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）

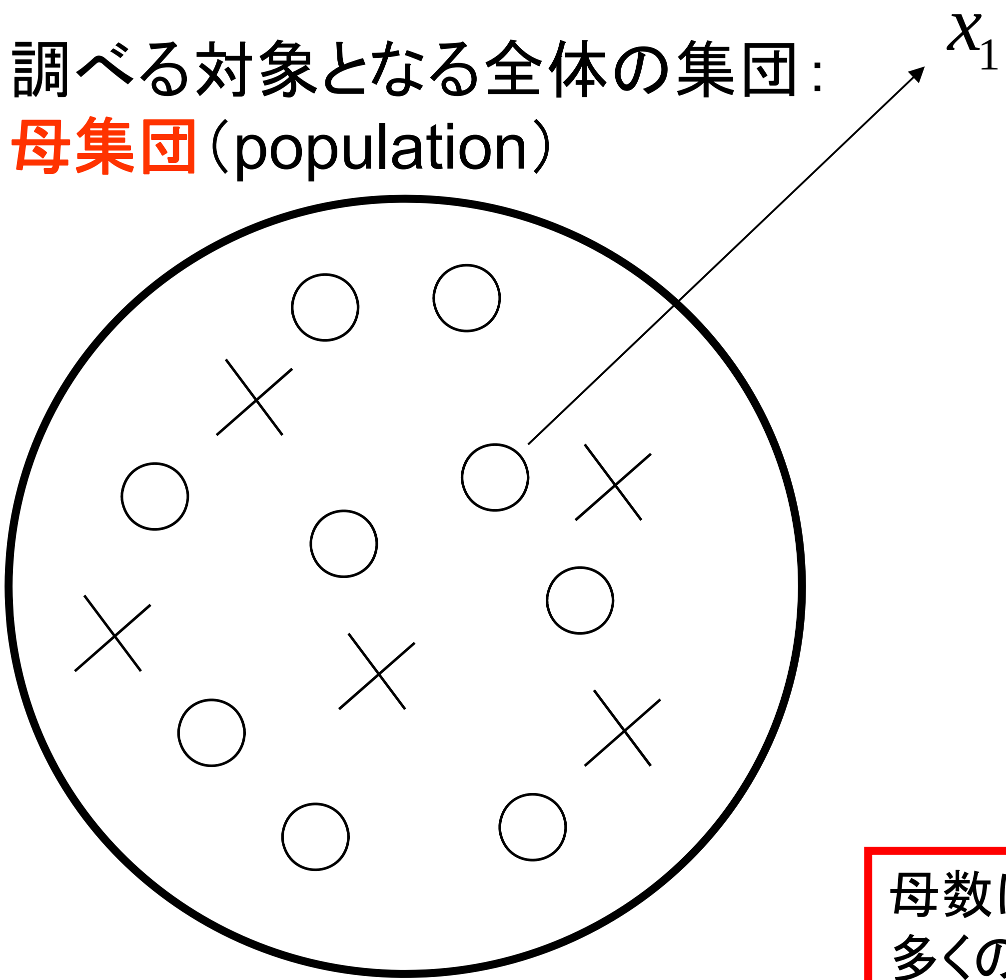
標本分布１

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)



