確率と統計(O)
「離散型確率分布の例(第6章)」

- 担当教員: 杉山 将 (計算工学専攻)
- 居室: W8E-406
- 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/

講義計画(シラバス)

- 確率と統計の基礎
- 確率変数，確率分布
- 積率，積率母関数
- 離散型の確率分布の例
- 連続型の確率分布の例
- 確率不等式， 擬似乱数
- 多次元の確率分布
- 大数の法則，中心極限定理
- 統計的推定，仮説検定

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(discrete uniform distribution):
  N 個の事象が等確率で起こる

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

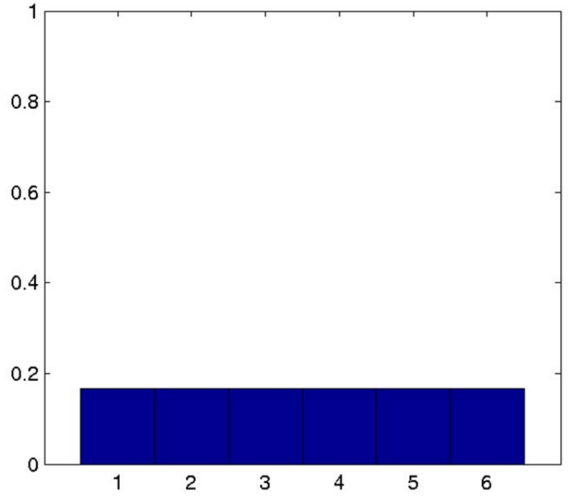
- 期待値:

$$E(X) = \frac{N+1}{2}$$

- 分散:

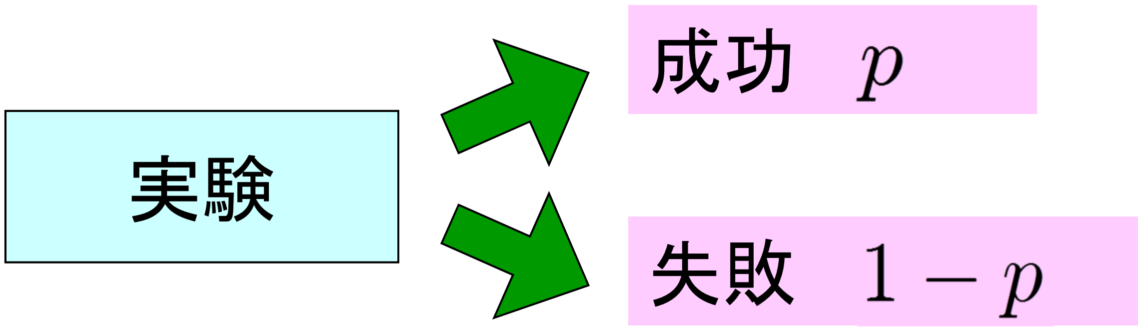
$$V(X) = \frac{N^2-1}{12}$$

- 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p, 失敗する確率が$1-p$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: ${}_nC_x$

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

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

- 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.

二項分布の性質

- 積率母関数:

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

$$q = 1 - p$$

証明:二項定理

$$\sum_{x=0}^{n} {}_nC_x p^x q^{n-x} = (p+q)^n$$

より

$$M_X(t) = \sum_{x=0}^{n} e^{tx} {}_nC_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} {}_nC_x (pe^t)^x q^{n-x} = (pe^t + q)^n$$

(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: $V(X) = np(1-p)$

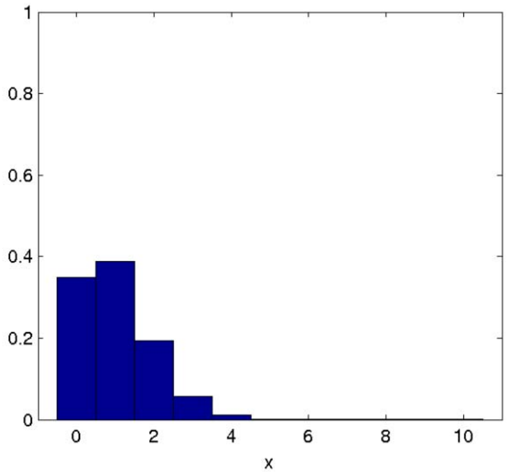
分散は $p = 0.5$ のとき最大になる．これは，成功と失敗の確率が五分五分のときに予想が難しいという直感と合っている

