

九州大学 工学部地球環境工学科 船舶海洋システム工学コース

海事統計学 第8回 (担当:木村)

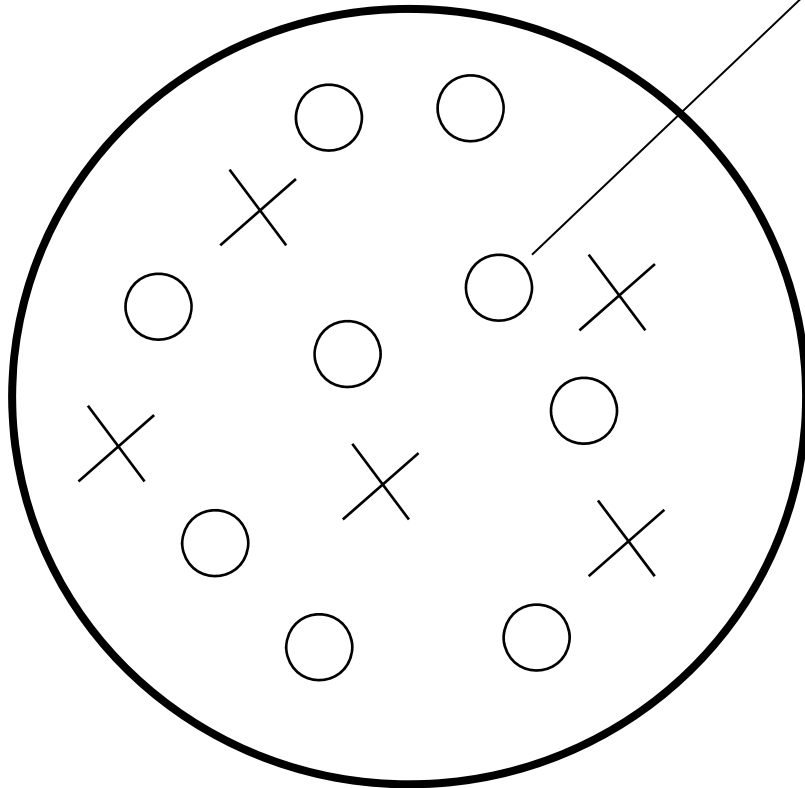
標本分布1

授業の資料および演習・例題等の解答は

<http://sysplan.nams.kyushu-u.ac.jp/gen/index.html>

母集団と標本

調べる対象となる全体の集団：
母集団 (population)



x_1

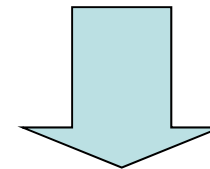
・母集団から無作為に抽出されたデータ: **標本** (sample)

・標本は、抽出する度に異なった値をとる → 「**標本**」= **確率変数**

・標本の確率分布 =



・母集団分布を特徴付けるパラメータ:
(例: 平均・分散・相関係数など)
母数 (population parameter)

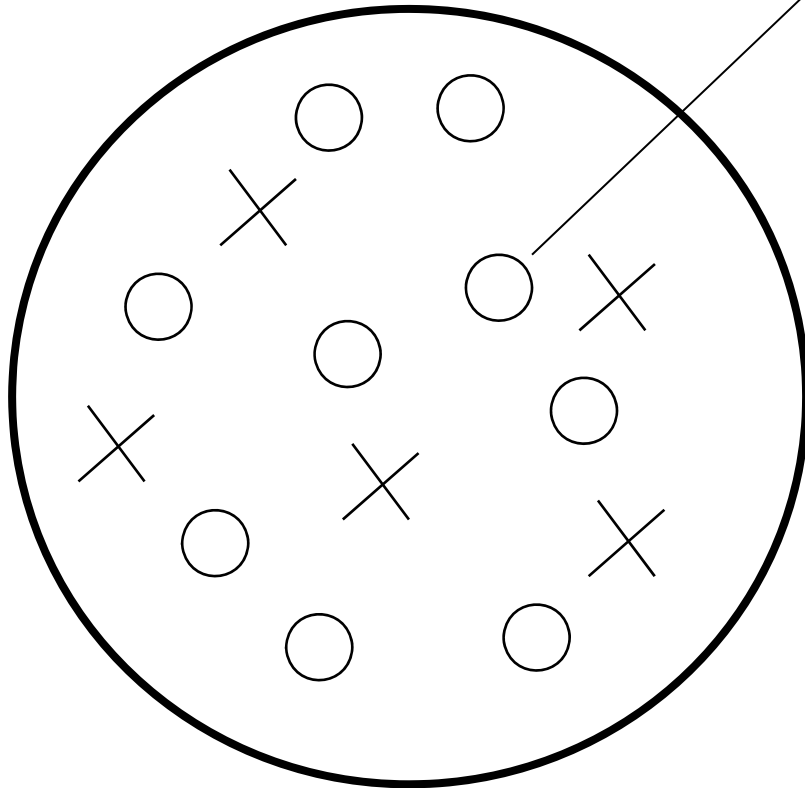


母数は、一般に未知
多くの場面において、母数を知る必要有

母集団全体を調べれば良いが...

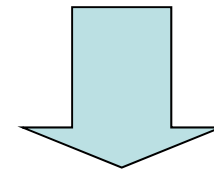
母集団と標本

調べる対象となる全体の集団：
母集団 (population)



x_1

- ・母集団から無作為に抽出されたデータ: **標本** (sample)
- ・標本は、抽出する度に異なった値をとる → 「**標本**」=確率変数
- ・標本の確率分布 = **母集団分布**
- ・母集団分布を特徴付けるパラメータ:
(例: 平均・分散・相関係数など)
母数 (population parameter)

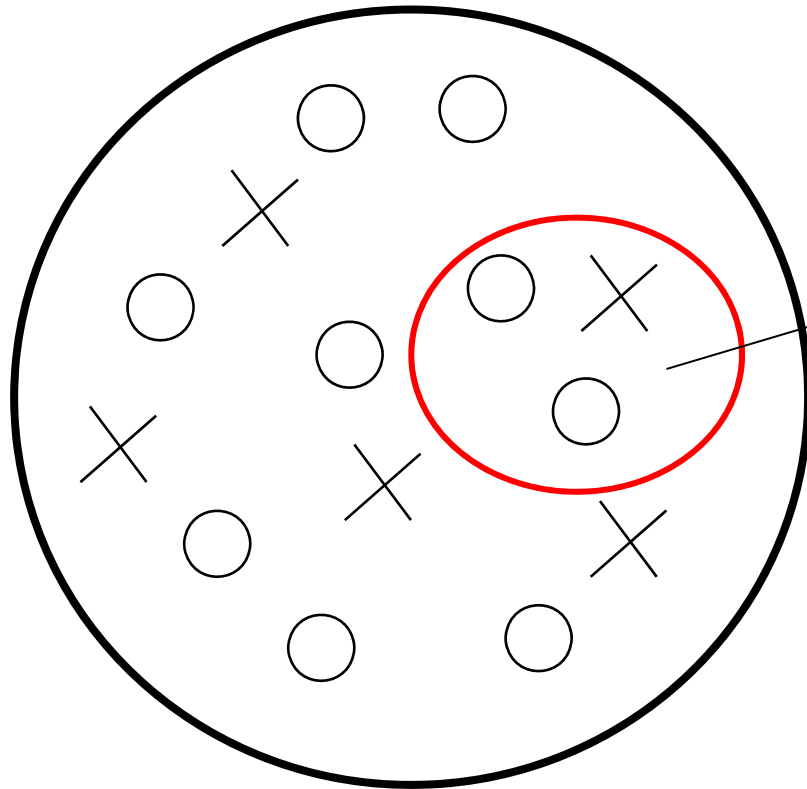


母数は、一般に未知
多くの場面において、母数を知る必要有

母集団全体を調べれば良いが...

母集団と標本

調べる対象となる全体の集団：
母集団 (population)



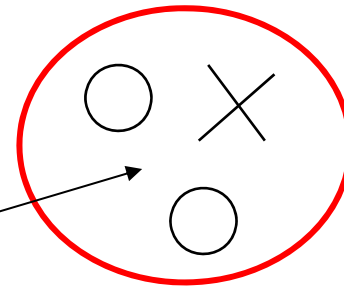
母集団のとある性質を示す
パラメータ：母数 p

一般に、**母集団全体を調べることは不可能**

(例) 強度や寿命テスト、

これから発生するであろう未来の事象など

母集団の中から選ばれる一部分
の集まり：**標本** (sample)



x_1, x_2, \dots, x_n

標本について調べる
ことで、母数を推測
→ 統計的推論 \hat{p}

標本から計算されるパラメータ：(標本)統計量

標本平均

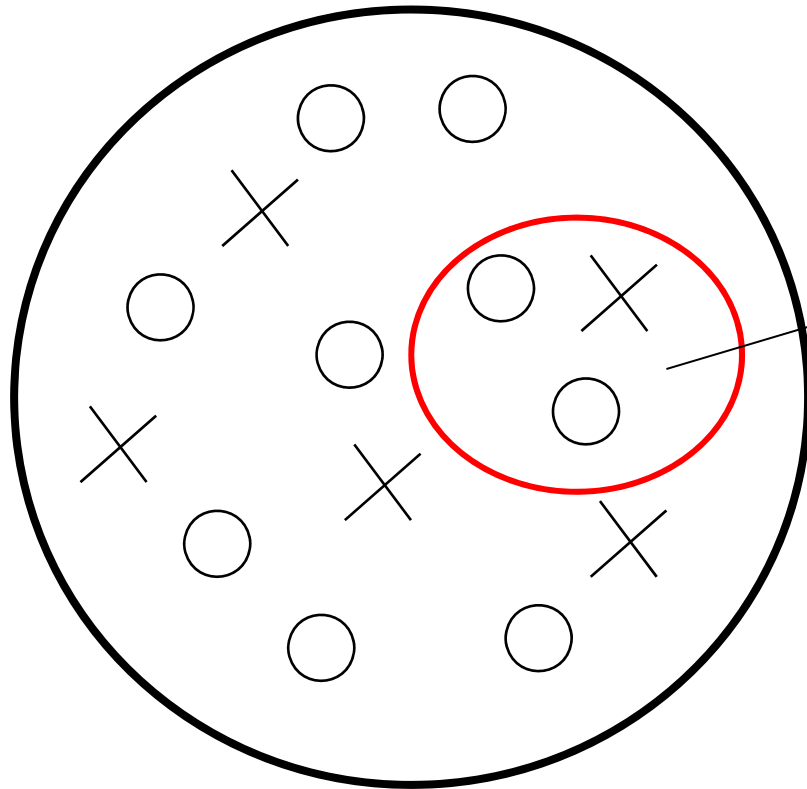
標本分散

不偏分散

統計量は、標本毎に異なる値をとる確率変数
この統計量の分布を標本分布という

母集団と標本

調べる対象となる全体の集団：
母集団 (population)