・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 =

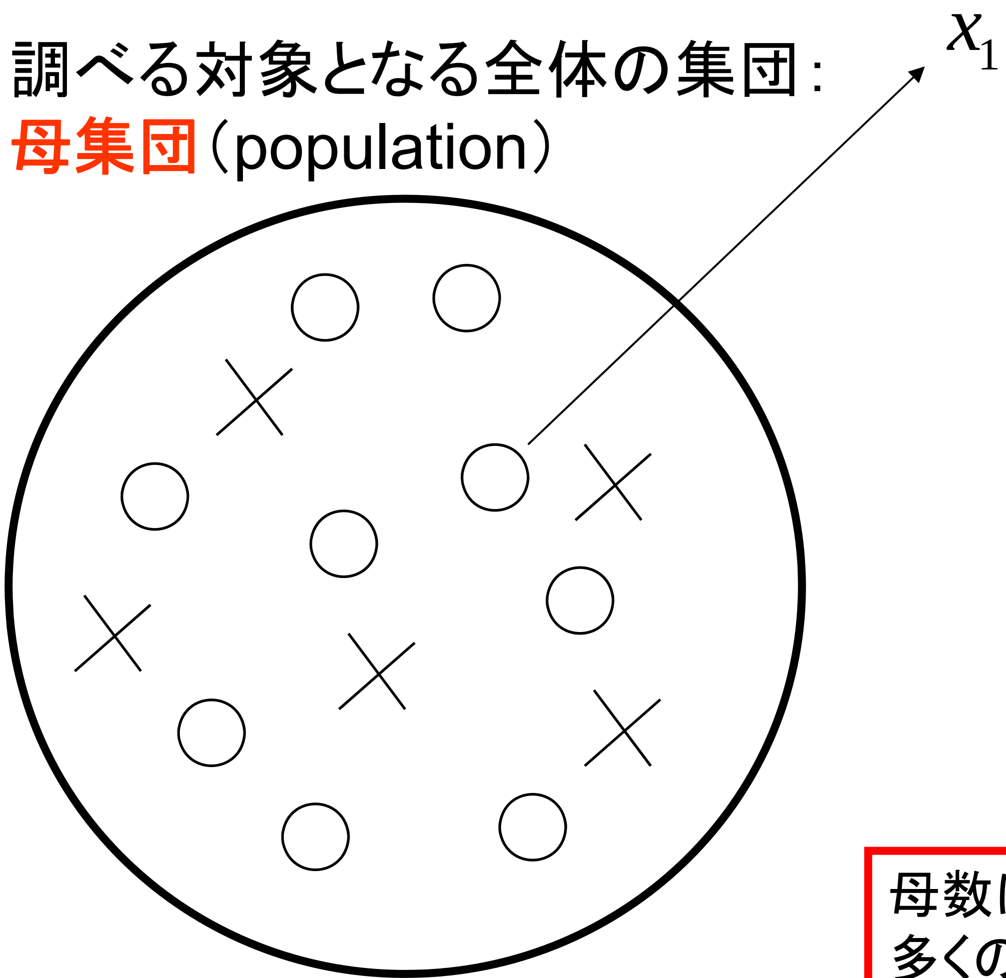
・母集団分布を特徴付けるパラメータ:
(例:平均・分散・相関係数など)
母数 (population parameter)

