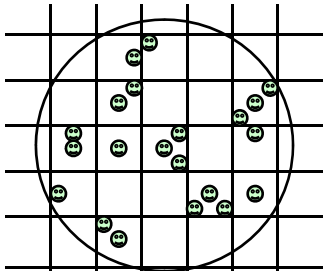
離散型分布 discrete distribution

- 二項分布とポアソン分布
  - 例：単一時間に発生する事故件数は?
    - 一日 m件の事故が発生したとする．これを1時間毎，1分毎，1秒毎と縮めていき，1刻みに1件の事故が発生するようにし(同時刻に2件発生することはないとする)，その刻み数をnとする．
    - すると，この話は n個の刻みの中で 1件事故が発生するかしないかとみなすことができる．即ち，n個の刻みの中から事件が発生した x個の刻みの個数を考えることになる
    - 事故発生率は $p = m / n =$ 一定！

      (ポアソン分布の期待値)
    - このときの $m = n\,p$ は二項分布の期待値！

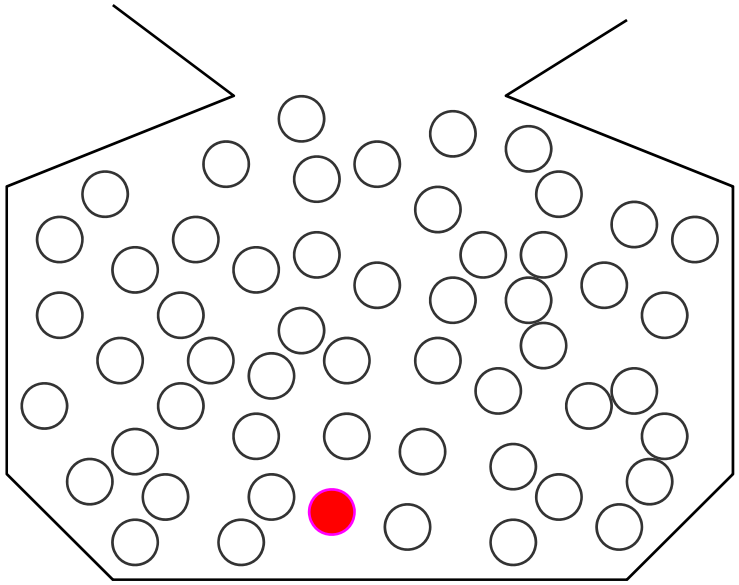
離散型分布 discrete distribution

- ポアソン分布の例
  - 例1：飛行機事故（事故はめったに起きない）
    - 飛行機事故の確率1/10万，飛行機搭乗回数を1万回としたとき，一度も事故にあわない確率は?
  - 例2：大量生産品の不良品数（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：爆撃命中数（めったに当たらない）
    - 第二次大戦中のドイツ軍の砲弾命中精度は$\lambda=0.93$のポアソン分布に従うという．1000発打って1発当たる確率は?
  - 例4：薬の副作用
    - 副作用の確率が1/200の薬を5000人が服用したとき，30人以上に副作用が出る確率は?
  - 例5：生物・植物の生態・繁茂状況を示す分布
    - 単位面積あたりのバクテリアの個数



演習4

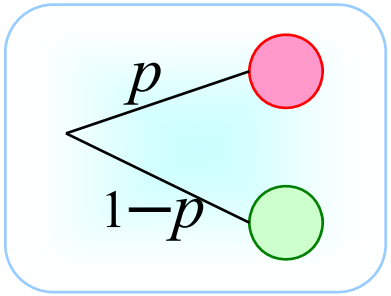
- ポアソン分布を求めよう
  - 赤玉1個，白玉99個入っている袋から1つ取り出しては戻すという行為を5回行ったとき，赤球が出る回数の確率分布（ポアソン分布）を求めてみよう！



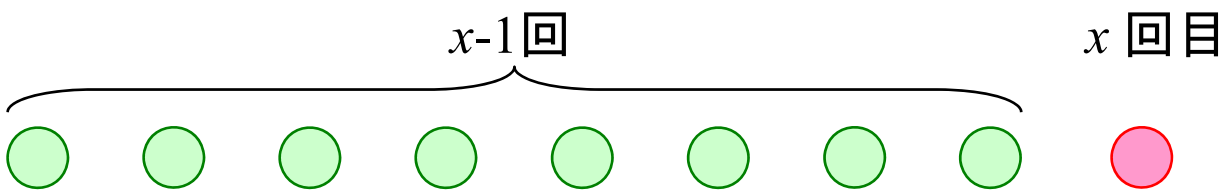
離散型分布 discrete distribution

- <u>幾何分布 geometric distribution</u>
  - ベルヌーイ試行において，試行回数を決めずに初めてある事象が起こるまでの試行回数を x とするときの X=x の確率分布

$$f(x) = p(1-p)^{x-1} \quad (x = 1,2,\cdots)$$



幾何数列（等比数列）の形なので，幾何分布とよばれる

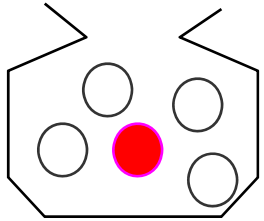


  - 幾何分布は，時間を離散的に(1,2,3,…)考えるとき，初めて何かが起こるまで<u>待つ時間の長さの</u><u>確率分布</u>〔（離散的な）待ち時間分布〕

離散型分布 discrete distribution

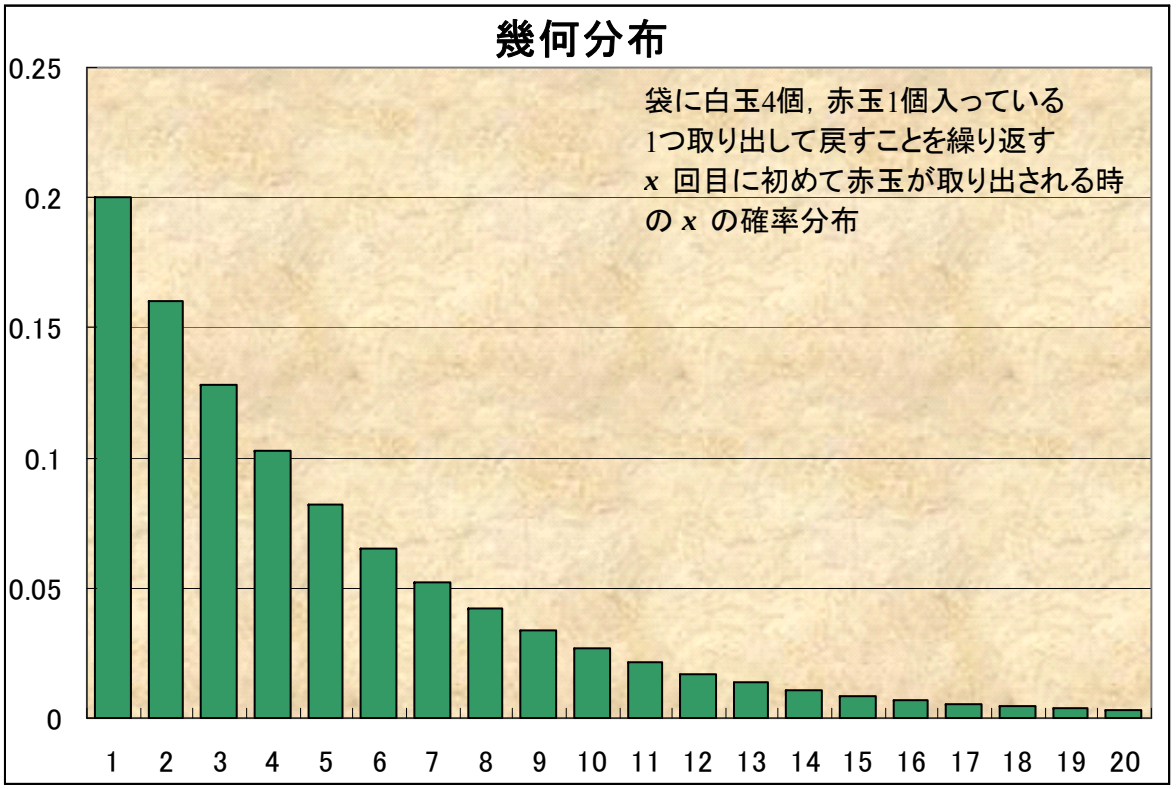
- 幾何分布
  - 確率分布

$$f(x) = p(1-p)^{x-1},\ (x = 1,2,\cdots)$$



  - 期待値・分散

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



例では…

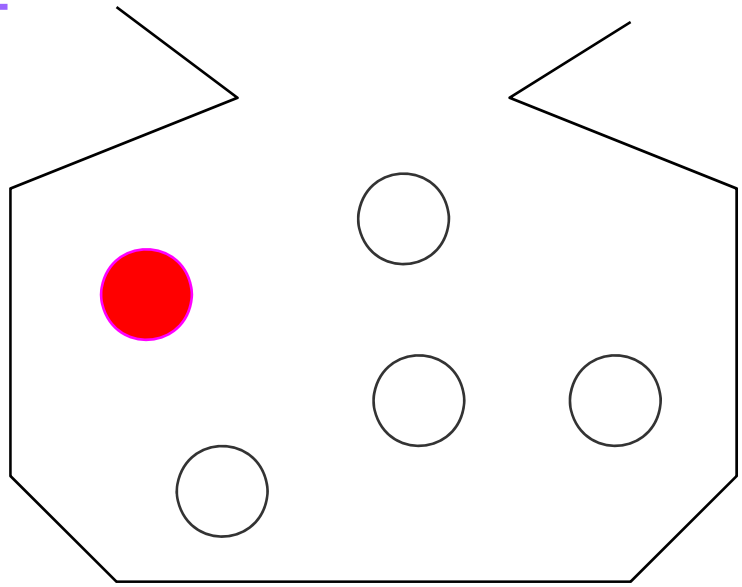
$$\begin{cases} E(X) = \dfrac{1}{1/5} = 5.00, \\ V(X) = \dfrac{(1-1/5)}{(1/5)^2} = 20.00 \end{cases}$$