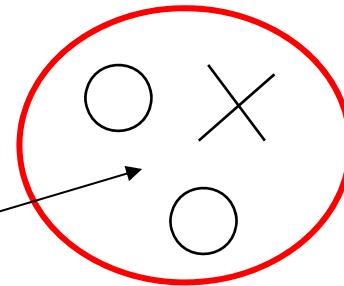
母集団のとある性質を示す
パラメータ：母数 p

一般に、母集団全体を調べることは不可能

(例) 強度や寿命テスト、

これから発生するであろう未来の事象など

母集団の中から選ばれる一部分
の集まり：**標本** (sample)



x_1, x_2, \dots, x_n

標本について調べる
ことで、母数を推測
→ 統計的推論 \hat{p}

標本から計算されるパラメータ：(標本)統計量

標本平均

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

標本分散

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

不偏分散

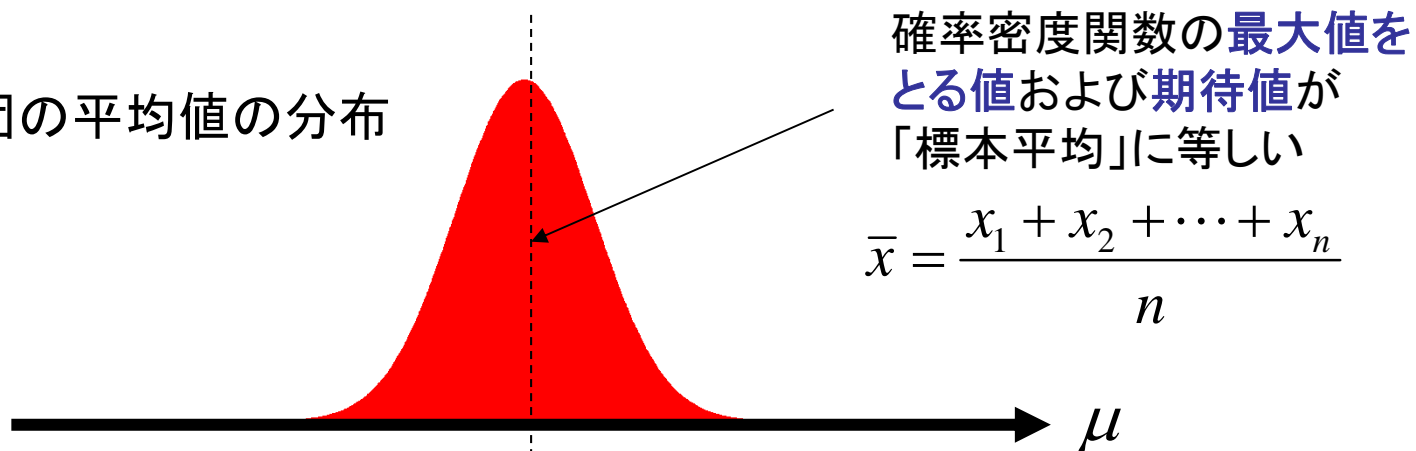
$$U^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

統計量は、標本毎に異なる値をとる確率変数
この統計量の分布を標本分布という

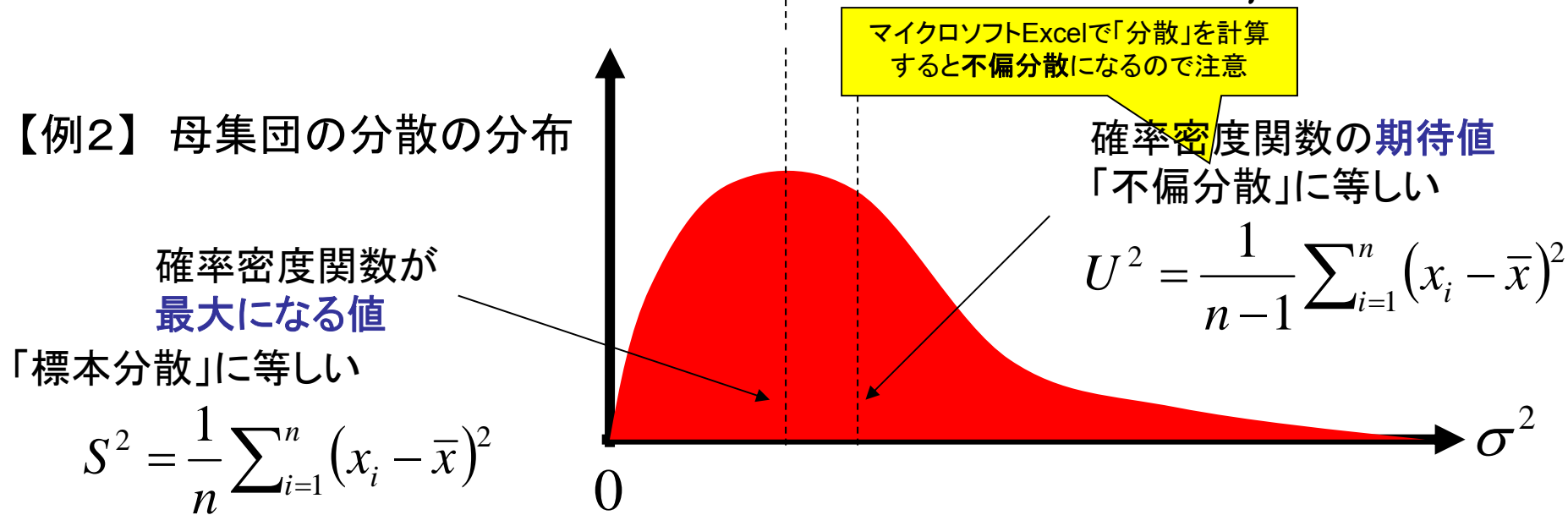
補足：「標本分散」と「不偏分散」の違い

講義後半で詳しく述べるが、サンプルが与えられると、母数パラメータの考えられうる値は「確率(密度)分布」として得られる

【例1】 母集団の平均値の分布



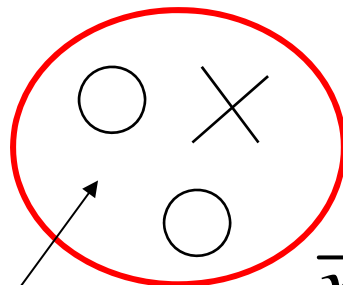
【例2】 母集団の分散の分布



母集団と標本

標本 (sample)

$$x_1, x_2, \dots, x_n$$



標本平均 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

\bar{X} を確率変数とした標本分布を考える

期待値 $E\{\bar{x}\} = E\left\{\frac{1}{n}(x_1 + x_2 + \dots + x_n)\right\}$

$$= \frac{1}{n}(E\{x_1\} + E\{x_2\} + \dots + E\{x_n\})$$

$$= \square$$

分散 $Var\{\bar{x}\} = Var\left\{\frac{1}{n}(x_1 + x_2 + \dots + x_n)\right\}$

$$= \frac{1}{n^2} Var\{x_1 + x_2 + \dots + x_n\}$$

$$= \frac{1}{n^2} (Var\{x_1\} + Var\{x_2\} + \dots + Var\{x_n\})$$

$$= \square$$

母集団 (population)

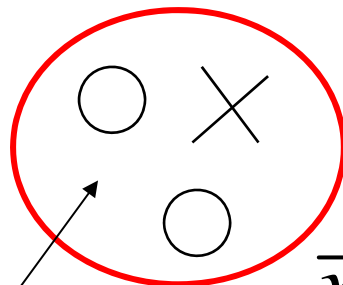
母集団の期待値 μ
 母集団の分散 σ^2
 (これらは母数で、
 一般に未知数)

平均値 μ 分散 σ^2 の母集団からとられた
 大きさ n の標本平均 \bar{x} の
 期待値は
 分散は

母集団と標本

標本 (sample)

$$x_1, x_2, \dots, x_n$$



標本平均 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$

\bar{X} を確率変数とした標本分布を考える

期待値 $E\{\bar{x}\} = E\left\{\frac{1}{n}(x_1 + x_2 + \dots + x_n)\right\}$

$$= \frac{1}{n}(E\{x_1\} + E\{x_2\} + \dots + E\{x_n\})$$

$$= \mu$$

分散 $Var\{\bar{x}\} = Var\left\{\frac{1}{n}(x_1 + x_2 + \dots + x_n)\right\}$

$$= \frac{1}{n^2} Var\{x_1 + x_2 + \dots + x_n\}$$

$$= \frac{1}{n^2} (Var\{x_1\} + Var\{x_2\} + \dots + Var\{x_n\})$$

$$= \frac{\sigma^2}{n}$$

サンプル数 n が増えれば
小さくなる

母集団 (population)

母集団の期待値 μ
母集団の分散 σ^2
(これらは母数で、
一般に未知数)

平均値 μ 分散 σ^2 の母集団からとられた
大きさ n の標本平均 \bar{x} の
期待値は μ
分散は $\frac{\sigma^2}{n}$

大数の法則

$x_1, x_2, \dots, x_n, \dots$ がそれぞれ確率変数で、

$$m_n = E\{x_1 + x_2 + \dots + x_n\} = E\{x_1\} + E\{x_2\} + \dots + E\{x_n\}$$

$$V_n = \text{Var}\{x_1 + x_2 + \dots + x_n\}$$

が存在し、 $\frac{V_n}{n^2} \rightarrow 0 \quad (n \rightarrow \infty)$ が成立するならば、

$$\frac{(x_1 + x_2 + \dots + x_n) - m_n}{n}$$

は0に確率収束する。

(大数の弱法則)

つまり、

サンプルを多数とれば、
得られる平均値は母平均(真の期待値)
に近づく

大数の法則

$x_1, x_2, \dots, x_n, \dots$ がそれぞれ確率変数で、

$$m_n = E\{x_1 + x_2 + \dots + x_n\} = E\{x_1\} + E\{x_2\} + \dots + E\{x_n\}$$

$$V_n = \text{Var}\{x_1 + x_2 + \dots + x_n\}$$

が存在し、 $\frac{V_n}{n^2} \rightarrow 0 \quad (n \rightarrow \infty)$ が成立するならば、

$$\frac{(x_1 + x_2 + \dots + x_n) - m_n}{n} \quad \text{は0に確率収束する。}$$

(大数の弱法則)

つまり、

サンプルを多数とれば、
得られる平均値は母平均(真の期待値)
に近づく

中心値極限定理

Central limit theorem

$x_1, x_2, \dots, x_n, \dots$ が互いに独立確率変数で、それらが全て

期待値 μ 分散 σ^2 の に従うとき、

平均値 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ の分布は n が大きくなると

正規分布 に近づく

すなわち

標準化

は標準正規分布 $N(0,1)$ に近づく

$x_1, x_2, \dots, x_n, \dots$ が互いに独立確率変数で、それらが全て

期待値 μ 分散 σ^2 の**正規分布**に従うとき、
 n の大きさにかかわらず上記の正規分布になる

中心値極限定理

Central limit theorem

$x_1, x_2, \dots, x_n, \dots$ が互いに独立確率変数で、それらが全て期待値 μ 分散 σ^2 の **任意の分布** に従うとき、

平均値 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ の分布は n が大きくなると

正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に近づく

すなわち

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

標準化

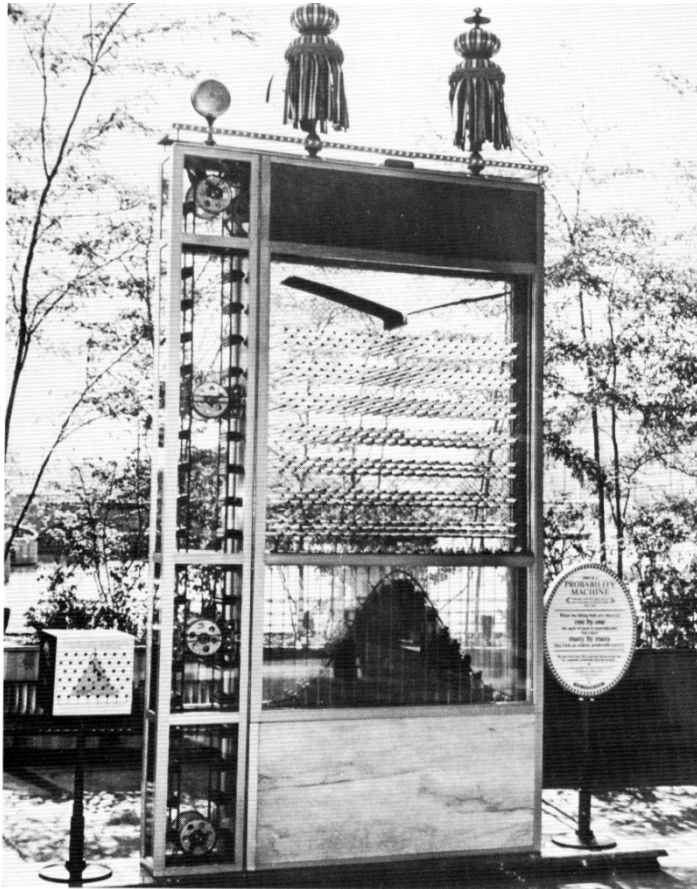
は標準正規分布 $N(0,1)$ に近づく

$x_1, x_2, \dots, x_n, \dots$ が互いに独立確率変数で、それらが全て

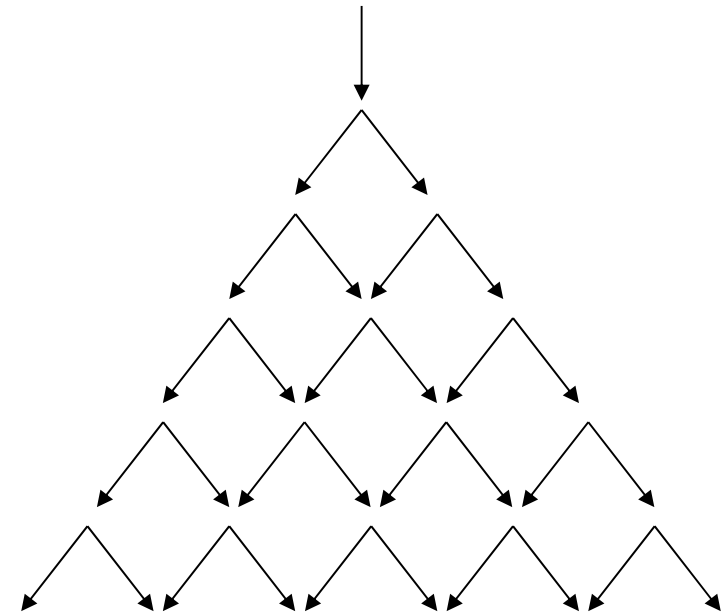
期待値 μ 分散 σ^2 の **正規分布** に従うとき、
 n の大きさにかかわらず上記の正規分布になる