- 演習: 上記の期待値と分散を証明せよ

二項分布の例

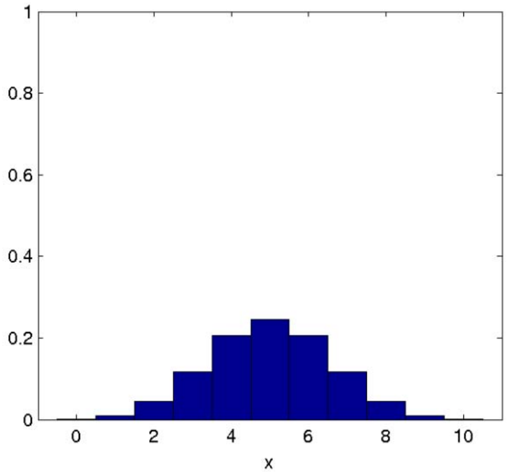
■ 実験の成功率を変えたとき

$n = 10$



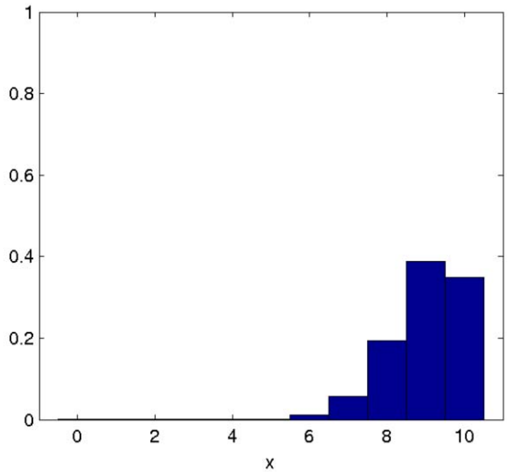
$p = 0.1$

成功率が低い



$p = 0.5$

五分五分

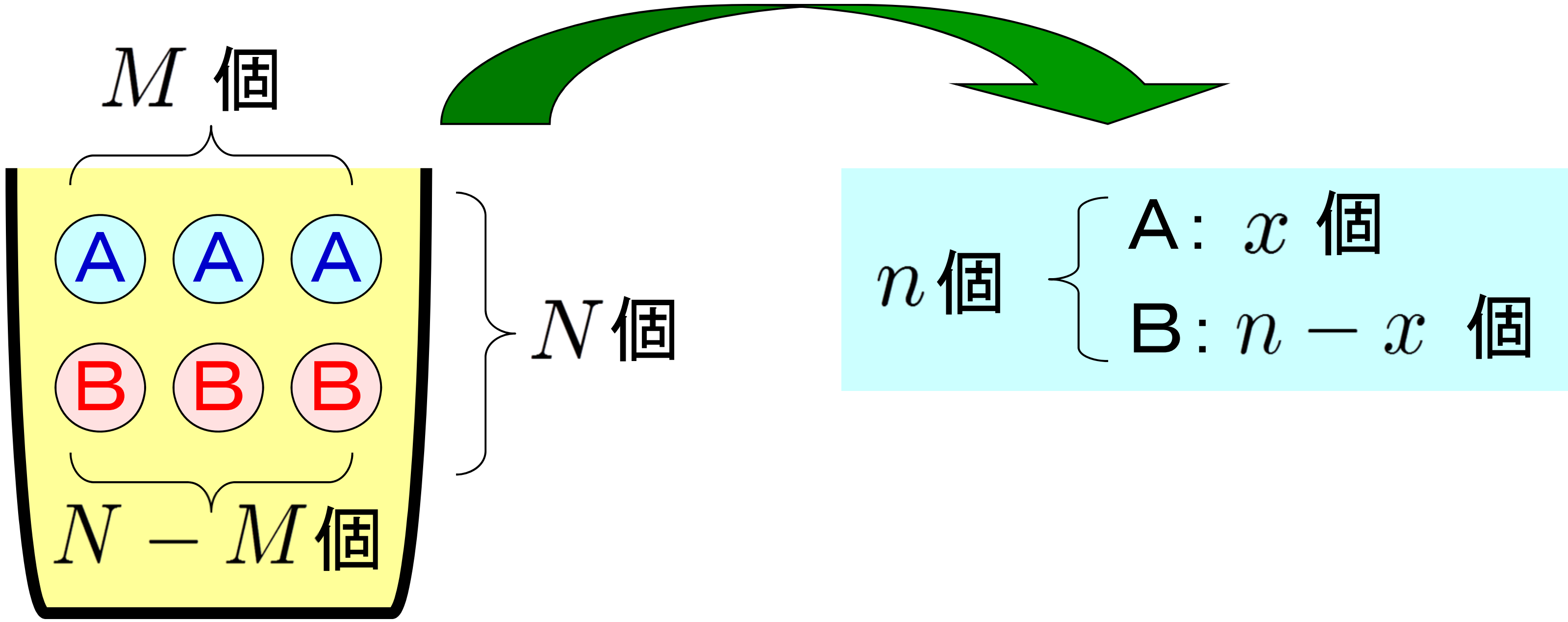


$p = 0.9$

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.

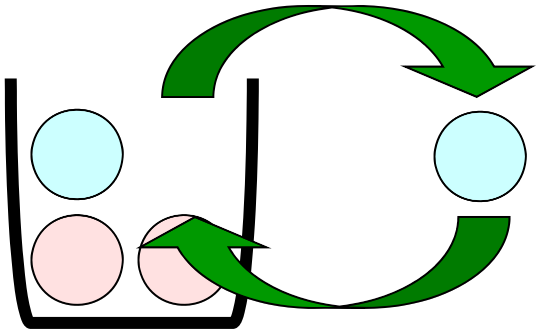


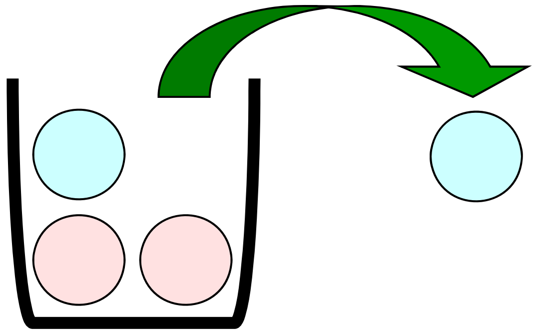
復元抽出と非復元抽出（続き）

- 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 個取り出す総組み合わせ数：${}_{N}C_{n}$

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

“ ${}_{N}C_{n}$ 回のうちの ${}_{M}C_{x} \times {}_{N-M}C_{n-x}$ 回”

- 名称は微分方程式論の超幾何級数に由来.

超幾何分布の性質

- xは 0 から n までの好きな値を取れるとは限らない.
  - xの最小値: $x_{min} = \max\{0, n-(N-M)\}$
  - xの最大値: $x_{max} = \min(n, M)$
- 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. 分散の証明は宿題.

