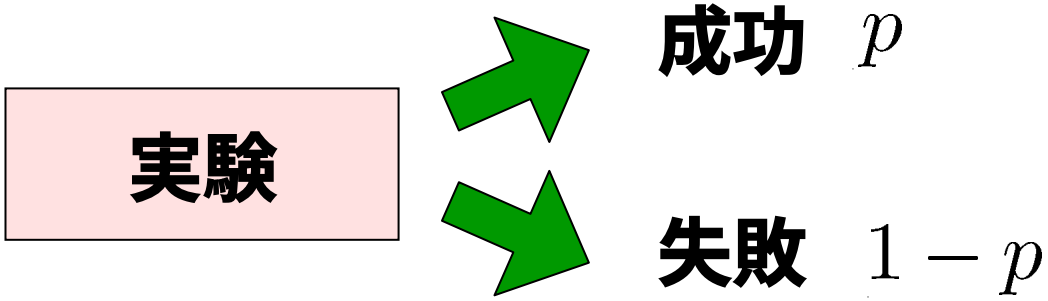


$M = 90$

Aが多い

ベルヌーイ試行

- ある実験の成功する確率と失敗の確率がそれぞれ p と $1-p$.



- この実験を同じ条件で独立に n 回繰り返す
- これを**ベルヌーイ試行**(Bernoulli trials)という
- 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、ベルヌーイ分布

- n回のベルヌーイ試行に対して，実験が x 回成功して $n-x$ 回失敗する確率は

$$f(x) = {}_nC_x p^x (1-p)^{n-x}$$
$$\text{for } x = 0, 1, \ldots, n$$

- これを**二項分布**(Binomial distribution)とよび，$Bi(n,p)$ で表す
- $n = 1$ のとき，$Bi(1,p)$ を特に**ベルヌーイ分布**(Bernoulli distribution)と呼ぶ

二項分布 (続き)

- $f(x)$ が確率関数であることの証明
  - $f(x) \geq 0$ は明らか
  - 二項定理より

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

が成り立つ．$p+q=1$ を用いれば

$$\sum_x f(x) = 1$$

が得られる

二項分布の性質

- **期待値:**

$$E(X) = np$$

- **分散:**

$$V(X) = npq \qquad q = 1 - p$$

- **積率母関数:**

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

- **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**

二項分布の性質(証明)

- 二項定理より積率母関数は

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

積率母関数の1階，2階微分はそれぞれ

$$M'_X(t) = npe^t(pe^t + q)^{n-1}$$
$$M''_X(t) = npe^t(pe^t + q)^{n-1} + n(n-1)p^2e^{2t}(pe^t + q)^{n-2}$$

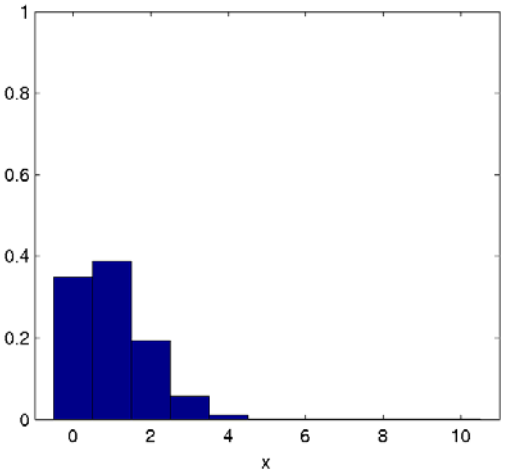
これより

$$E[X] = M'_X(0) = np$$
$$V(X) = E[X^2] - \{E[X]\}^2 = M''_X(0) - M'_X(0)^2$$
$$= np + n(n-1)p^2 - n^2p^2 = np(1-p)$$

二項分布の例

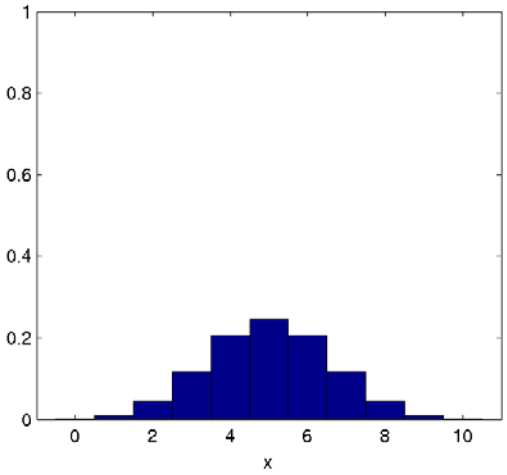
- 実験の成功率を変えたとき

$n = 10$



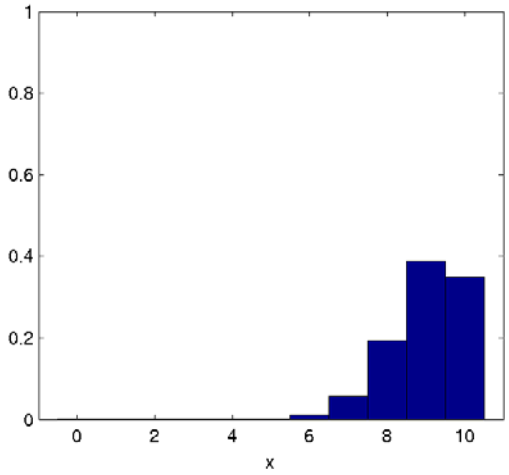
$p = 0.1$

成功率が低い



$p = 0.5$

五分五分

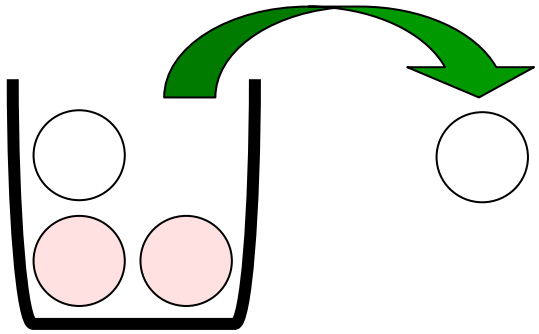


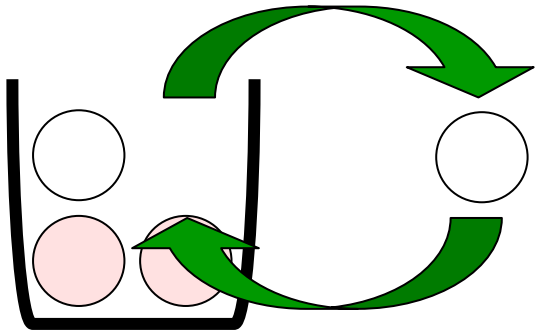
$p = 0.9$

成功率が高い

復元抽出と非復元抽出

- 超幾何分布では，n 個の玉を袋から取り出すときに，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．これを**非復元抽出**(sampling without replacement)という．
- 二項分布では，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．これを**復元抽出**(sampling with replacement)という．





小レポート

- 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 $${}_NC_n = \sum_{x=0}^{n} {}_MC_x \; {}_{N-M}C_{n-x}$$ も役立つ．