正規分布による二項分布の近似

二項分布 $B(n, p)$ は n が大きくなるとき、
正規分布 $N(np, np(1-p))$ に近づく

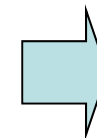


各段において、右または左へ
 $1/2$ の確率で落ちていく



どこに落ちるかは二項分布に従う

独立なベルヌーイ分布の確率変数の足し合わせ = 二項分布



正規分布へ

【復習】 二項分布

1回の試行で、ある事象の起こる確率が p
この試行を n 回行う

ベルヌイ試行
ベルヌイ分布

確率変数 x : 事象の起こった回数
確率分布 $P(x)$ → 二項分布 $B(n, p)$

(例) 3回サイコロを投げたときの1の目が出る回数

x	0	1	2	3
$P(x)$	$\left(\frac{5}{6}\right)^3 = \frac{125}{216}$	$3\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 = \frac{75}{216}$	$3\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) = \frac{15}{216}$	$\left(\frac{1}{6}\right)^3 = \frac{1}{216}$

一般に、確率 p を持つ事象が n 回の試行中 x 回起こることを考えると、

このときの確率は $p^x (1-p)^{n-x}$

$\underbrace{\text{○ ○ … ○}}_{x \text{ 回}} \quad \underbrace{\text{× × … ×}}_{n-x \text{ 回}}$

n 回の試行中に x 回起こるパターンの組合せ (combination) は、

$${}_n C_x = \frac{n!}{x!(n-x)!} \quad \text{よって} \quad P(x) = {}_n C_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

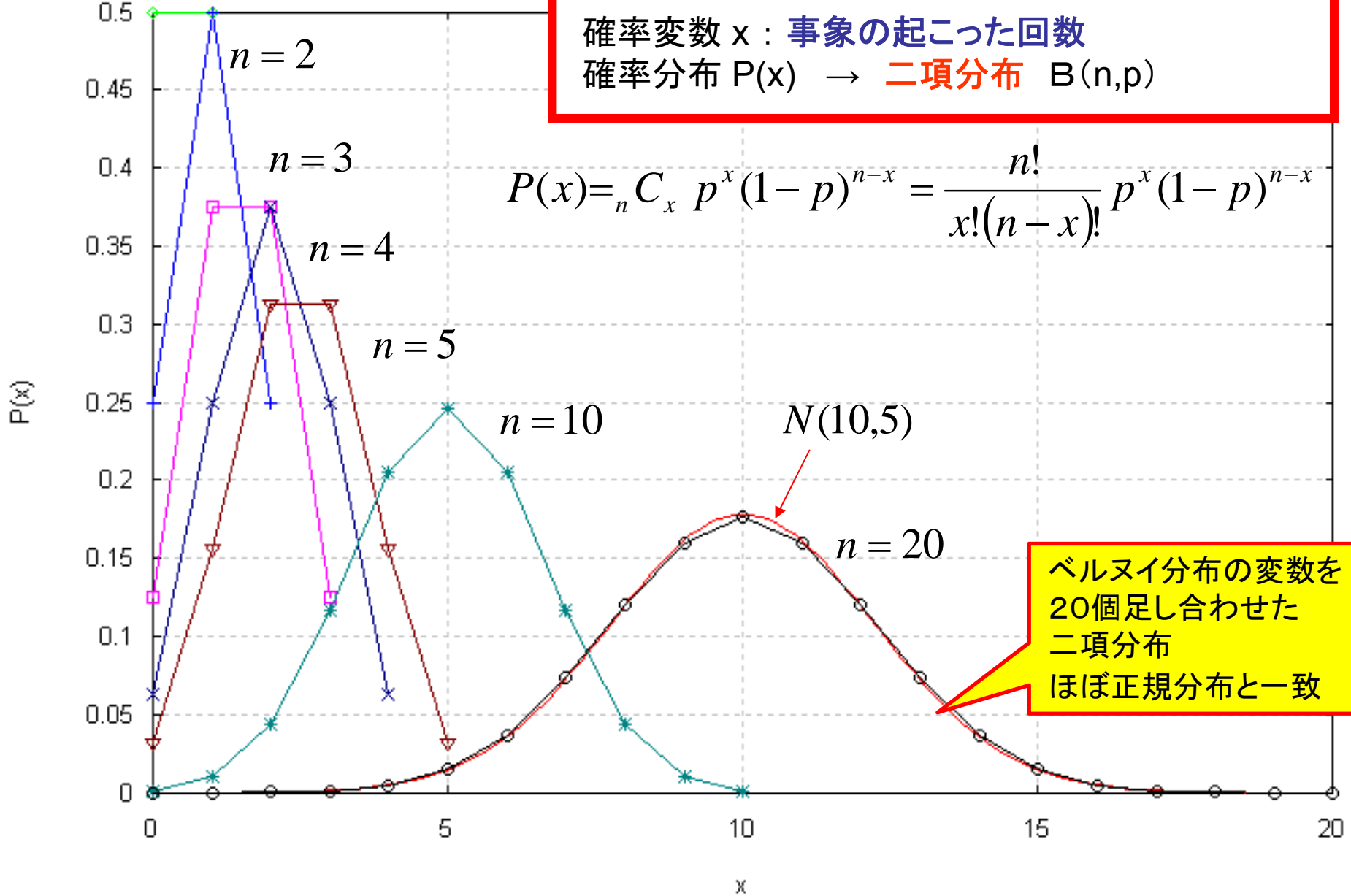
【復習】 二項分布の極限

$n = 1$ $p = 0.5$

1回の試行で、ある事象の起こる確率が p
 この試行を n 回足し合わせる \searrow ベルヌイ試行

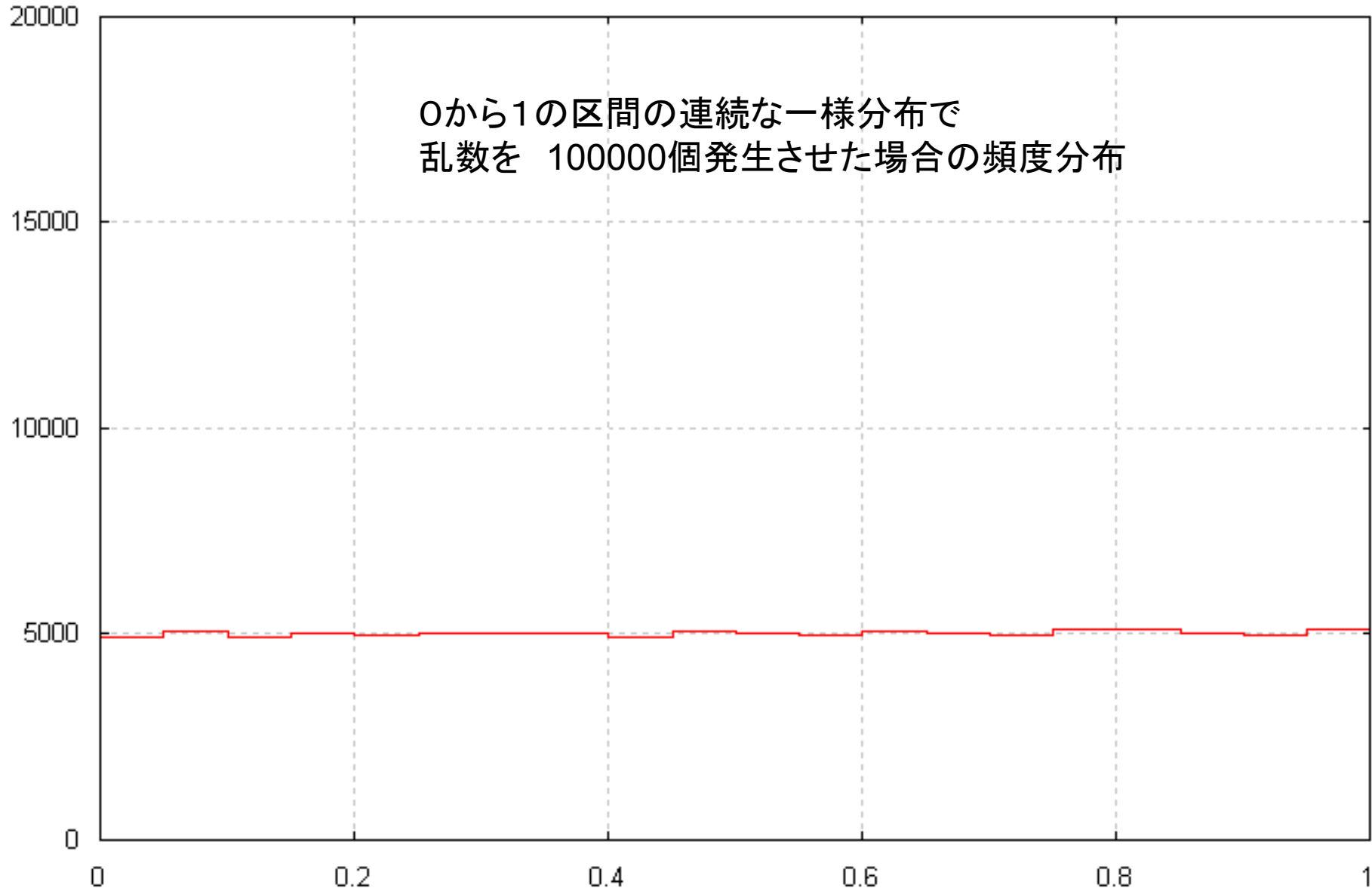
確率変数 x : 事象の起こった回数
 確率分布 $P(x)$ \rightarrow 二項分布 $B(n,p)$