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）



・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　＝ $\boxed{\text{母集団分布}}$

・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）

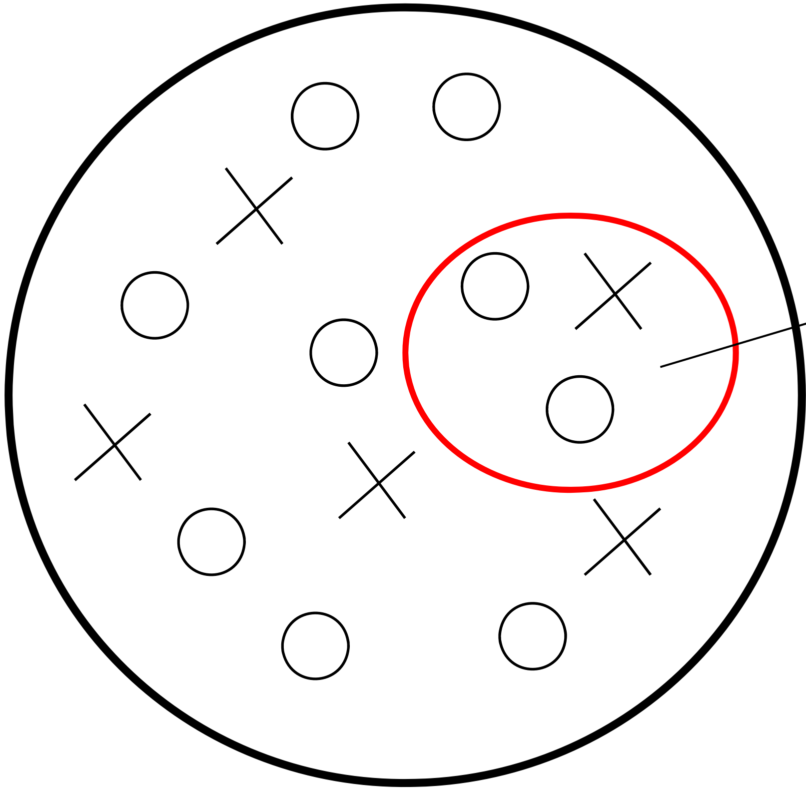
母数は、一般に未知
多くの場面において、母数を知る必要有

母集団全体を調べれば良いが…

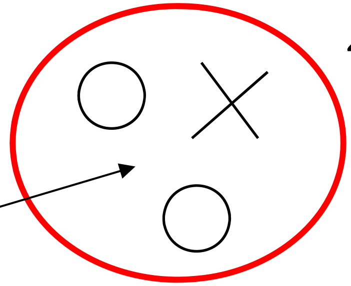
母集団と標本

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

調べる対象となる全体の集団：
母集団（population）



母集団の中から選ばれる一部分の集まり：　**標本**（sample）



$x_1, x_2, \cdots, x_n$

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

母集団のとある性質を示すパラメータ：　母数　p

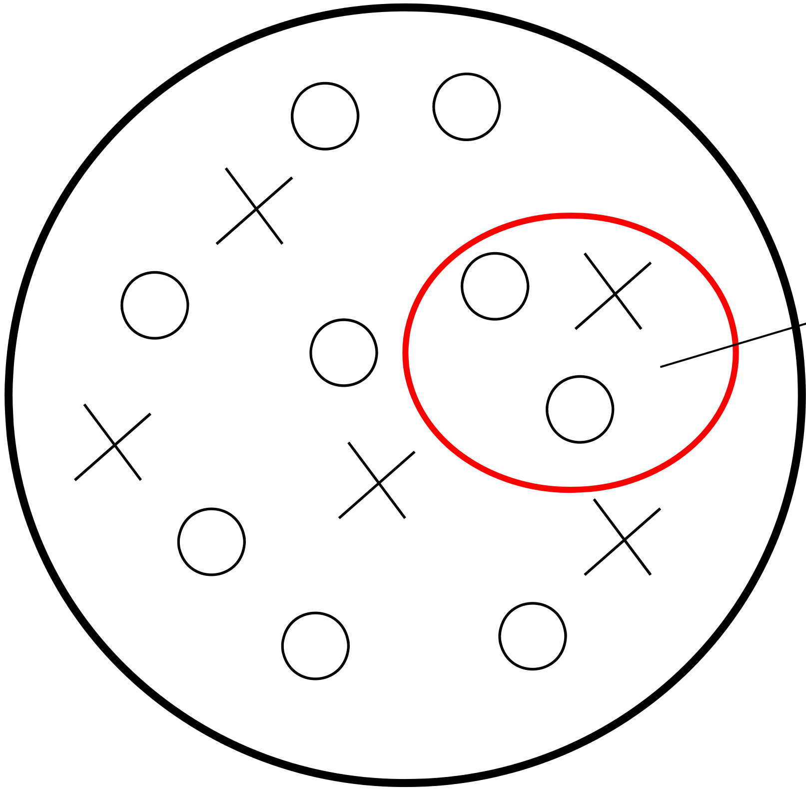
統計量は、標本毎に異なる値をとる確率変数
この統計量の分布を標本分布という

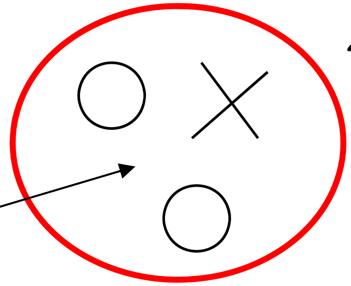
母集団と標本

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

調べる対象となる全体の集団：
母集団（population）

母集団の中から選ばれる一部分の集まり： **標本**（sample）





$x_1, x_2, \cdots, x_n$