離散型分布 discrete distribution

- 幾何分布の例
  - 例1：災害の到来
    - ある1年に風水害が起こる確率が1/25であるとする．風水害が起こるのは平均何年に1回か?
    - 上記と同じ災害が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：ドアを開けられる鍵を見つけよう！
    - n個の鍵束を持っている．かぎ束からひとつ鍵を取り出しドアを開けるとき，何回目で開くか? ただし，試した鍵は1回毎に鍵束に戻すこととする

演習5

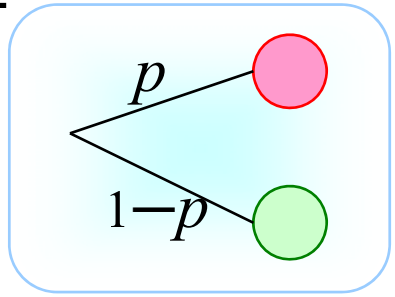
- 幾何分布を求めよう
  - 白玉4つ，赤玉1つが入っている袋がある．1つ取り出して元に戻すという試行を繰り返したとき，初めて赤玉が取り出される回の確率分布（幾何分布）を求めてみよう！　10回目に初めて赤玉が取り出される確率はどれだけか？

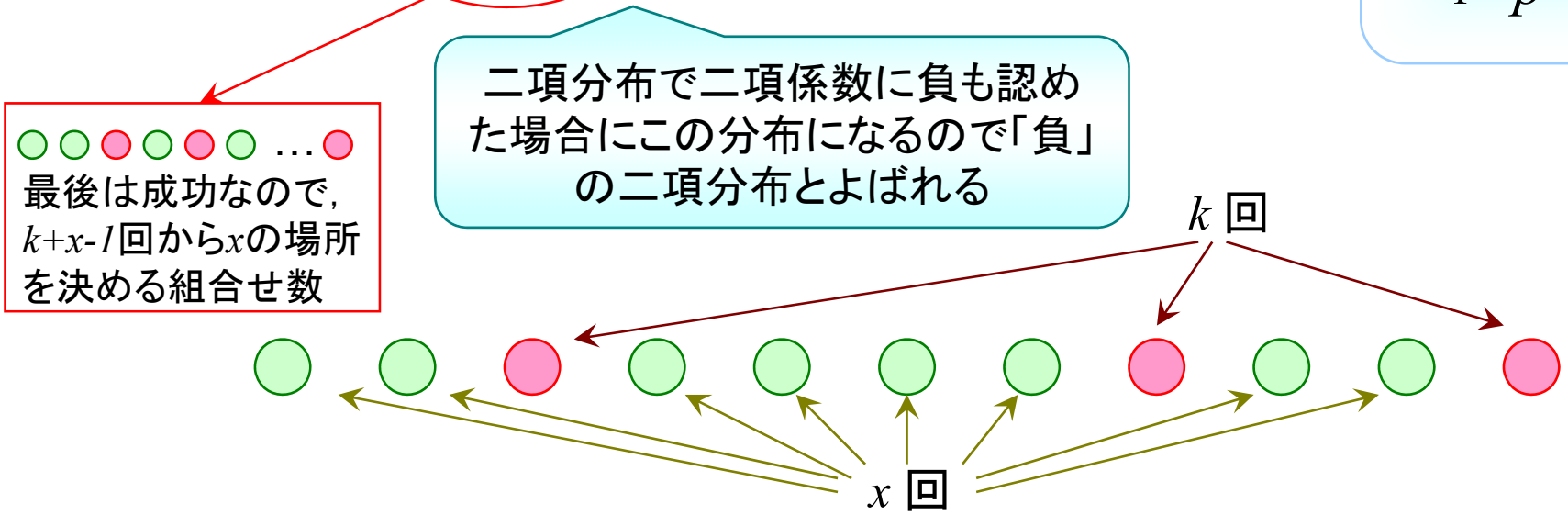


離散型分布 discrete distribution

- <u>負の二項分布 negative binomial distribution</u>
  - 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,…)の従う分布

$$f(x)={}_{k+x-1}C_x p^k (1-p)^x \quad (x=0,1,2,\cdots)$$



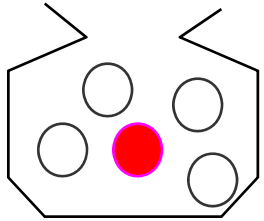


  - <u>k=1のときは幾何分布に等しい</u>ため幾何分布の一般化となっている

離散型分布 discrete distribution

- 負の二項分布
  - 確率分布

$$f(x) = {}_{k+x-1}C_x p^k (1-p)^x, \ (x = 0,1,2,\cdots)$$



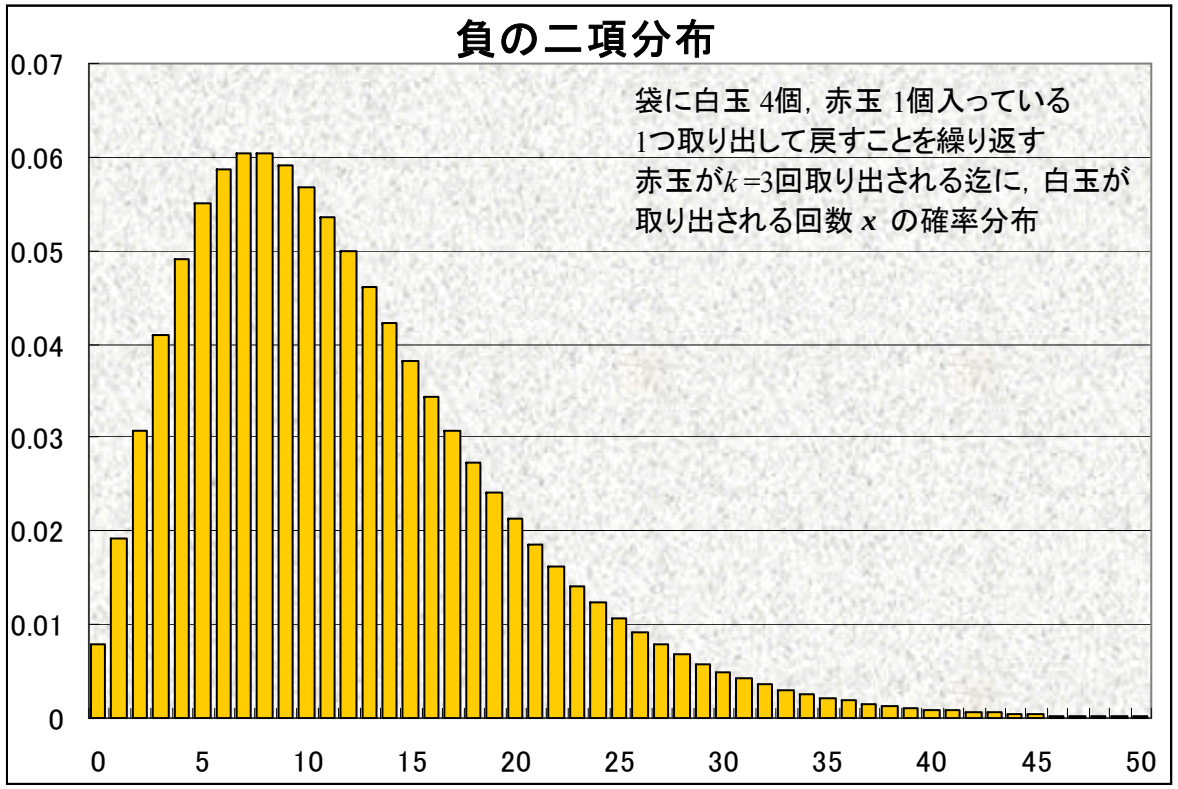
  - 期待値・分散

$$\begin{cases} E(X) = \dfrac{k(1-p)}{p}, \\ V(X) = \dfrac{k(1-p)}{p^2} \end{cases}$$