$$P(x) = {}_n C_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

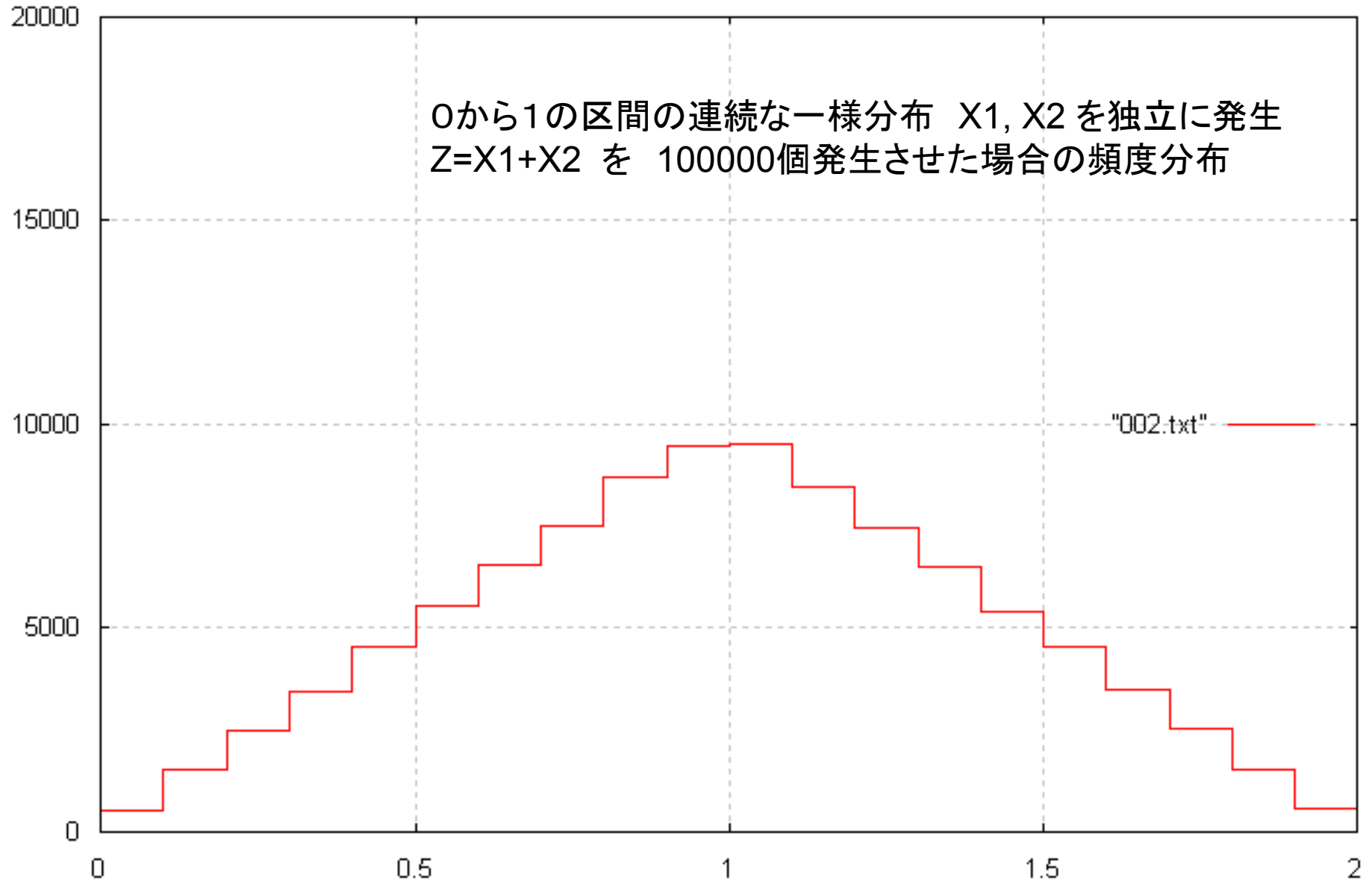


ベルヌイ分布の変数を
 20個足し合わせた
 二項分布
 ほぼ正規分布と一致

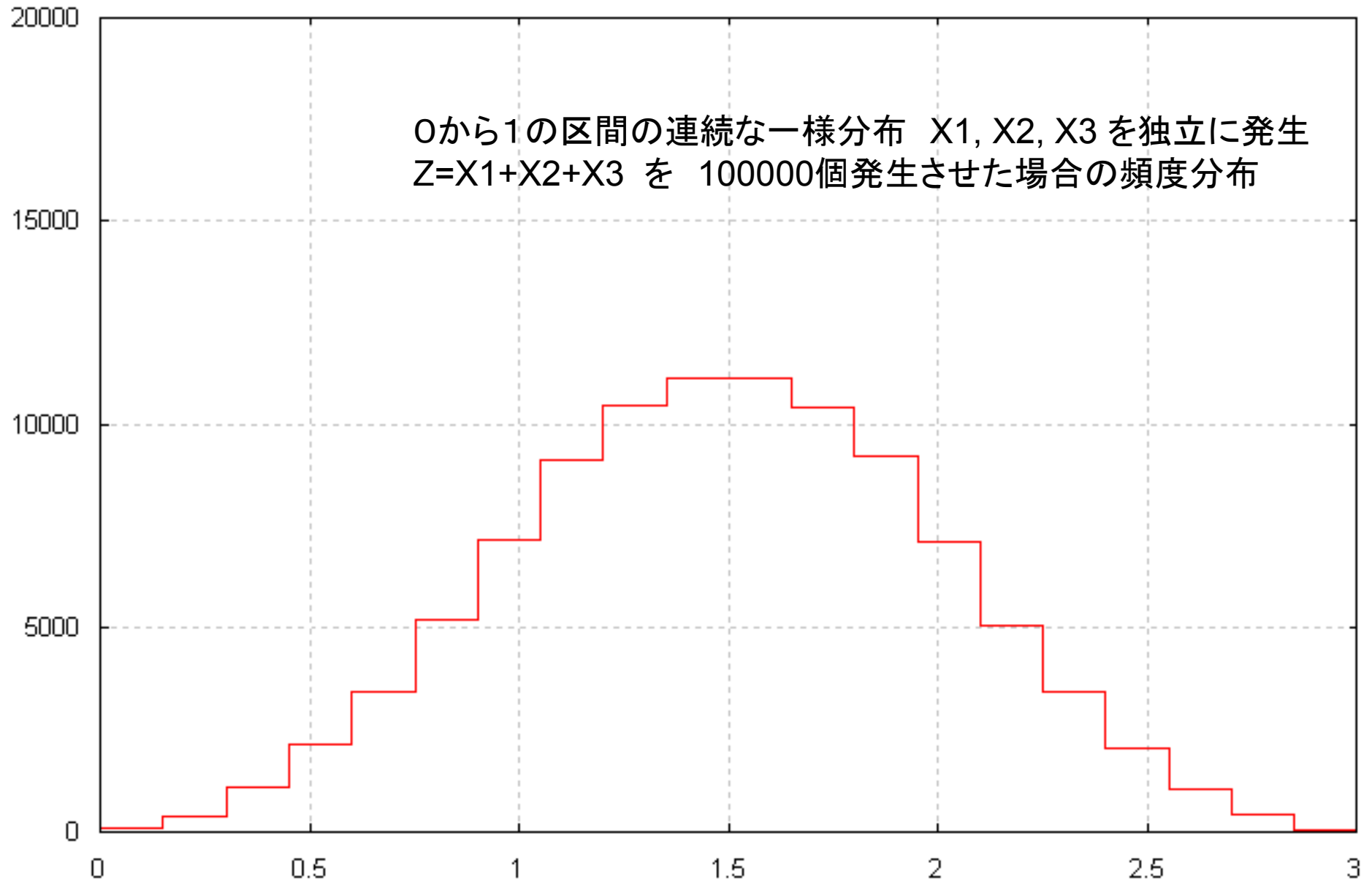
独立な確率変数の和の分布例(一様分布)



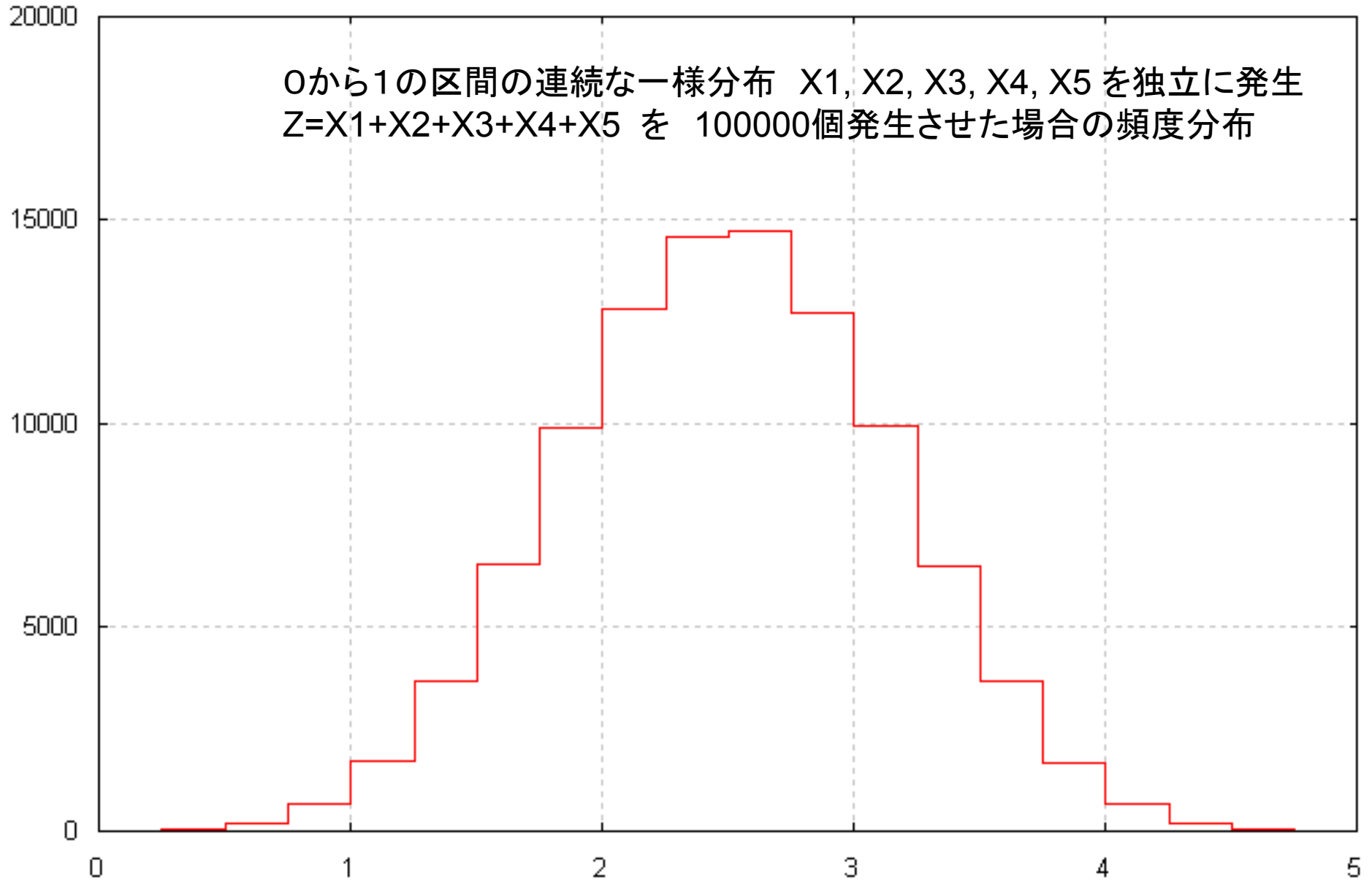
独立な確率変数の和の分布例(一様分布)



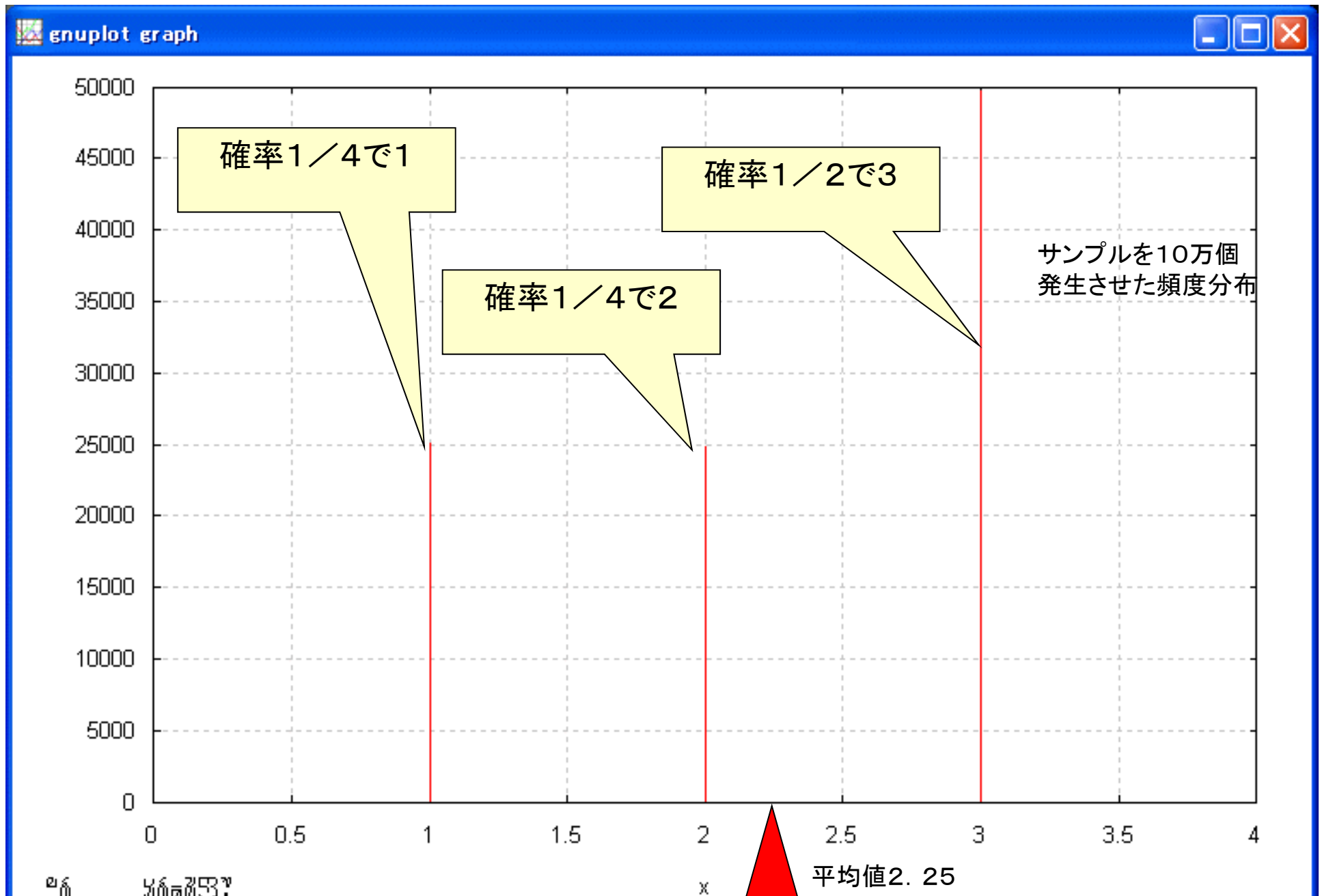
独立な確率変数の和の分布例(一様分布)