標本について調べることで、母数を推測
→ 統計的推論 $\hat{p}$

母集団のとある性質を示すパラメータ： 母数 p

標本から計算されるパラメータ： （標本）統計量

標本平均 $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$

標本分散 $S^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

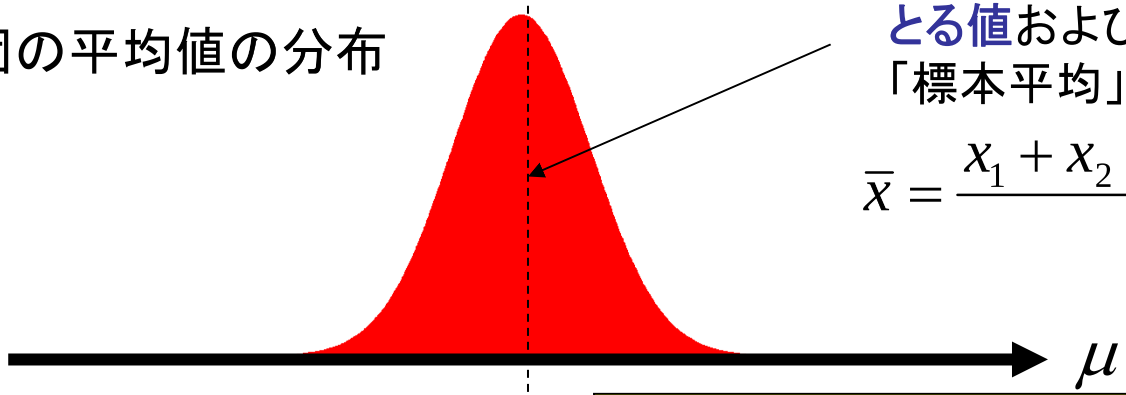
不偏分散 $U^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

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

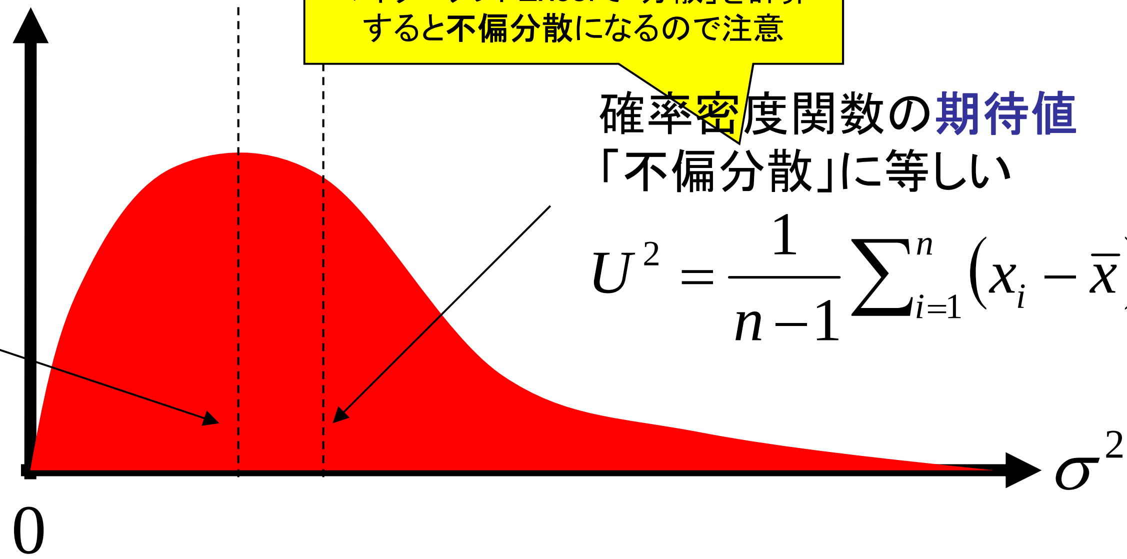


確率密度関数の**最大値をとる値**および**期待値**が
「標本平均」に等しい

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

マイクロソフトExcelで「分散」を計算
すると**不偏分散**になるので注意

【例２】 母集団の分散の分布



確率密度関数の**期待値**
「不偏分散」に等しい

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

確率密度関数が
最大になる値
「標本分散」に等しい

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

不偏分散の式 $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ において、なぜ 1/n ではなく1/(n-1) なのか？