幾何分布のk倍

例では…

$$\begin{cases} E(X) = \dfrac{3(1-1/5)}{1/5} = 12.00, \\ V(X) = \dfrac{3(1-1/5)}{(1/5)^2} = 60.00 \end{cases}$$

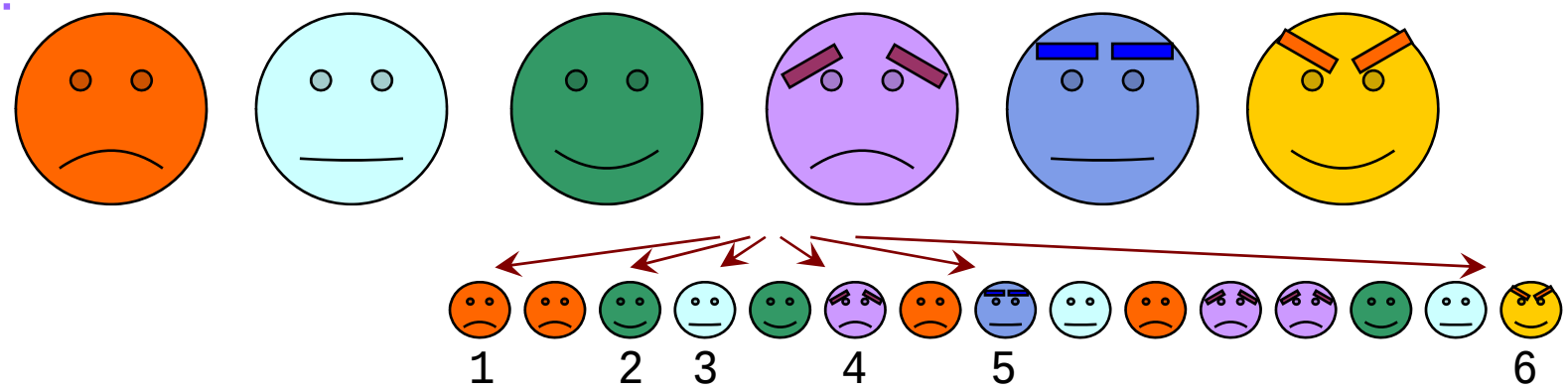


離散型分布 discrete distribution

- 負の二項分布の例
  - 例1：災害の到来
    - ある1年に風水害が起こる確率が1/25であるとする．風水害が起こるのは平均何年に1回か?
    - 上記と同じ災害が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繰り返したとき，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：シリーズものコレクター
    - 12種類のキャラクターが売られている．ただし，箱を開けるまで中にどれが入っているかはわからない．あるコレクターが全てのキャラクターを集めるためには何個買わねばならないか?

演習6

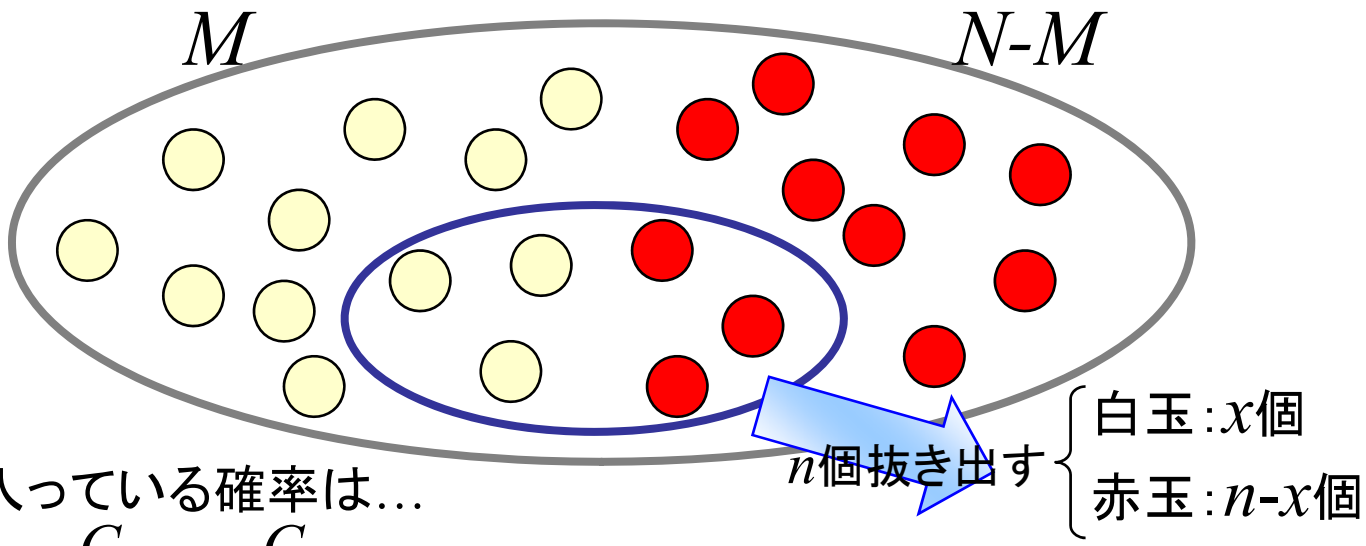
- 負の二項分布を求めよう
  - お菓子の付録に6種類のオマケがある．このオマケはそれぞれの箱に全種類がランダムに等確率に入っているとする．箱を開けるまで中身はわからない．黄色のオマケを3個そろえたい．そのために，お菓子を平均何個買わねばならないか？



注）全てのオマケを（少なくとも一つずつ）集める場合は，「クーポン収集問題」と呼ばれる問題になり，負の二項分布では計算できない．

離散型分布 discrete distribution

- <u>超幾何分布 hypergeometric distribution</u>
  - 例：白玉がM個，赤玉がN-M個（全部で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…

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

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

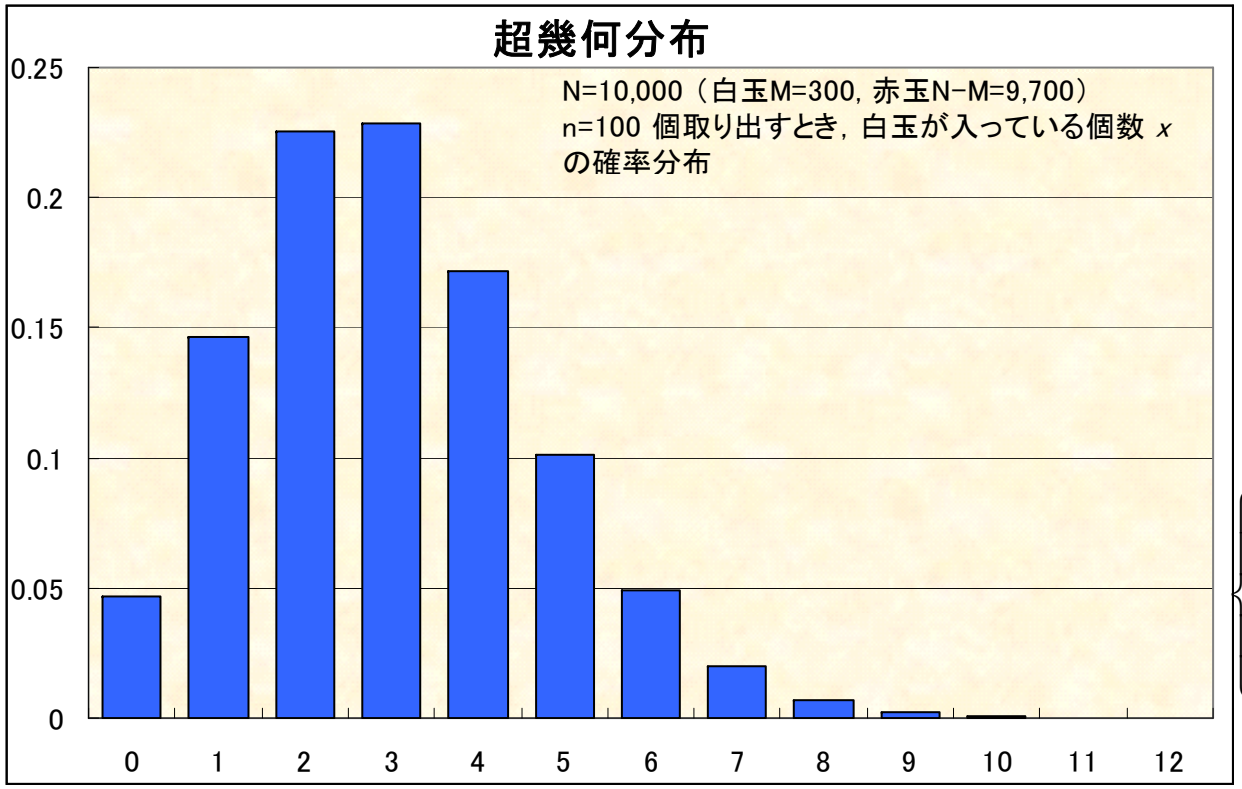
ただし，xの取り得る範囲は

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

離散型分布 discrete distribution

- 超幾何分布
  - 確率分布

$$f(x) = \frac{{}_M C_x \cdot {}_{N-M} C_{n-x}}{{}_N C_n} \qquad (x = \max\{0, n-(N-M)\}, \cdots, \min\{n, M\})$$