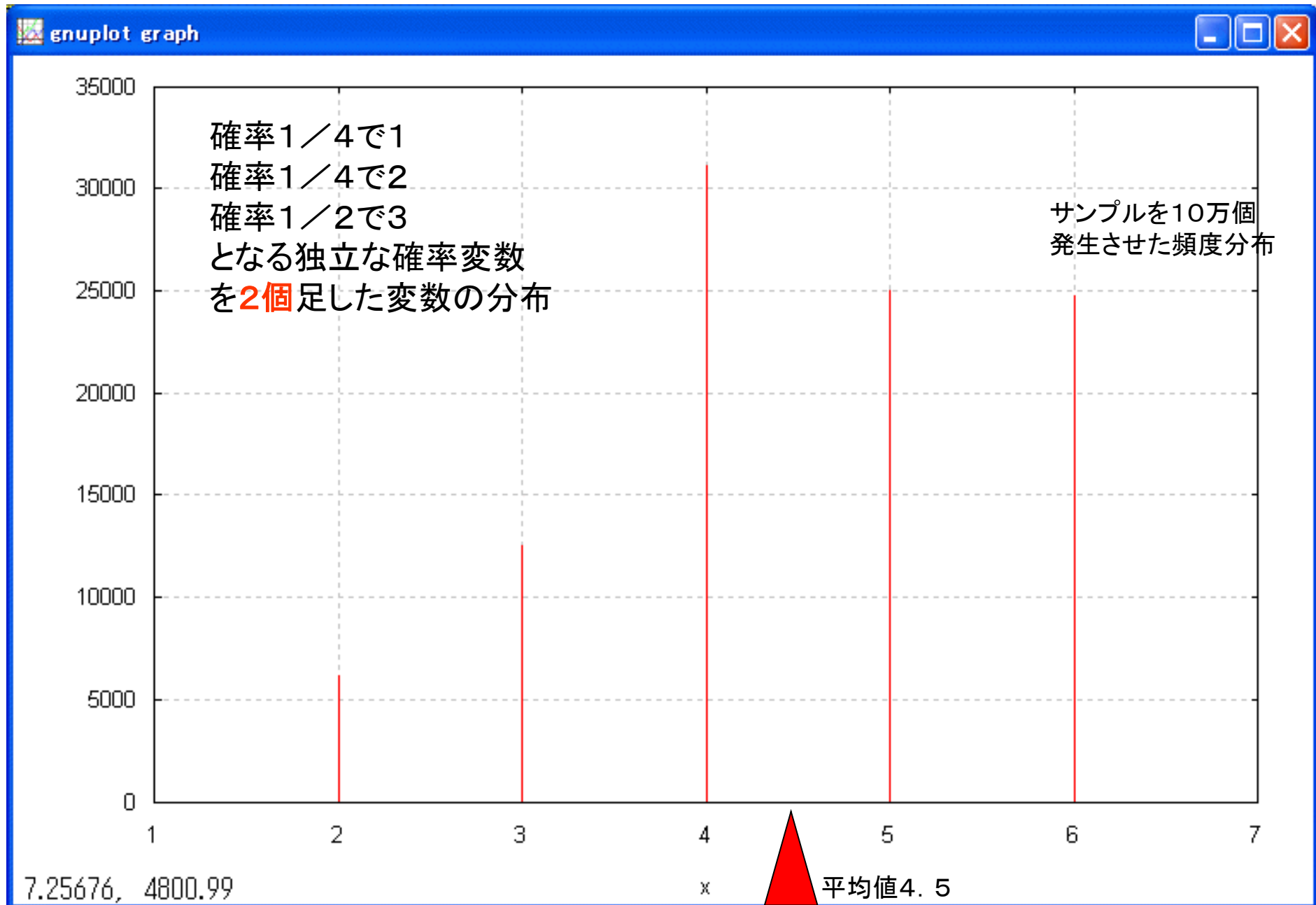
独立な確率変数の和の分布例(一様分布)



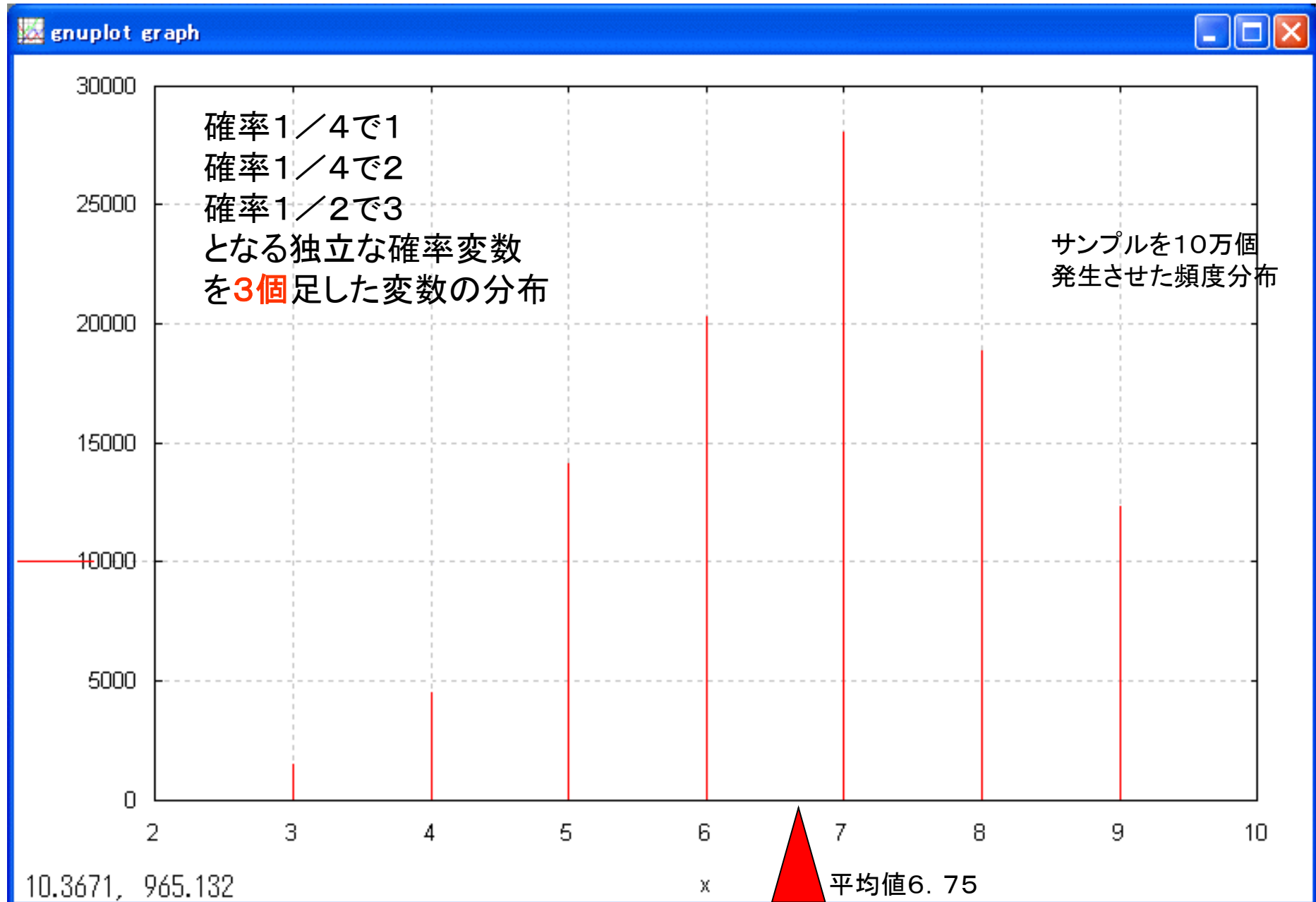
独立な確率変数の和の分布例(偏った離散分布)



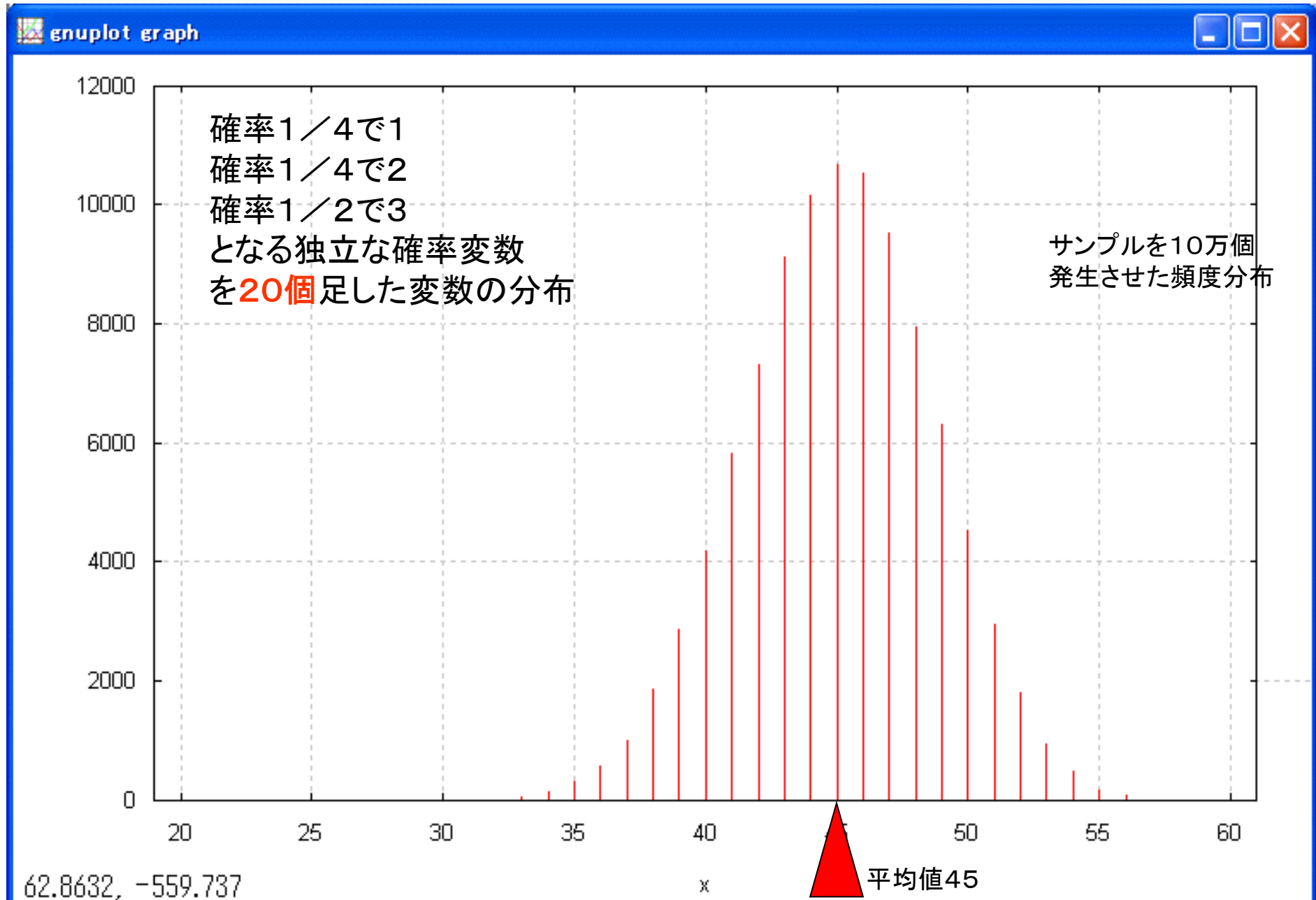
独立な確率変数の和の分布例(偏った離散分布)