超幾何分布の期待値の証明

$$E[X] = \frac{1}{{}_NC_n}\sum_{x=0}^{n} x\ {}_MC_x\ {}_{N-M}C_{n-x}$$

（$x=0$ の項はゼロ）

$$= \frac{1}{{}_NC_n}\sum_{x=1}^{n} x\ {}_MC_x\ {}_{N-M}C_{n-x}$$

$${}_MC_x = \frac{M}{x}{}_{M-1}C_{x-1}$$

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

$x \leftarrow x-1$

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

$${}_NC_n = \frac{N}{n}{}_{N-1}C_{n-1}$$

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

超幾何分布の期待値の証明(続き)

- 確率関数の性質 $\sum_x f(x) = 1$ より,

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

- この式で

$$ M \leftarrow M-1, N \leftarrow N-1, n \leftarrow n-1 $$

とおけば,

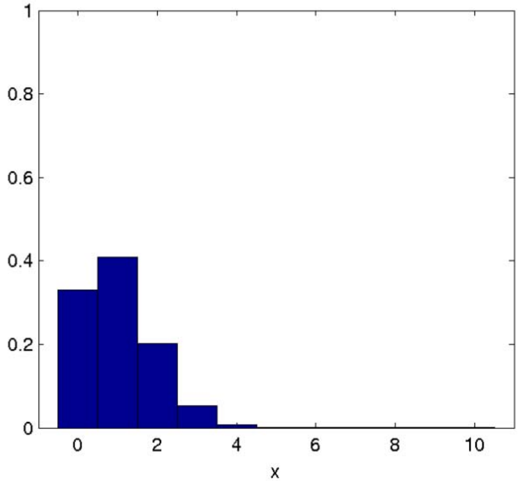
$$ {}_{N-1} C_{n-1} = \sum_{x=0}^{n-1} {}_{M-1} C_x \ {}_{N-M} C_{n-x-1} $$