- 期待値・分散

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

例では…

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

離散型分布 discrete distribution

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

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

二項分布の平均と分散

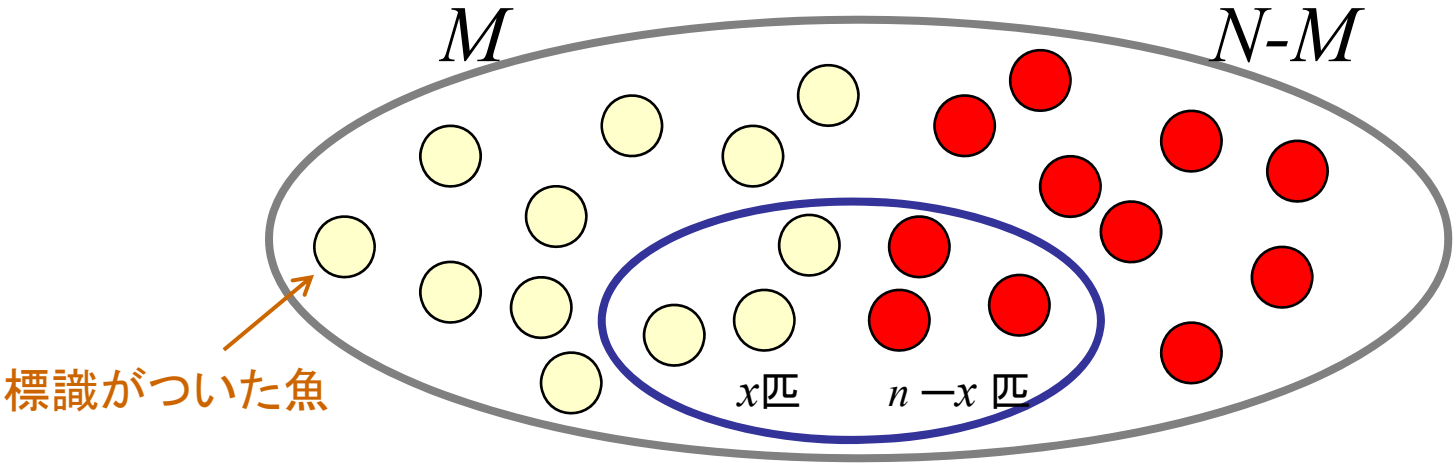
離散型分布 discrete distribution

標識再捕獲法
(mark-recapture method)
ともいう

- 超幾何分布の例
  - 例：資源調査「捕獲再捕獲法 capture-recapture method」
    - 湖の中の魚の個体数推定など

      湖に何匹の魚（N匹）がいるのか知りたいが，動くので難しい

      再捕獲により度数分布を書いてNを推定



- 例：ある湖の中に生息している対象魚について，200匹を捕獲し標識をつけた．さてしばらく後，湖から魚を10匹獲ったとき，標識がついている魚が2匹いた．この湖にはこの魚は何匹いると推定されるか?

Coffee Break!

Q1）1から100までの和は？

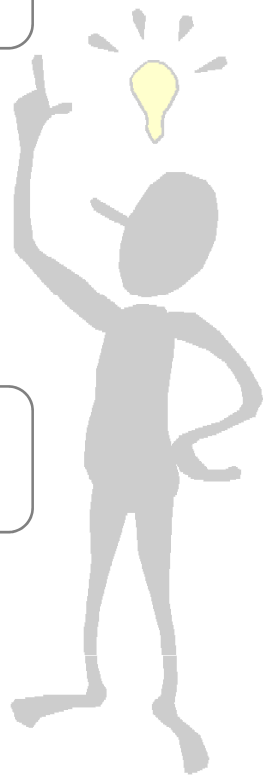
$$1 + 2 + 3 + \ldots + 100 = ?$$

Q2）1から100までの2乗和は？

$$1^2 + 2^2 + 3^2 + \ldots + 100^2 = ?$$

326から579までの和は？

42から283までの2乗和は？



1からnまでの和と2乗和は以下

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

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

$$\left[\begin{array}{l} \because) \, by \; (k+1)^3 - k^3 = 3k^2 + 3k + 1 \\ \sum_{k=1}^{n} \left\{ (k+1)^3 - k^3 \right\} = \sum_{k=1}^{n} \left\{ 3k^2 + 3k + 1 \right\} \end{array} \right]$$

確率密度関数

p. d. f. (probability density function)

連続(型)分布continuous distribution

★(連続)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

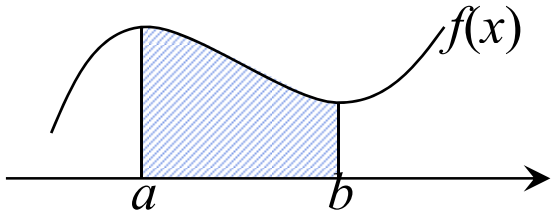
★ガンマ分布 Gamma distribution (χ^2分布, 指数分布)

★ベータ分布 Beta distribution

確率密度関数 p. d. f.

- 連続(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 は連続型の確率分布を持つという

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



ただし,

$$\begin{cases} f(x) \ge 0 (-\infty \le x \le \infty), \\ \int_{-\infty}^{\infty} f(x)dx = 1 \end{cases}$$

確率密度関数
probability density function

$\longleftrightarrow$

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

$$\begin{cases} f(x_k) \ge 0 \ (k = 1, 2, \cdots), \\ \sum_{k=1}^{\infty} f(x_k) = 1 \end{cases}$$

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

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

確率密度関数 p. d. f.

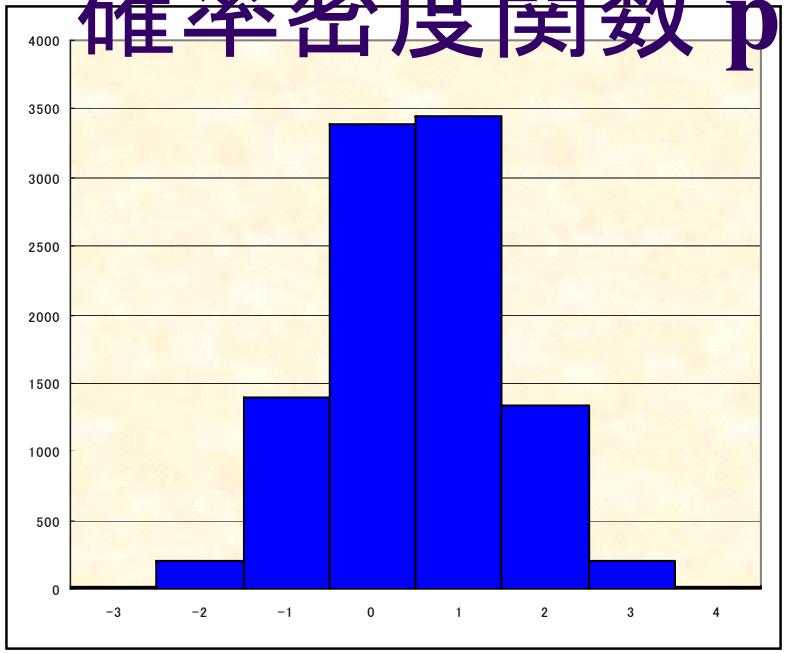
- 連続型確率変数の期待値と分散
  - 連続型確率変数 X の期待値

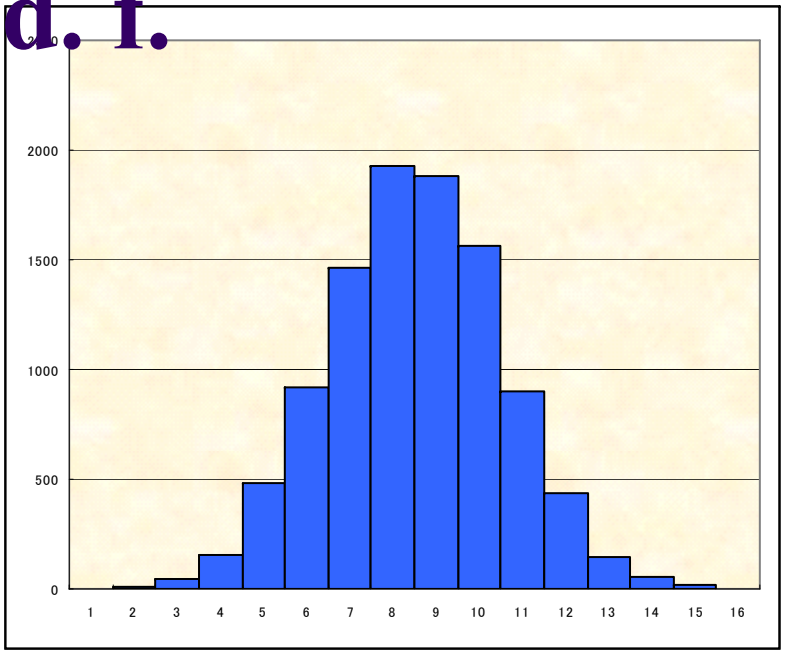
$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad \longleftrightarrow \quad \boxed{\sum_x x \cdot f(x)}$$