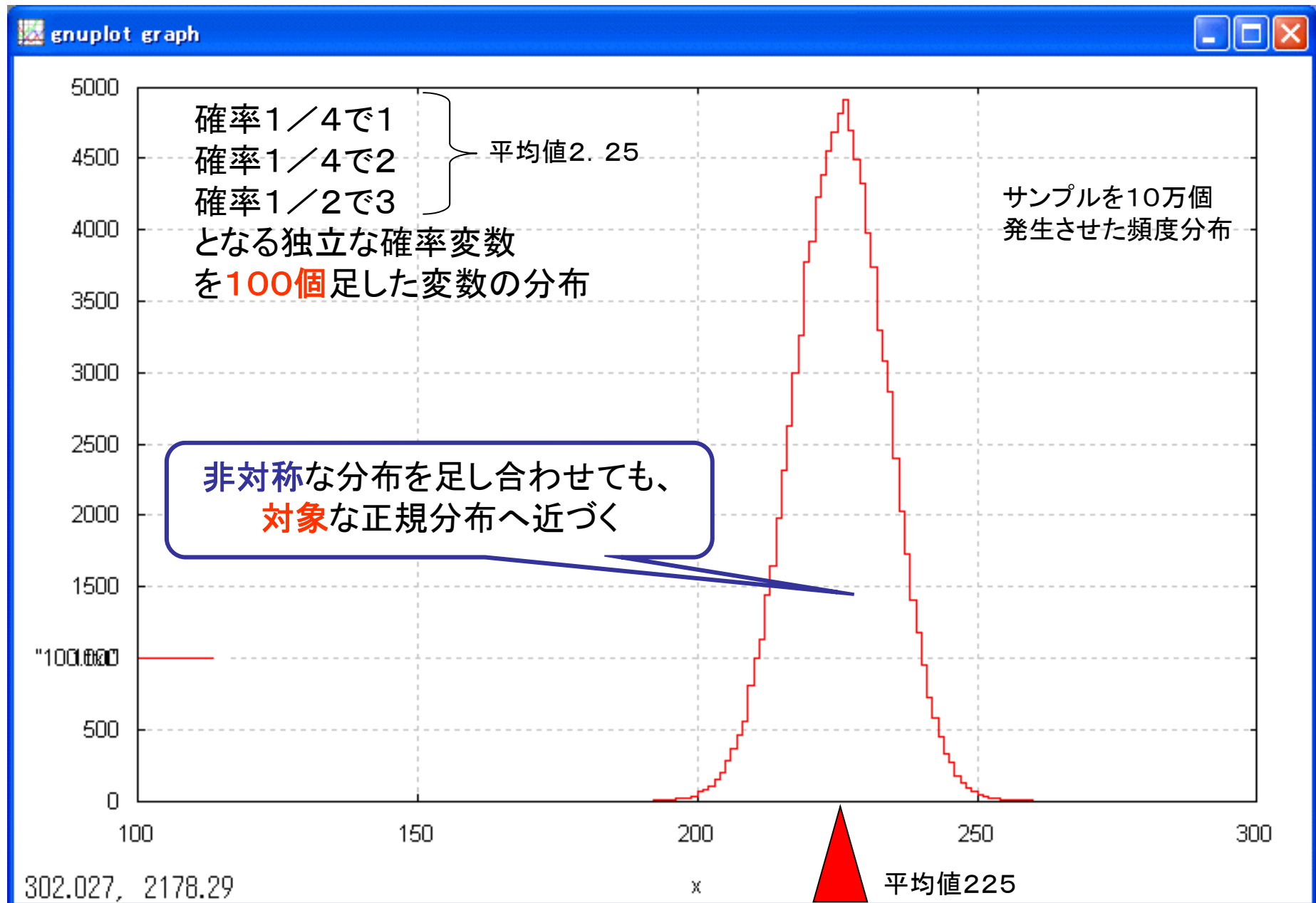
独立な確率変数の和の分布例(偏った離散分布)



独立な確率変数の和の分布例(偏った離散分布)



独立な確率変数の和の分布例(偏った離散分布)



【復習】2つの独立な確率変数の和の分布(一般型)

確率密度関数 $f_1(x)$ 信号
 $f_2(y)$ ノイズ

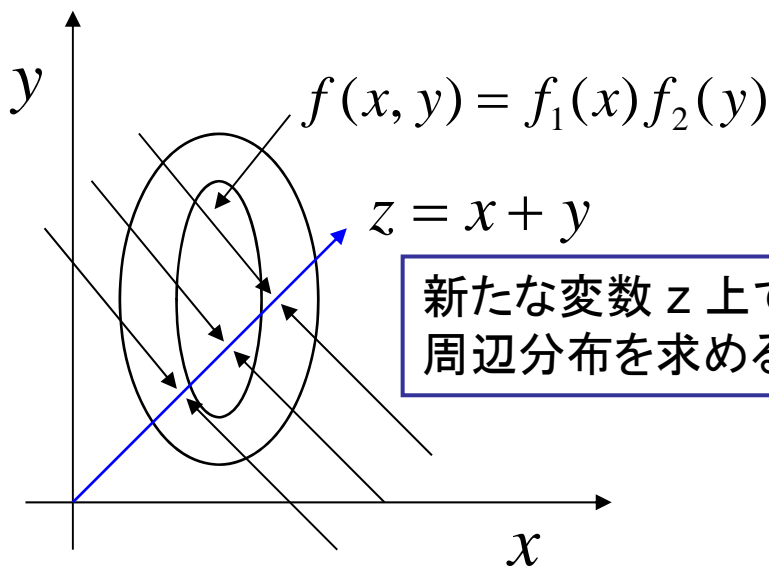
$$E\{x + y\} = E\{x\} + E\{y\}$$

独立な確率変数の和の期待値は、それぞれの期待値の和に等しい

$$\text{Var}\{x + y\} = \text{Var}\{x\} + \text{Var}\{y\}$$

独立な確率変数の和の分散は、それぞれの分散の和に等しい

確率変数 $z = x + y$ の確率密度分布関数は？



一般には $f_1(x)$ $f_2(y)$ とは
全く違う密度関数になる

$$f(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z - x) dx$$

たたみこみ積分 (convolution integral)

2つの独立な確率変数の和の分布(正規分布)

確率変数 X_1, X_2 がそれぞれ確率密度関数 $f_1(x) = N(\mu_1, \sigma_1^2)$ $f_2(x) = N(\mu_2, \sigma_2^2)$ に従うとき、これらの和 $Z = X_1 + X_2$ の確率密度関数は

$$f(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-x) dx = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{(z-x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2}} dx$$

たたみこみ積分

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{-\infty}^{\infty} e^{-\frac{1(\sigma_1^2+\sigma_2^2)}{2\sigma_1^2\sigma_2^2} \left(x - \frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_1^2+\sigma_2^2}\right)^2} dx$$

分散 = $\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$
の正規分布
を積分 = 1

$$= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}} \sqrt{2\pi} \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}$$

$$\frac{\int_{-\infty}^{\infty} e^{-\frac{1(\sigma_1^2+\sigma_2^2)}{2\sigma_1^2\sigma_2^2} \left(x - \frac{\sigma_2^2(z-\mu_1)+\sigma_1^2\mu_2}{\sigma_1^2+\sigma_2^2}\right)^2} dx}{\sqrt{2\pi} \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2}} e^{-\frac{(z-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}}$$

期待値 $\mu_1 + \mu_2$ 分散 $\sigma_1^2 + \sigma_2^2$
の正規分布
正規分布は特別な性質を持つ

$x_1, x_2, \dots, x_n, \dots$ が互いに独立確率変数で、それらが全て期待値 μ 分散 σ^2 の **任意の分布** に従うとき、

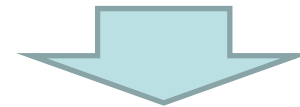
平均値 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ の分布は n が大きくなると

正規分布 $N(\mu, \frac{\sigma^2}{n})$ に近づく

分散

標準偏差 $\frac{\sigma}{\sqrt{N}}$

「1回の計測の誤差」= $\pm 2\sigma$
誤差を $1/10$ にするには？



$$\frac{2\sigma}{10} = \frac{2\sigma}{\sqrt{100}}$$

すなわち100回計測して平均をとれば誤差が $1/10$ になる

母比率の区間推定(大標本の場合)

調べる対象となる全体の集団:

母集団 (population)

標本 (sample)

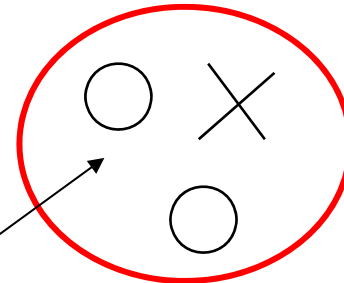
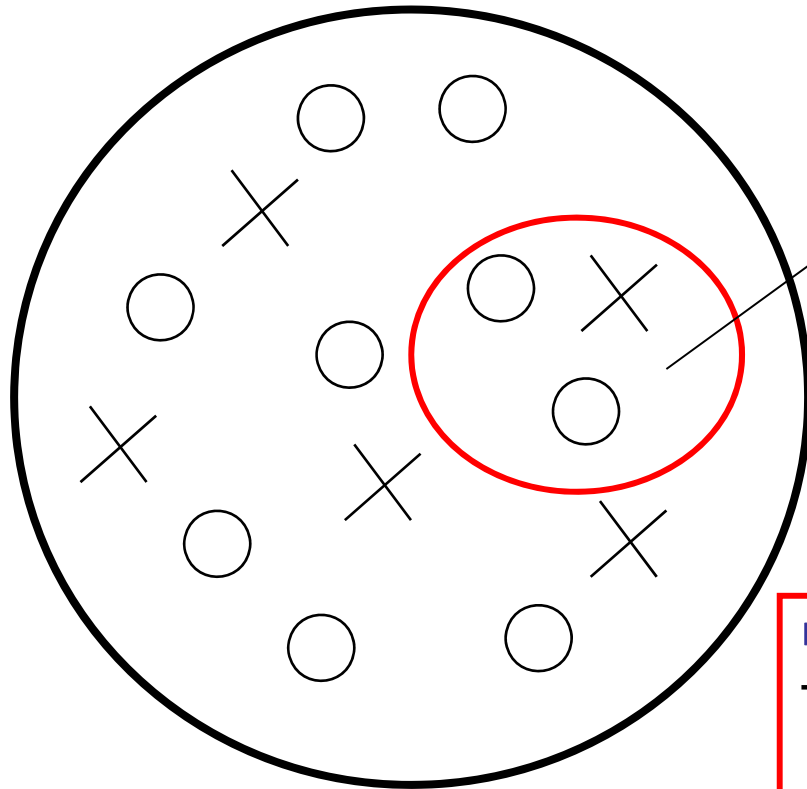
x_1, x_2, \dots, x_n

票を入れた場合1、
そうでなければ0を
とるような確率変数

x_i ベルヌイ分布

期待値 p

分散 $p(1-p)$



標本から計算される
パラメータ: (標本)統計量

標本平均 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ の

期待値は 分散

中心値極限定理より、

サンプル数 n が大きい場合、標本平均 \bar{x} の示す

分布は、正規分布

と見なせる

母集団 ある候補者に票を入れる
有権者の割合 p

これは母数 で、
一般に未知数

この性質を利用して標本より真の割合 p の
範囲を推定する → **母比率の区間推定**

母比率の区間推定(大標本の場合)

調べる対象となる全体の集団:

母集団 (population)