となり, $E(X) = \dfrac{nM}{N}$ を得る. （Q. E. D. ）

超幾何分布の例

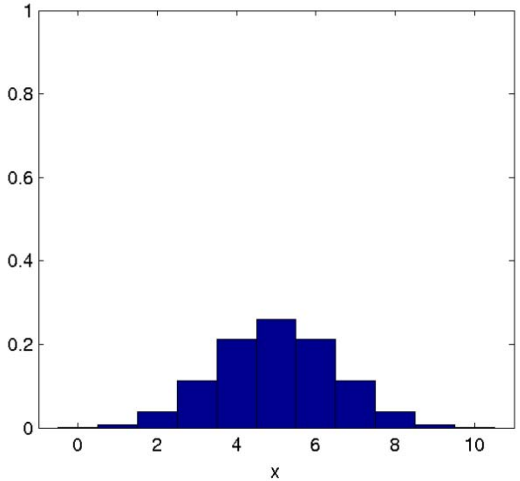
- 袋に入っている玉Aの数を変えたとき

$N = 100, n = 10$



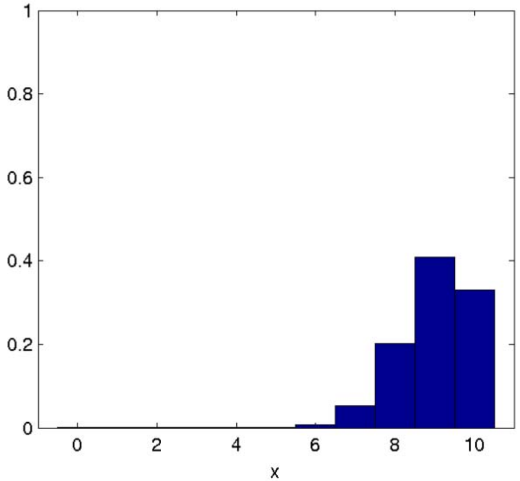
$M = 10$

Aが少ない



$M = 50$

A＝B



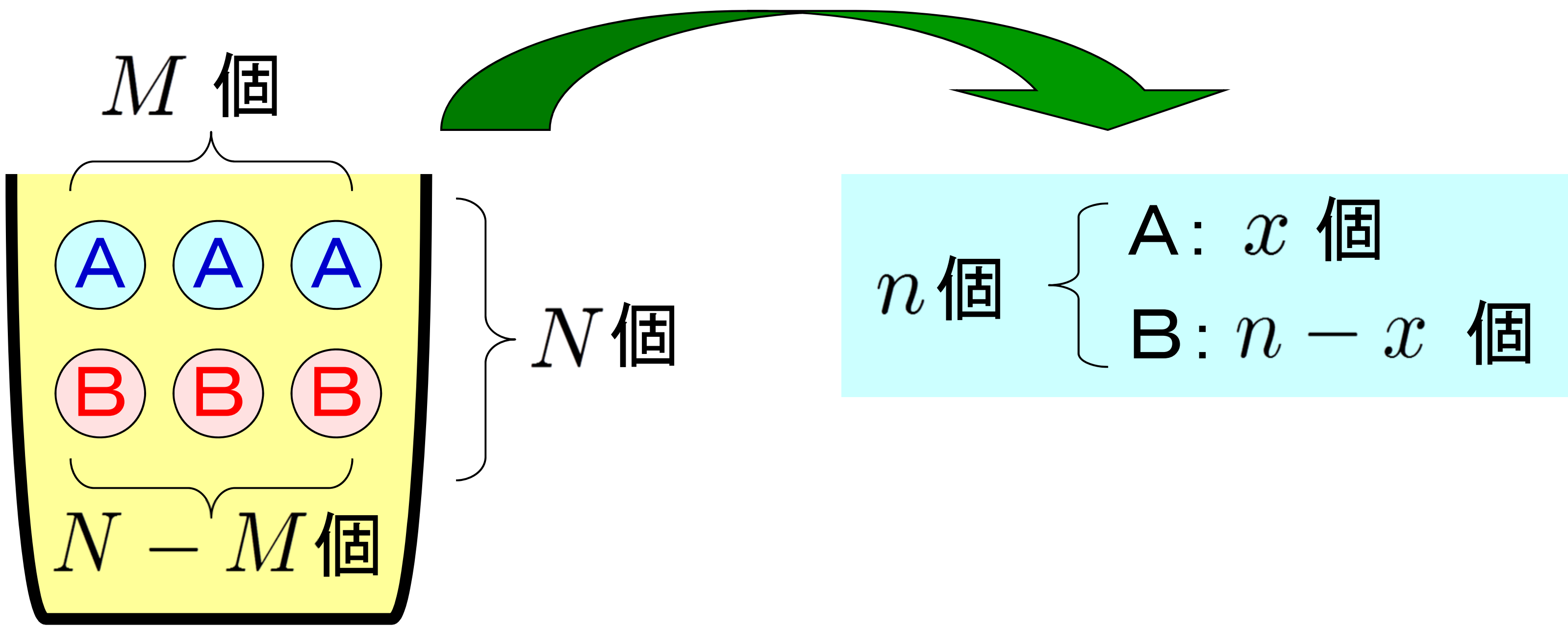
$M = 90$

Aが多い

- この例では，見た目は二項分布とそれほど変わらない．
- しかし，玉を戻さない超幾何分布の方が式は複雑．

超幾何分布と二項分布

- N が十分に大きいとき，玉を袋に戻しても戻さなくても大して変わらない．
- 実際，$N \to \infty$, $M/N \to p$ のとき，超幾何分布は二項分布と一致する．



まとめ

- 離散分布
  - 一様分布
  - 二項分布
  - 超幾何分布
- ベルヌーイ試行
- 復元抽出と非復元抽出

宿題

1. $\{a, a+1, \ldots, b\}$ $(a < b)$上の一様分布

$$f(x) = \frac{1}{b-a+1} \text{ for } x = a, a+1, \ldots, b$$

の期待値と分散が次式で与えられることを証明せよ.

$$E[X] = \frac{a+b}{2} \qquad V(X) = \frac{(b-a+1)^2 - 1}{12}$$

ヒント:

$$\sum_{x=1}^{n} x = \frac{n(n+1)}{2} \qquad \sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}$$

宿題(続き)

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,

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

が成り立つ. また恒等式

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

も役立つ.