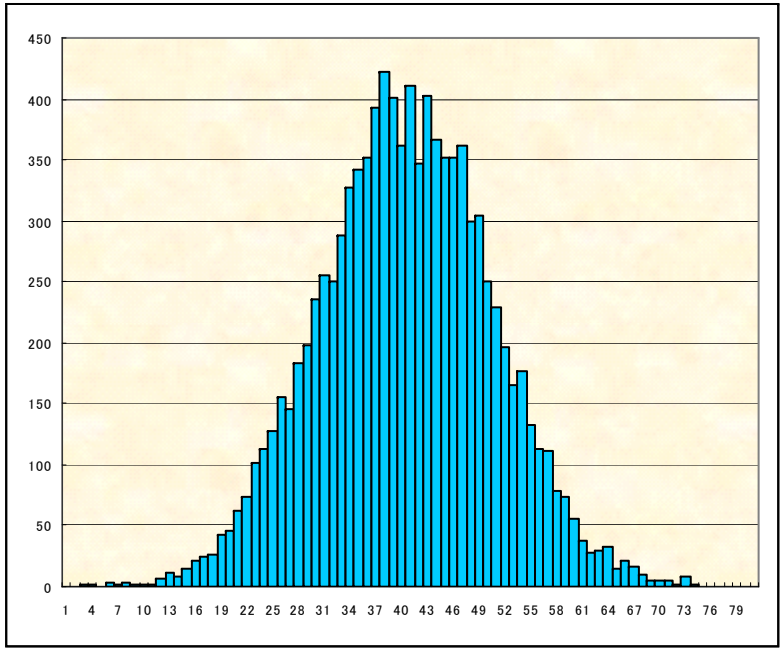
  - 連続型確率変数 X の分散

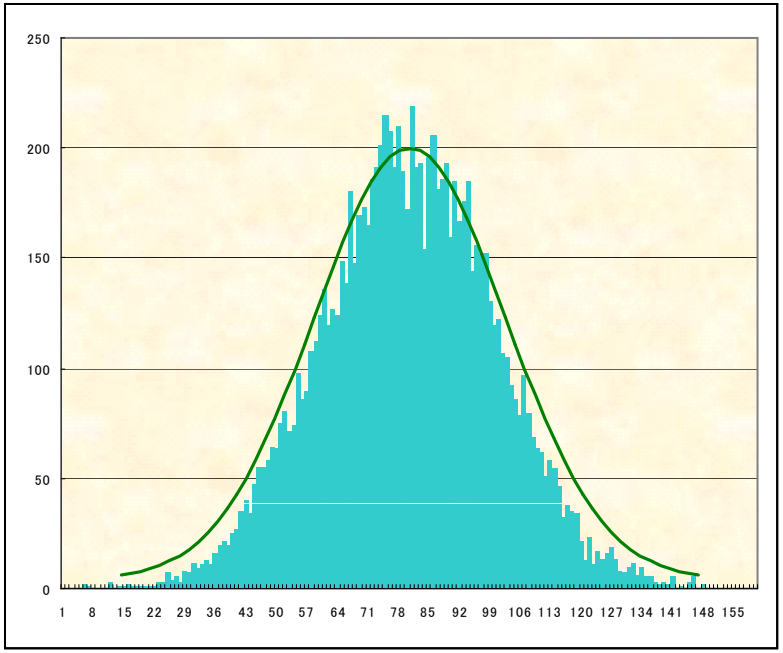
$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx \quad \longleftrightarrow \quad \boxed{\sum_x (x - E(X))^2 f(x)}$$

確率密度関数 p. d. f.


4000
3500
3000
2500
2000
1500
1000
500
0
−3 −2 −1 0 1 2 3 4


2000
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

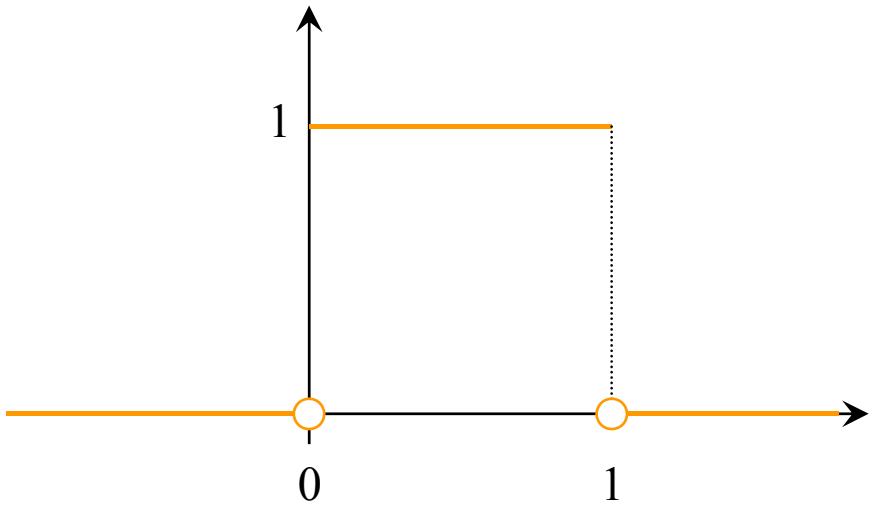

450
400
350
300
250
200
150
100
50
0
1 4 7 10 13 16 19 22 25 28 31 34 37 40 43 46 49 52 55 58 61 64 67 70 73 76 79


250
200
150
100
50
0
1 8 15 22 29 36 43 50 57 64 71 78 85 92 99 106 113 120 127 134 141 148 155

連続型分布 continuous distribution

- （連続）<u>一様分布 uniform distribution</u>
  - 確率密度関数

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



  - 累積分布関数

$$F(x)=\begin{cases}0 & (x<0)\\ x & (0\le x\le 1)\\ 0 & (1<x)\end{cases}$$

  - 期待値・分散

$$\begin{cases} E(X)=\int_{-\infty}^{\infty} x\cdot f(x)dx=\int_0^1 x\cdot 1dx=\left[\frac{x^2}{2}\right]_0^1=\left\{\frac{1^2}{2}-\frac{0^2}{2}\right\}=\frac{1}{2} \\ V(X)=\int_{-\infty}^{\infty}(x-E(X))^2\cdot f(x)dx=\int_0^1(x-\frac{1}{2})^2\cdot 1dx=\left[\frac{(x-1/2)^3}{3}\right]_0^1 \\ \quad=\left\{\frac{(1-1/2)^3}{3}-\frac{(0-1/2)^3}{3}\right\}=\frac{1/8}{3}-\frac{-1/8}{3}=\frac{1}{12} \end{cases}$$