標本 (sample)

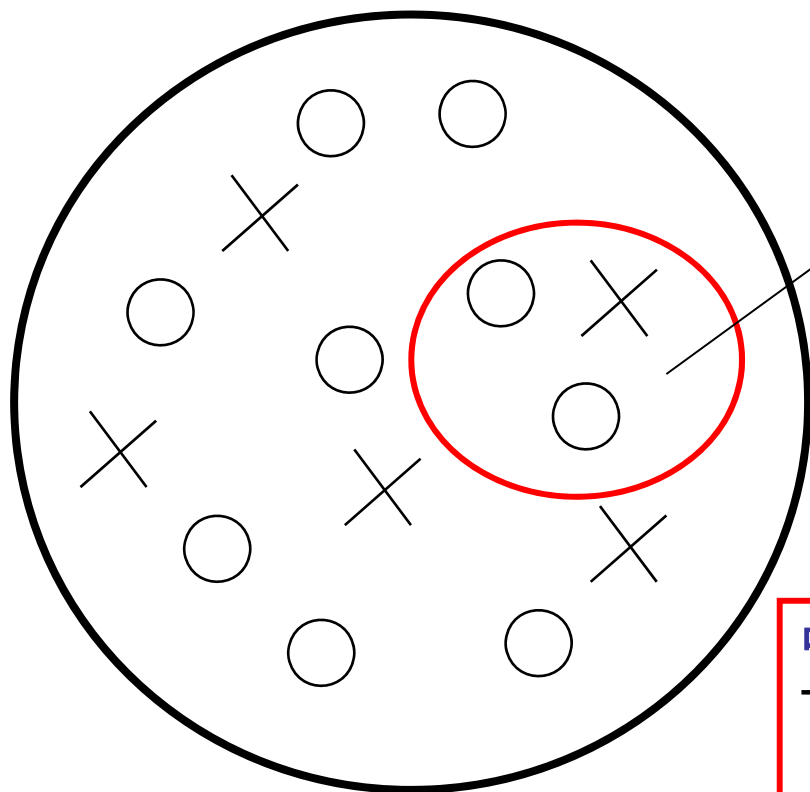
$$x_1, x_2, \dots, x_n$$

票を入れた場合1、
そうでなければ0を
とるような確率変数

x_i ベルヌイ分布

期待値 p

分散 $p(1-p)$



標本から計算される
パラメータ: (標本)統計量

標本平均 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ の

標本比率

期待値は p 分散

$$\frac{p(1-p)}{n}$$

母集団 ある候補者に票を入れる
有権者の割合 p

これは母数 **母比率** で、
一般に未知数

中心値極限定理より、
サンプル数 n が大きい場合、標本平均 \bar{x} の示す
分布は、正規分布 $N(p, \frac{p(1-p)}{n})$ と見なせる

この性質を利用して標本より真の割合 p の
範囲を推定する → **母比率の区間推定**

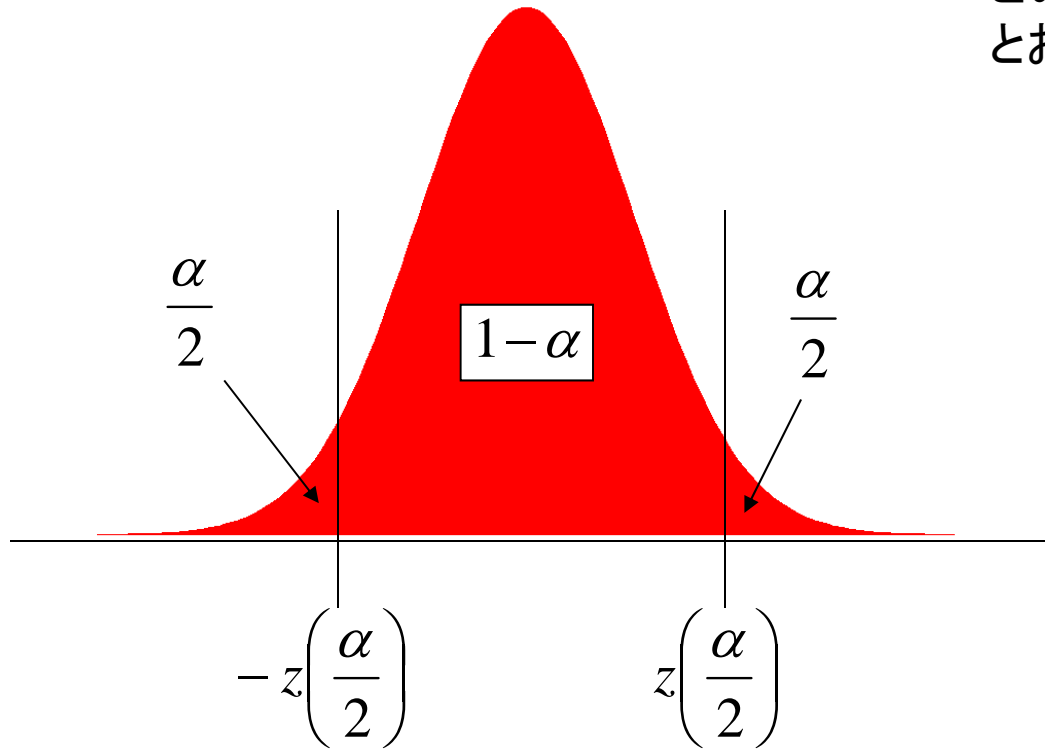
標本平均 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ の分布は、期待値 p 標準偏差 $\sqrt{\frac{p(1-p)}{n}}$ の正規分布

標準化



の分布は $N(0, 1)$ の標準正規分布 → 正規分布表を利用

信頼係数を $1 - \alpha$
 この信頼係数を満たす区間の上界をとくと、 $z\left(\frac{\alpha}{2}\right)$



$$P\left(-z\left(\frac{\alpha}{2}\right) \leq \boxed{} \leq z\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha$$

が成り立つ。この不等式

$$-z\left(\frac{\alpha}{2}\right) \leq \boxed{} \leq z\left(\frac{\alpha}{2}\right)$$

は、以下のように書き換えられる:

$$(\bar{x} - p)^2 \leq \left(z\left(\frac{\alpha}{2}\right)\right)^2 \frac{p(1-p)}{n}$$

標本平均 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ の分布は、期待値 p 標準偏差 $\sqrt{\frac{p(1-p)}{n}}$ の正規分布

標準化

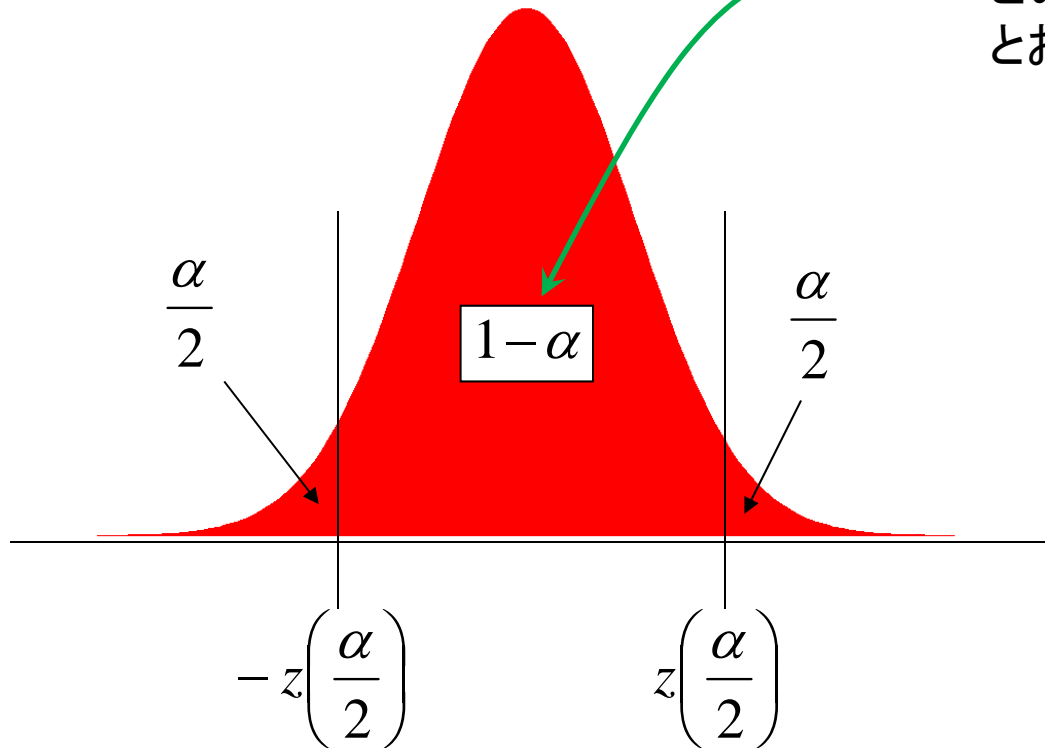
$$z = \frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

の分布は $N(0, 1)$ の標準正規分布 → 正規分布表を利用

信頼係数を $1 - \alpha$

この信頼係数を満たす区間の上界をとくと、

$$z\left(\frac{\alpha}{2}\right)$$



$$P\left(-z\left(\frac{\alpha}{2}\right) \leq \frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha$$

が成り立つ。この不等式

$$-z\left(\frac{\alpha}{2}\right) \leq \frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z\left(\frac{\alpha}{2}\right)$$

は、以下のように書き換えられる:

$$(\bar{x} - p)^2 \leq \left(z\left(\frac{\alpha}{2}\right)\right)^2 \frac{p(1-p)}{n}$$

$$(\bar{x} - p)^2 \leq \left(z \left(\frac{\alpha}{2} \right) \right)^2 \frac{p(1-p)}{n}$$

この式は以下のように変形できる:

$$\boxed{\phantom{(\bar{x} - p)^2 \leq \left(z \left(\frac{\alpha}{2} \right) \right)^2 \frac{p(1-p)}{n}}} \leq 0$$

下に凸のpに関する
2次関数が負になる
範囲を求めればよい

$$p_1 \leq p \leq p_2 \quad \text{ただし}$$

$$p_1 = \frac{\bar{x} + \frac{\left(z \left(\frac{\alpha}{2} \right) \right)^2}{2n} - \frac{z \left(\frac{\alpha}{2} \right)}{\sqrt{n}} \sqrt{\bar{x}(1-\bar{x}) + \frac{1}{4n} \left(z \left(\frac{\alpha}{2} \right) \right)^2}}{1 + \frac{1}{n} \left(z \left(\frac{\alpha}{2} \right) \right)^2}$$

標本数 n が大きいとき、

$\frac{1}{n} \left(z \left(\frac{\alpha}{2} \right) \right)^2$ を0として無視すると

$$p_2 = \frac{\bar{x} + \frac{\left(z \left(\frac{\alpha}{2} \right) \right)^2}{2n} + \frac{z \left(\frac{\alpha}{2} \right)}{\sqrt{n}} \sqrt{\bar{x}(1-\bar{x}) + \frac{1}{4n} \left(z \left(\frac{\alpha}{2} \right) \right)^2}}{1 + \frac{1}{n} \left(z \left(\frac{\alpha}{2} \right) \right)^2}$$



$$(\bar{x} - p)^2 \leq \left(z \left(\frac{\alpha}{2} \right) \right)^2 \frac{p(1-p)}{n} \quad \text{この式は以下のように変形できる:}$$

$$\left(1 + \frac{1}{n} \left(z \left(\frac{\alpha}{2} \right) \right)^2 \right) p^2 - \left(2\bar{x} + \frac{1}{n} \left(z \left(\frac{\alpha}{2} \right) \right)^2 \right) p + \bar{x}^2 \leq 0$$

下に凸のpに関する
2次関数が負になる
範囲を求めればよい

$$p_1 \leq p \leq p_2 \quad \text{ただし}$$

$$p_1 = \frac{\bar{x} + \frac{\left(z \left(\frac{\alpha}{2} \right) \right)^2}{2n} - \frac{z \left(\frac{\alpha}{2} \right)}{\sqrt{n}} \sqrt{\bar{x}(1-\bar{x}) + \frac{1}{4n} \left(z \left(\frac{\alpha}{2} \right) \right)^2}}{1 + \frac{1}{n} \left(z \left(\frac{\alpha}{2} \right) \right)^2}$$