真のμ

$$E\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2\right\} = \sum_{i=1}^{n}E\left\{(x_i - \bar{x})^2\right\} = \sum_{i=1}^{n}E\left\{(x_i - \mu - (\bar{x} - \mu))^2\right\}$$

$$= \sum_{i=1}^{n}E\left\{(x_i - \mu)^2\right\} + \sum_{i=1}^{n}E\left\{(\bar{x} - \mu)^2\right\} + 2\sum_{i=1}^{n}E\left\{(x_i - \mu)(\bar{x} - \mu)\right\}$$

$$\boxed{\sigma^2} \qquad \boxed{\sigma^2/n}$$

$$\boxed{\begin{aligned} &2\sum_{i=1}^{n}E\left\{(x_i - \mu)(\bar{x} - \mu)\right\} = 2E\left\{\sum_{i=1}^{n}(x_i - \mu)(\bar{x} - \mu)\right\} \\ &= 2E\left\{n(\bar{x} - \mu)(\bar{x} - \mu)\right\} = 2nE\left\{(\bar{x} - \mu)^2\right\} = 2n(\sigma^2/n) = 2\sigma^2 \end{aligned}}$$

よって

$$E\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2\right\} = n\sigma^2 + \sigma^2 - 2\sigma^2 = (n-1)\sigma^2$$

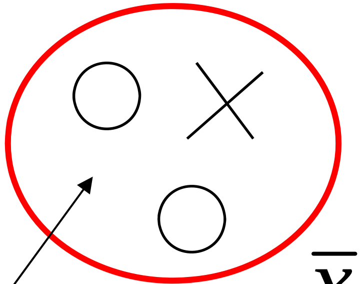
よって $E\left\{\hat{\sigma}^2\right\} = \sigma^2$ 　平均的には真の分散に等しくなる

これに対して、

1／nで割った値	$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$	は確率密度が最大 （最ユウ推定値）

母集団と標本

標本（sample）　$x_1, x_2, \cdots, x_n$



標本平均　$\overline{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$

$\overline{X}$ を確率変数とした標本分布を考える

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）

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

$$= \frac{1}{n}\left(E\{x_1\} + E\{x_2\} + \cdots E\{x_n\}\right)$$

$$= \boxed{\quad\quad}$$

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

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

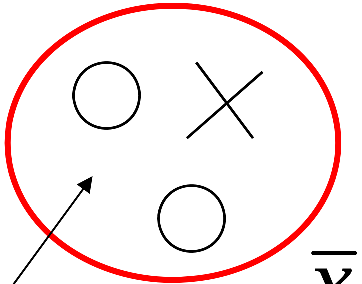
$$= \frac{1}{n^2}\left(Var\{x_1\} + Var\{x_2\} + \cdots Var\{x_n\}\right)$$

$$= \boxed{\quad\quad}$$

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 $\overline{X}$ の
期待値は
分散は

母集団と標本

標本（sample）　$x_1, x_2, \cdots, x_n$



標本平均　$\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$

$\bar{x}$ を確率変数とした標本分布を考える

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）

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

$$= \frac{1}{n}\left(E\{x_1\} + E\{x_2\} + \cdots E\{x_n\}\right)$$

$$= \boxed{\mu}$$

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

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

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

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

サンプル数 n が増えれば小さくなる

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　σ^2/n

大数の法則

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

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

$$V_n = Var\{x_1 + x_2 + \cdots + x_n\}$$

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

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

は0に確率収束する。

（大数の弱法則）

つまり、$\boxed{\text{サンプルを多数とれば、得られる平均値は母平均（真の期待値）に近づく}}$

大数の法則

$x_1, x_{2,} \cdots, x_n, \cdots$ がそれぞれ確率変数で、

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

$$V_n = Var\{x_1 + x_2 + \cdots + x_n\}$$

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

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

中心値極限定理　Central limit theorem

$x_1, x_2, \cdots, x_n, \cdots$ 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 $\boxed{\qquad\qquad}$ に従うとき、

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

正規分布 $\boxed{\qquad\qquad}$ に近づく

標準化

すなわち $\boxed{\qquad\qquad}$ は標準正規分布 N(0,1) に近づく

$x_1, x_2, \cdots, x_n, \cdots$ 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

中心極限定理　Central limit theorem

$x_1, x_2, \cdots, x_n, \cdots$　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 $\boxed{\text{任意の分布}}$ に従うとき、