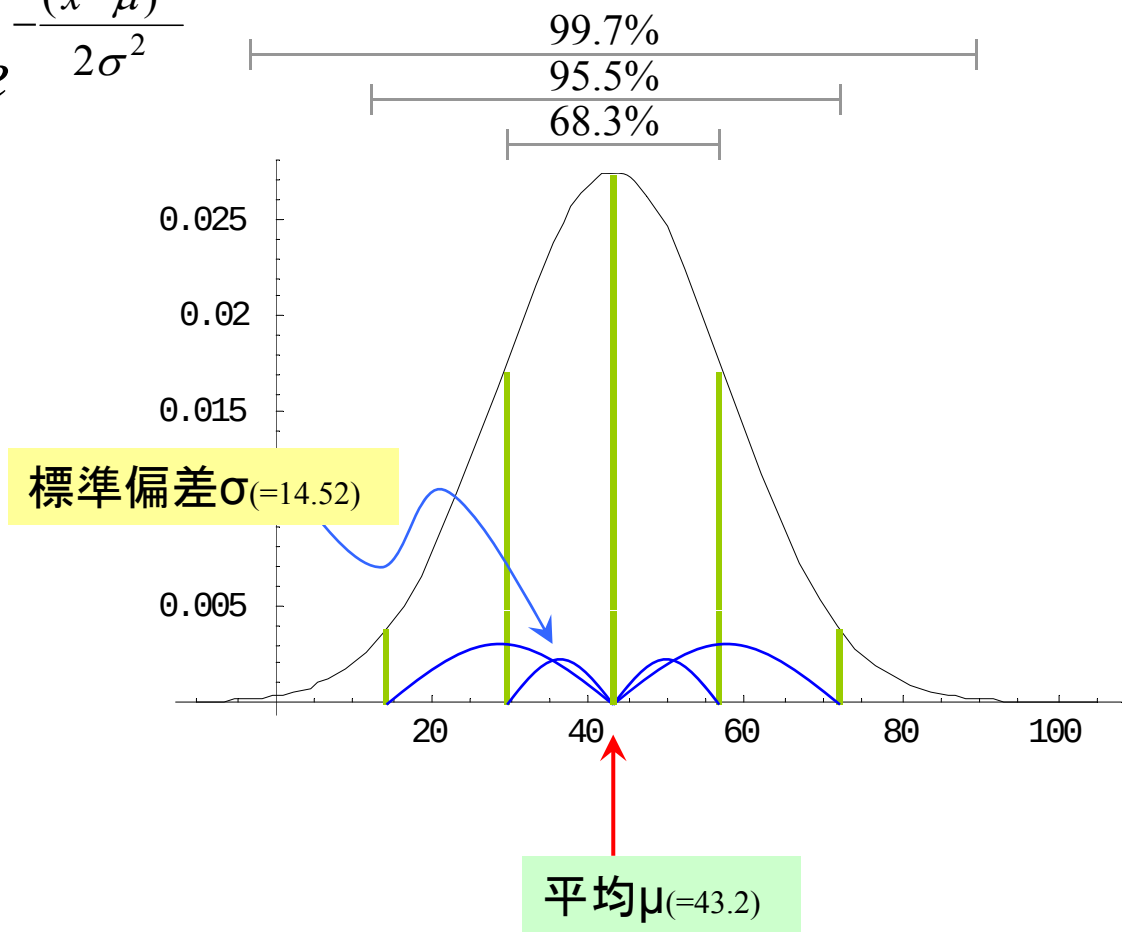
連続型分布 continuous distribution

- <u>正規分布 normal distribution</u>
  - 確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

平均μ, 分散σ^2

$$\boxed{N(\mu, \sigma^2)}$$



連続型分布 continuous distribution

- 標準正規分布 standard normal distribution
  - 確率変数の標準化
    - 平均μ，分散σ^2の正規分布に従う確率変数Xについて

$$X \to Z = \frac{X-\mu}{\sigma}$$

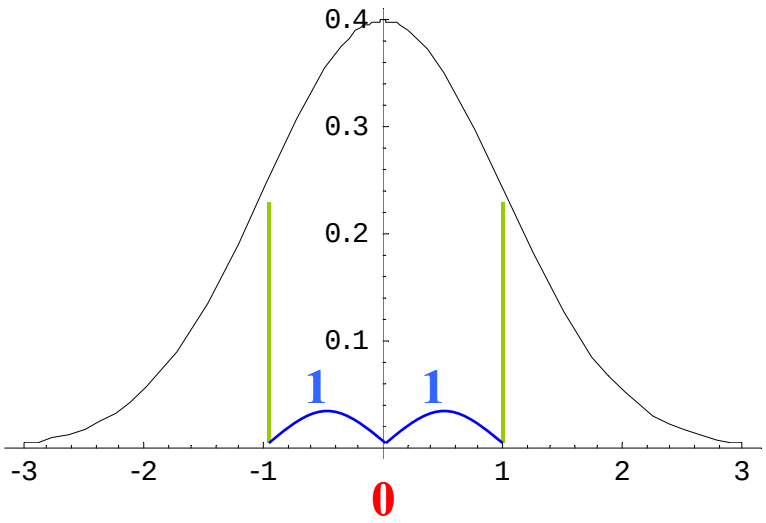
標準正規分布

確率変数Zは，平均0，分散1の正規分布に従う

  - 確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$N(0,1)$$



連続型分布 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近づく

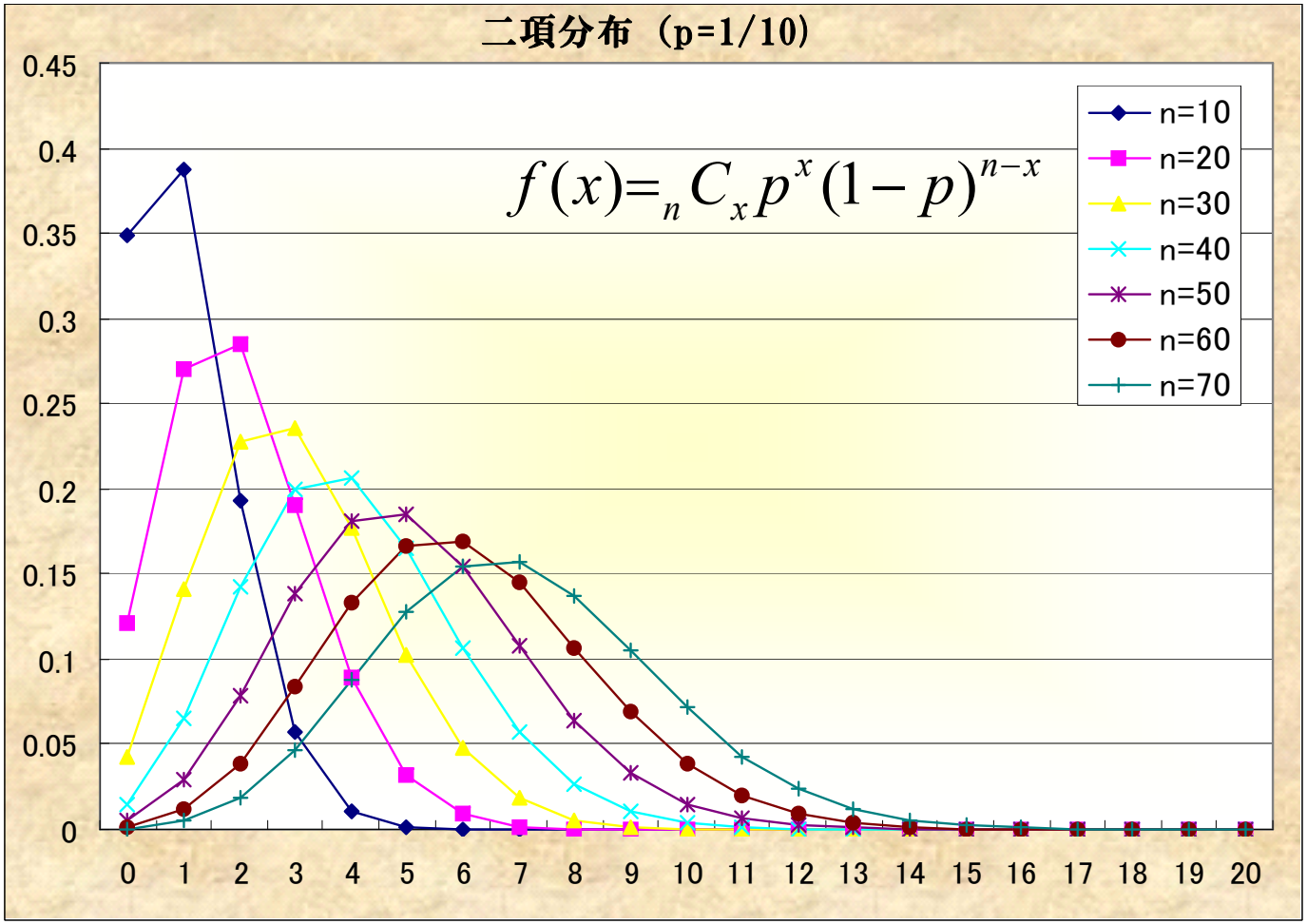
$$\boxed{Bi(n,p)} \xrightarrow{n \to \infty} \boxed{N(\mu,\sigma^2)} \quad \begin{cases} \mu = np \\ \sigma^2 = np(1-p) \end{cases}$$

  - 試行回数nが一定の時に，確率pを0.5に近づけると，二項分布は正規分布に近づく

$$\boxed{Bi(n,p)} \xrightarrow[p \to 0.5]{n = c,} \boxed{N(\mu,\sigma^2)}$$

連続型分布 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近づく



連続型分布 discrete distribution

- 正規分布による二項分布の近似
  - 例:内閣支持率
    - 500人の人に内閣を支持するかどうか聞いたところ，275人が指示すると答えた.