標本数 n が大きいとき、

$\frac{1}{n} \left(z \left(\frac{\alpha}{2} \right) \right)^2$ を0として無視すると

$$p_2 = \frac{\bar{x} + \frac{\left(z \left(\frac{\alpha}{2} \right) \right)^2}{2n} + \frac{z \left(\frac{\alpha}{2} \right)}{\sqrt{n}} \sqrt{\bar{x}(1-\bar{x}) + \frac{1}{4n} \left(z \left(\frac{\alpha}{2} \right) \right)^2}}{1 + \frac{1}{n} \left(z \left(\frac{\alpha}{2} \right) \right)^2}$$

$$p_1 = \bar{x} - z \left(\frac{\alpha}{2} \right) \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

$$p_2 = \bar{x} + z \left(\frac{\alpha}{2} \right) \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

母比率の大標本区間推定(まとめ)

調べる対象となる全体の集団:

母集団 (population)

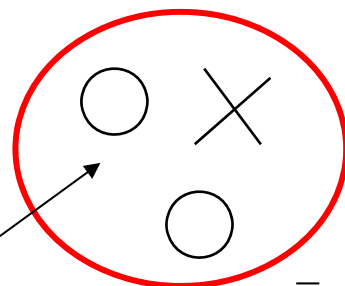
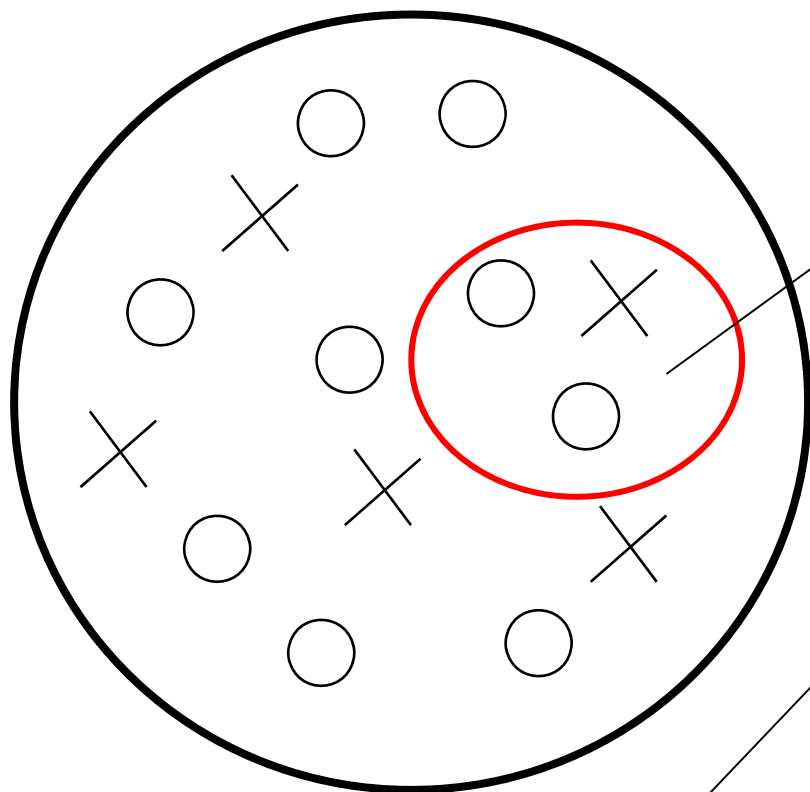
標本 (sample)

$$x_1, x_2, \dots, x_n$$

ただし n は大きな数

標本平均 (**標本比率**)

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$



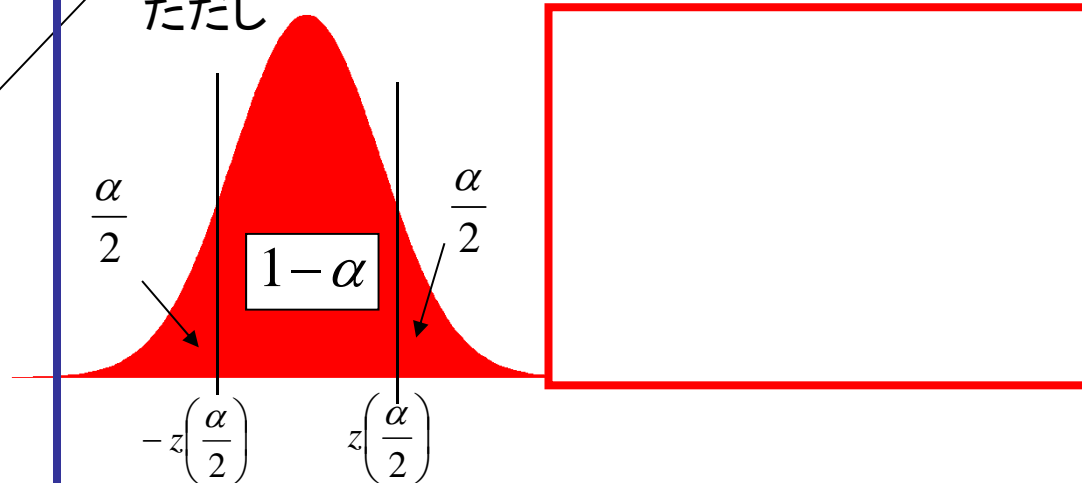
$$z = \frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

の分布は $N(0, 1)$
(標準正規分布)

母集団 ある候補者に票を入れる
有権者の割合 p

これは母数 (**母比率**) で、
一般に未知数

よって真の割合 p の範囲は **信頼係数** $1 - \alpha$
のとき $p_1 \leq p \leq p_2$
ただし



母比率の大標本区間推定(まとめ)

調べる対象となる全体の集団:

母集団 (population)

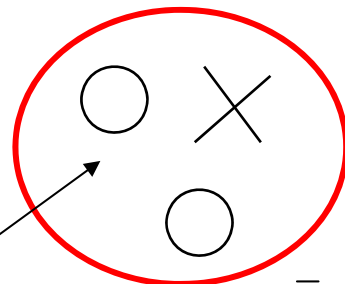
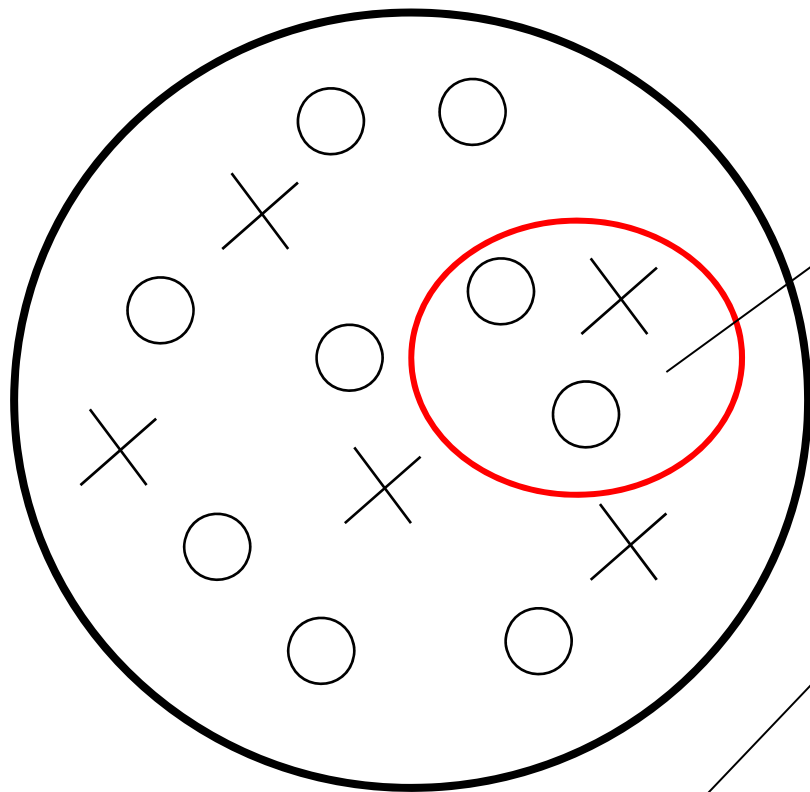
標本 (sample)

$$x_1, x_2, \dots, x_n$$

ただし n は大きな数

標本平均 (**標本比率**)

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$



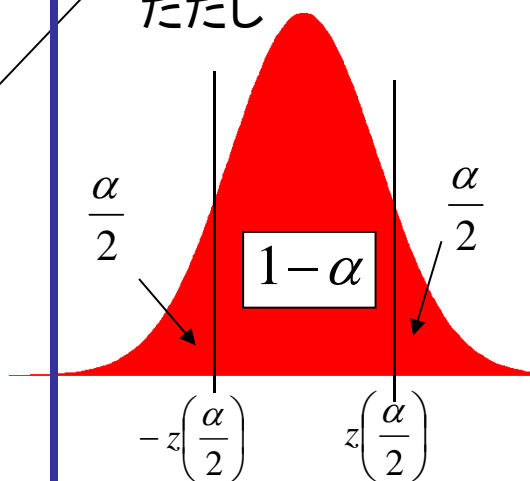
$$z = \frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

の分布は $N(0, 1)$
(標準正規分布)

母集団 ある候補者に票を入れる
有権者の割合 p

これは母数 (**母比率**) で、
一般に未知数

よって真の割合 p の範囲は **信頼係数** $1 - \alpha$
のとき $p_1 \leq p \leq p_2$
ただし



$$p_1 = \bar{x} - z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

$$p_2 = \bar{x} + z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

母比率の大標本区間推定

補足

【注意】 標本数 n は大きな数でなければならない

実用上は $np > 5$ and $n(1-p) > 5$ のとき正規分布に近似できる

ただし p は母比率

【練習問題】

- (1) 内閣の支持率の世論調査のため、無作為に2500人を選んだら、内閣の支持者が500人だった。このとき、内閣の支持率 p の95%信頼区間を求めよ。

【練習問題】

- (1) 内閣の支持率の世論調査のため、無作為に2500人を選んだら、内閣の支持者が500人だった。このとき、内閣の支持率 p の95%信頼区間を求めよ。

$$\bar{x} = \frac{500}{2500} = \frac{1}{5} \quad \text{正規分布表より } 1 - \alpha = 0.95 \quad \text{となるような } z\left(\frac{\alpha}{2}\right) = ?$$

【練習問題】

- (1) 内閣の支持率の世論調査のため、無作為に2500人を選んだら、内閣の支持者が500人だった。このとき、内閣の支持率 p の95%信頼区間を求めよ。

$$\bar{x} = \frac{500}{2500} = \frac{1}{5} \quad \text{正規分布表より } 1 - \alpha = 0.95 \quad \text{となるような } z\left(\frac{\alpha}{2}\right) = 1.96$$

$$\left[\bar{x} - 1.96 \times \sqrt{\frac{\bar{x}(1 - \bar{x})}{2500}}, \quad \bar{x} + 1.96 \times \sqrt{\frac{\bar{x}(1 - \bar{x})}{2500}} \right] = [0.184, \quad 0.216]$$

【演習問題】

2018.05.08

学籍番号

氏名

[問] 母比率 p を標本比率 $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$ から信頼係数95%で推定したい。

- (1) すべての \bar{x} の値に対して $\bar{x}(1 - \bar{x}) \leq 0.25$ が成り立つことを示せ。
- (2) 信頼区間の幅を 0.04 以下にするには、標本の大きさ n をいくら以上にとれば良いか。(1)の式を利用して導け。

【演習問題】

学籍番号

氏名

[問] 母比率 p を標本比率 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ から信頼係数95%で推定したい。

(1) すべての \bar{x} の値に対して $\bar{x}(1 - \bar{x}) \leq 0.25$ が成り立つことを示せ。

$\bar{x}(1 - \bar{x})$ は0と1を通り上に凸な2次関数で、 $\bar{x} = \frac{1}{2}$ のとき最大値 $1/4$

(2) 信頼区間の幅を 0.04 以下にするには、標本の大きさ n をいくら以上にとれば良いか。

信頼区間の幅は、 $2 \times z\left(\frac{\alpha}{2}\right) \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}} \leq 0.04$

よって

$$2 \times 1.96 \sqrt{\frac{0.25}{n}} \leq 0.04$$

式変形すると

$$2401 \leq n$$

正規分布表より $1 - \alpha = 0.95$ となるような $z\left(\frac{\alpha}{2}\right) = 1.96$