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

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

標準化

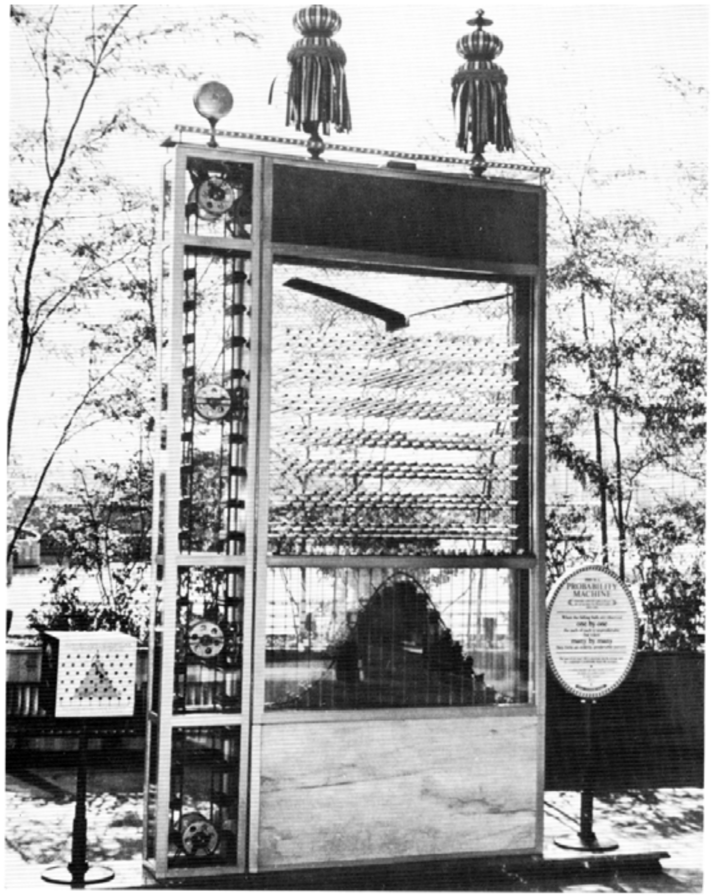
すなわち $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}}$ は標準正規分布 N(0,1) に近づく

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

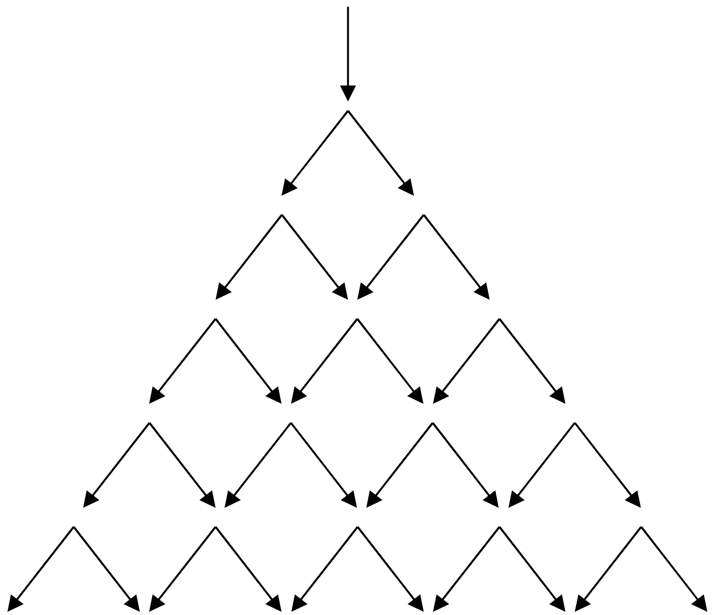
期待値 μ 分散 σ^2 の正規分布に従うとき、
n の大きさにかかわらず上記の正規分布になる

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 → ベルヌイ試行
この試行を n 回行う

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)

ベルヌイ試行
ベルヌイ分布

（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 回おこることを考えると、

$$\underbrace{\bigcirc\ \bigcirc \cdots \bigcirc}_{x\text{ 回}}\quad \underbrace{\times\ \times \cdots \times}_{n-x\text{ 回}}$$