$$\text{内閣支持率:}\quad p=\frac{275}{500}=0.55$$

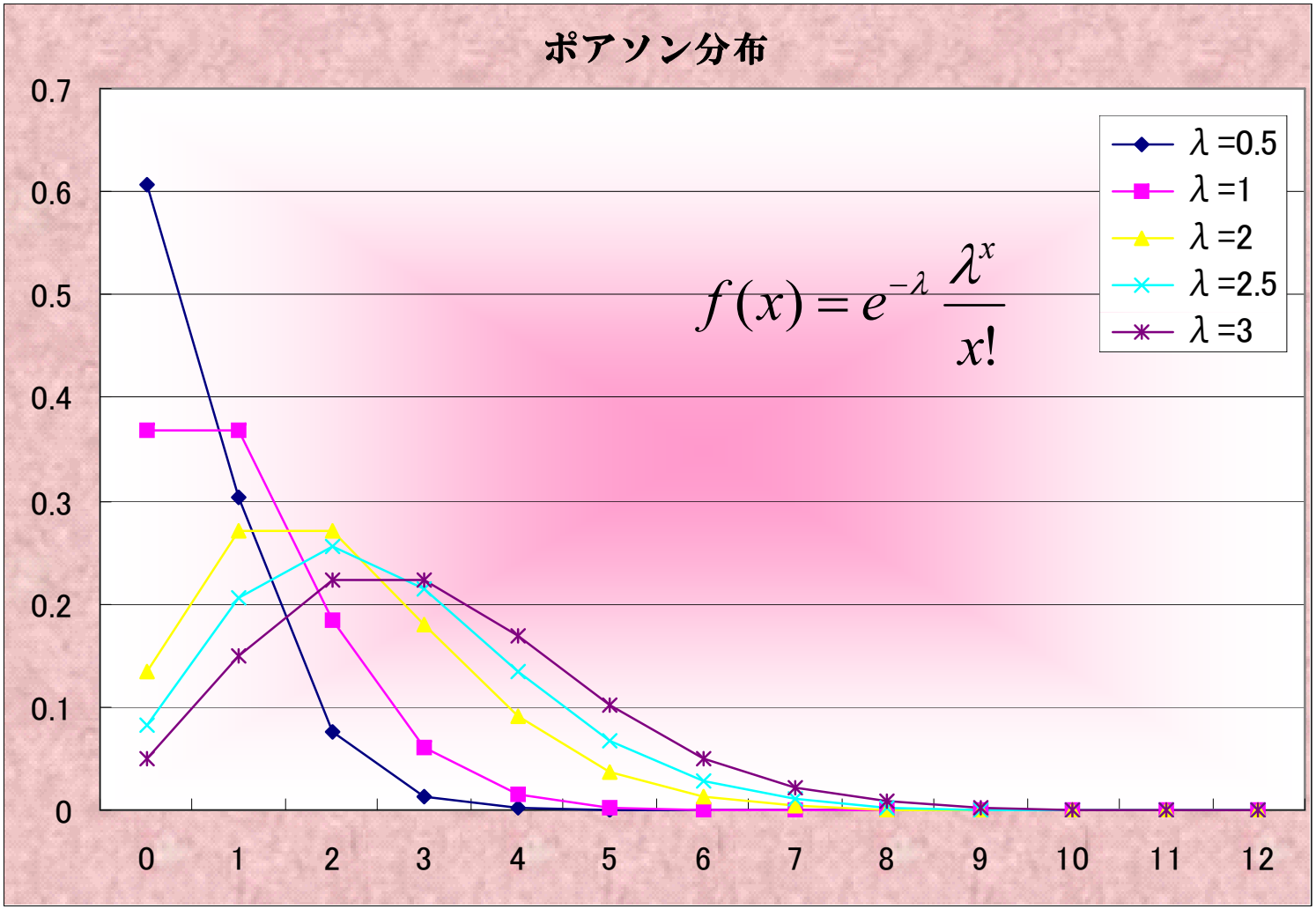
    - 内閣支持率を p (不支持率 q = 1-p) とすると，これは二項分布となる.
    - 点推定では内閣支持率は55%である．正規分布近似を考えると，$\bar{x}=np=500\times0.55=275$

$$\sigma=\sqrt{npq}=\sqrt{500\times0.55\times0.45}\cong\sqrt{124}\cong11$$

    - より，95%信頼区間における区間推定では，内閣支持率は $275\pm1.96\times11\Leftrightarrow253\leq x\leq297$ より50.6%〜59.4%

連続型分布 discrete distribution

- ポアソン分布から正規分布へ…



連続型分布 continuous distribution

- <u>指数分布 exponential distribution</u> $Ex(\lambda)$

  ポアソン分布に従って起きる事象の生起間隔を表現

  - 確率密度関数

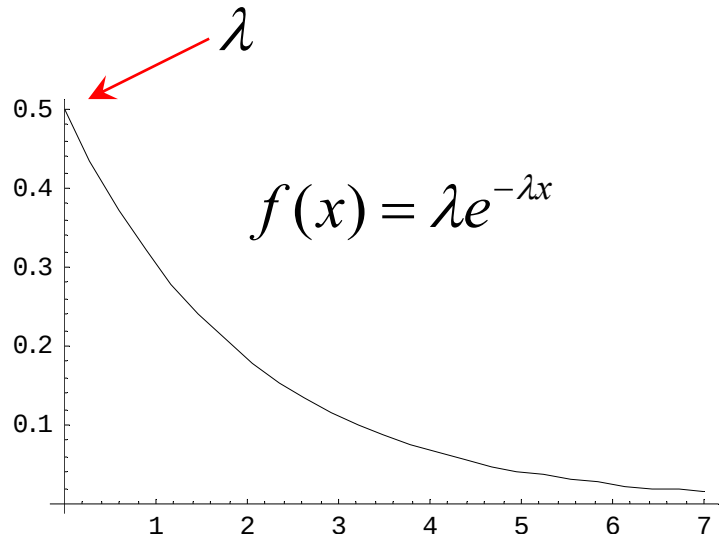
$$f(x)=\begin{cases}\lambda e^{-\lambda x} & (x\geq 0)\\ 0 & o.w.\end{cases}$$

  - 累積分布関数

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

  - 期待値・分散

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



連続型分布 continuous distribution

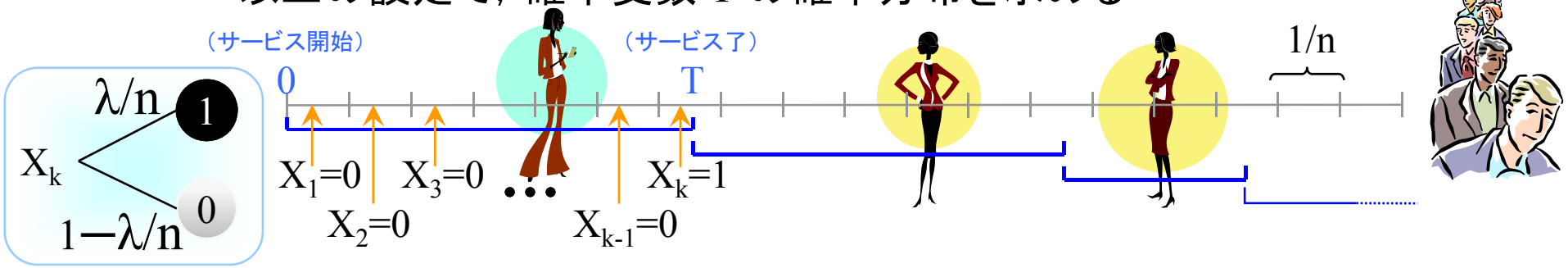
- 指数分布 exponential distribution
  - 例：サービスの待ち時間（チケット売場の列）

単位時間をn等分し，ある区間kで誰かが購入を終了したら $X_k=1$，そうでないなら $X_k=0$ とする確率変数 $X_1, X_2, X_3, \ldots$ を考える．

チケット販売開始時点 = 0, 1人目の購入終了時点 = T とする．

$X_1, X_2, X_3, \ldots$ はパラメータ $p = \lambda/n$ のベルヌーイ試行に従うとする．

→ 以上の設定で，確率変数 T の確率分布を求める



区間N+1で購入終了 → $T \fallingdotseq (1/n)N$ （nが十分大きい時）

任意の $t>0$ に対し，kを，$k \leqq nt < k+1$ を満たす整数とする

N+1は幾何分布に従う

$$\rightarrow P(0 \le T \le t) \approx P(0 \le N \le nt) = \sum_{i=0}^{k} (1-\lambda/n)^i \lambda/n$$

累積分布関数

確率密度関数

$$= 1-(1-\lambda/n)^{k+1} \approx 1-(1-\lambda/n)^{nt} \xrightarrow[n\to\infty]{} 1-e^{-\lambda t} = \int_0^t \lambda e^{-\lambda x} dx$$

連続型分布 continuous distribution

- ガンマ分布 Gamma distribution　$Ga(\alpha,\lambda)$

  - 確率密度関数

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

  - ガンマ関数 Gamma function

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

$$\Gamma(n)=(n-1)!\qquad (n\in \boldsymbol{N})$$

nが自然数の時

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

連続型分布 continuous distribution

- <u>ベータ分布 Beta distribution</u>

$Be(\alpha, \beta)$

  - 確率密度関数

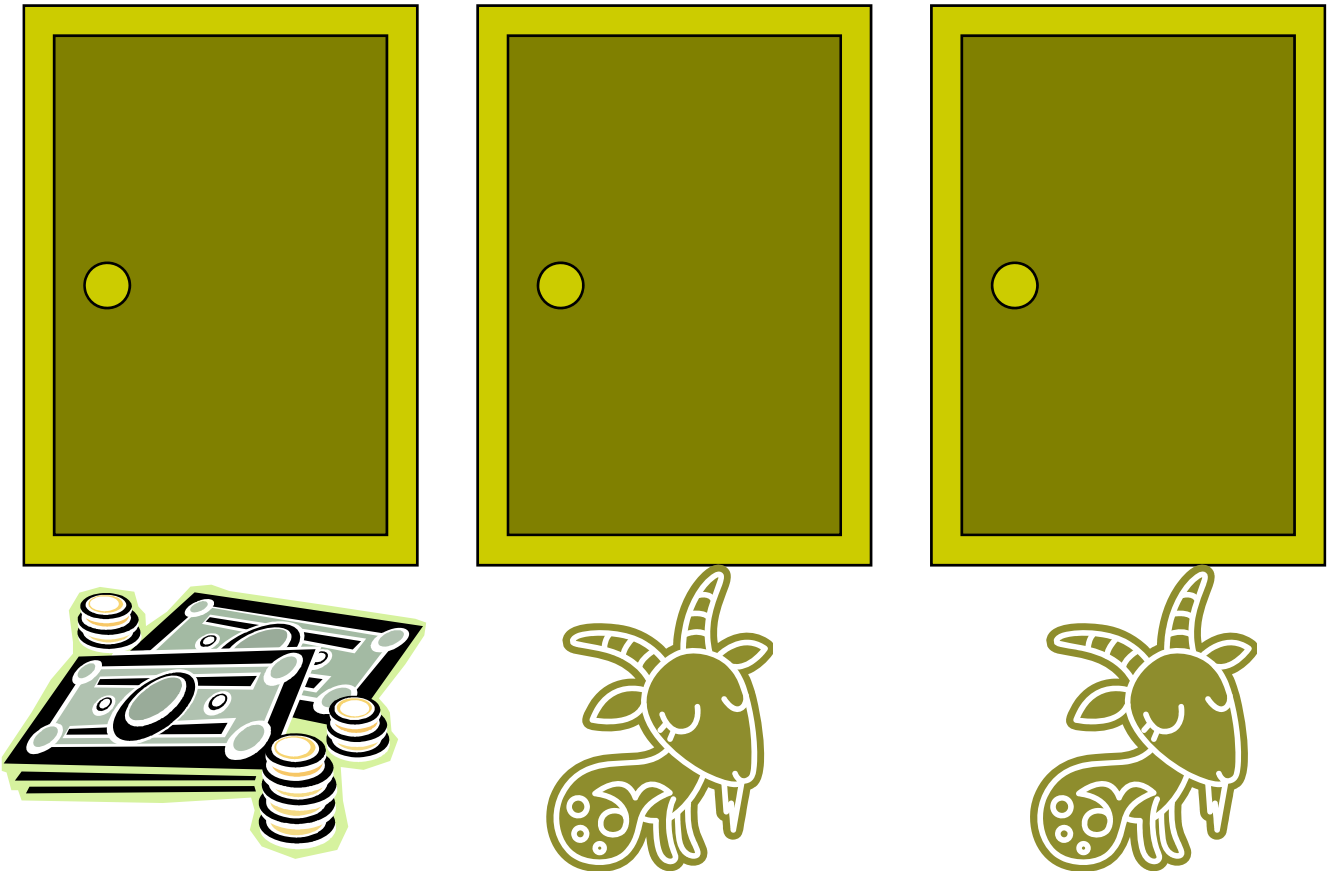
$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \quad (0 < x < 1)$$

  - 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
確率的直感



3枚の扉の向こうに
- 百万ドル（当たり）
- 山羊（はずれ）
- 山羊（はずれ）

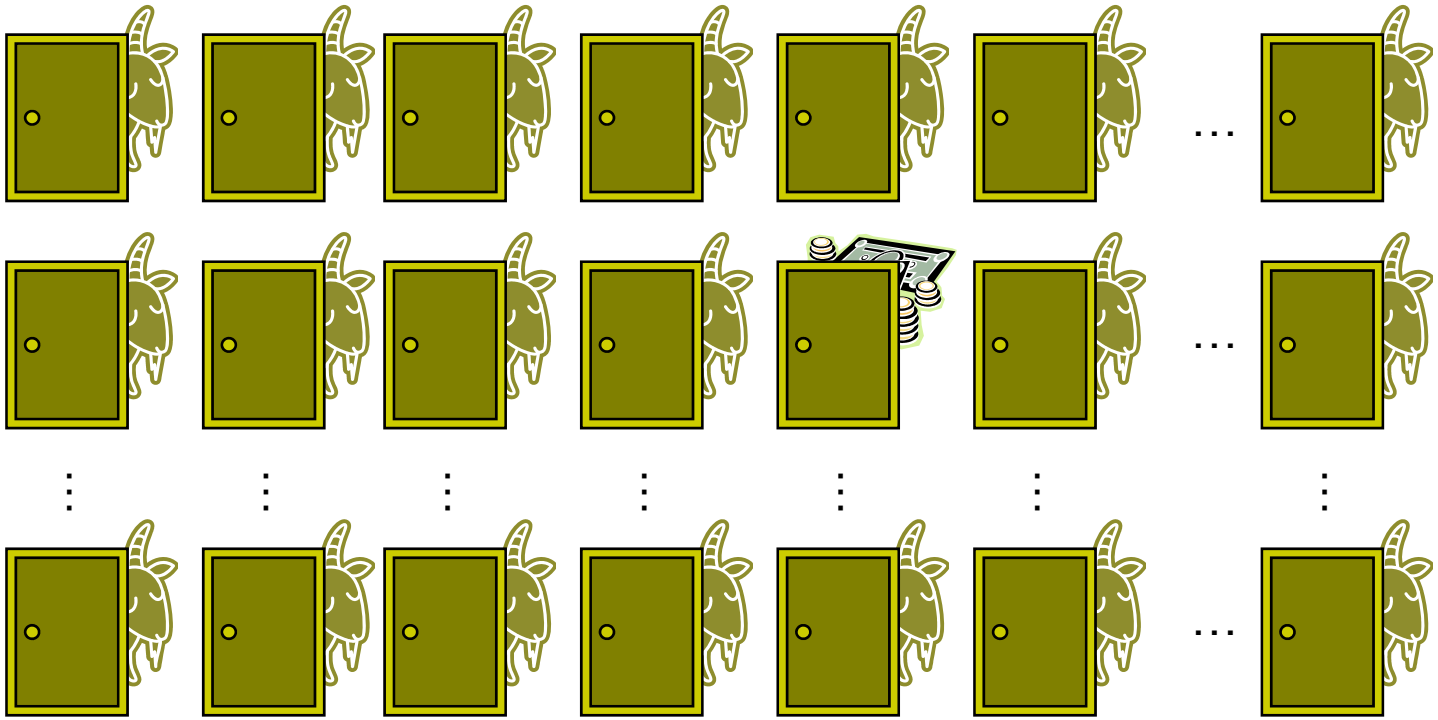
が隠されているよ．あなたは扉を1つだけ選んでいいのよ．

ところで，あなたが選ばなかった2つの扉のうち，山羊の扉を開くから，それを見た後で，開く扉を変えてもいいよ．

さぁ，どうする？

Coffee Break!

Monty-Hole Dilemma



どうしても納得いかない人のため，扉の数を増やしてみましょう！

最初に選ぶ扉が100万もあったらどうかしら？

100万の扉からあなたが1つを選んだ後で，残り99万9999の扉のうち山羊（はずれ）の99万9998の扉を開いて見せます．

それでもあなたは，最初の選択を変えない？　あなたの最初の選択は神懸かり的な幸運に恵まれているのかしら？

参考文献

- 東京大学教養学部統計学教室編 「統計学入門」東京大学出版会（1991）
- 村上雅人「なるほど統計学」海鳴社（2002）
- G.Blom, L.Holst, D.Sqndell / 森真 訳「確率論へようこそ」シュプリンガー・フェアラーク東京（1995,2005[新装版]）
- 丹慶勝市「図解雑学 統計解析」ナツメ社（2003）

- 東京大学教養学部統計学教室編 「自然科学の統計学」東京大学出版会（1992）
- 白石修二「例題で学ぶ Excel統計入門」森北出版（2001）
- J.Matousek, J.Nesetril / 根上生也・中本敦浩 訳「離散数学への招待 上」シュプリンガー・フェアラーク東京（2002）
- 徳山豪「工学基礎 離散数学とその応用」数理工学社（2003）
- B.Schechter / グラベルロード訳「My Brain is Open」共立出版（2003）