このときの確率は $p^x(1-p)^{n-x}$

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

$${}_nC_x = \frac{n!}{x!(n-x)!}$$

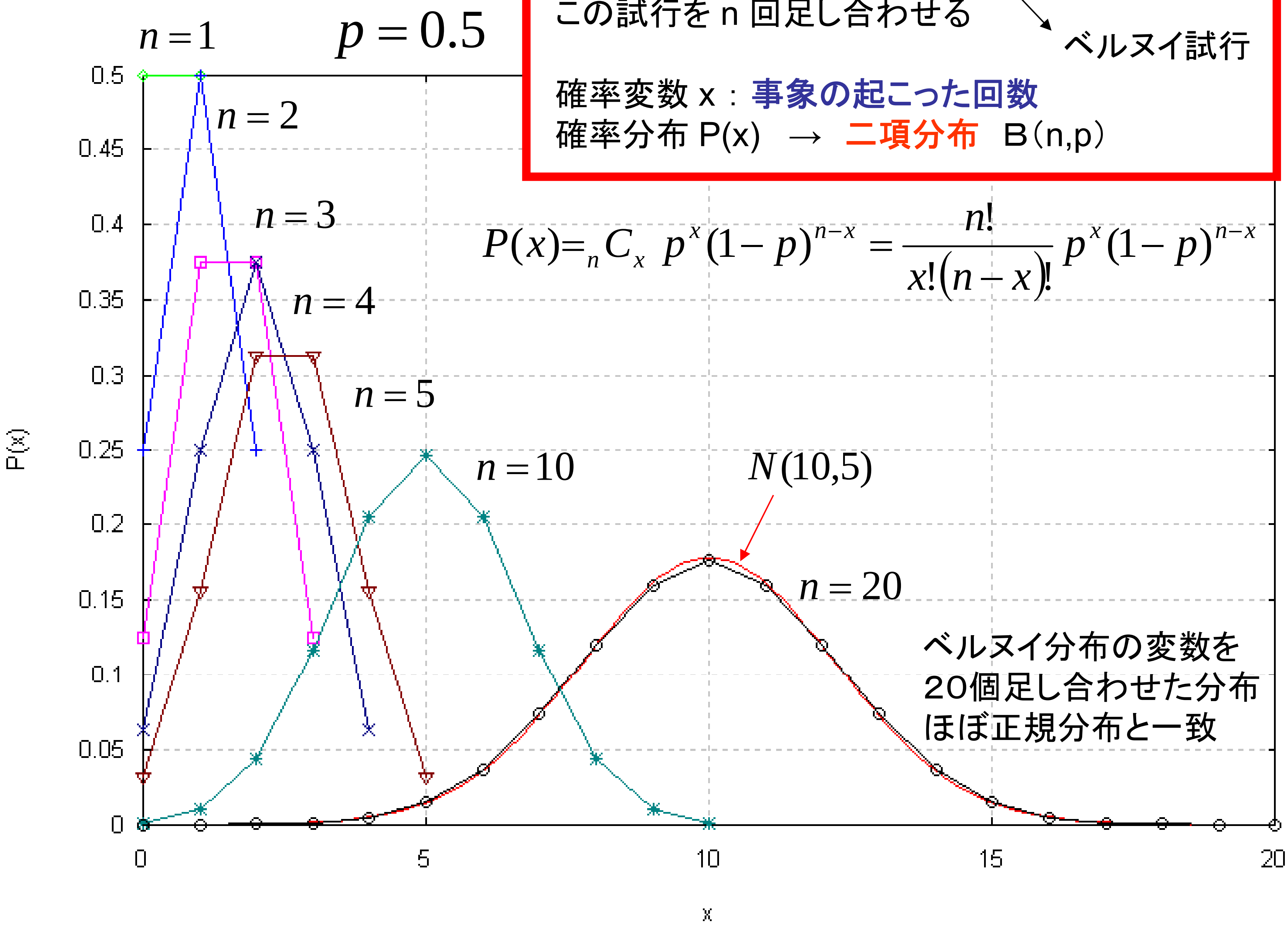
よって $P(x) = {}_nC_x\, p^x(1-p)^{n-x} = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$

【復習】 二項分布の極限

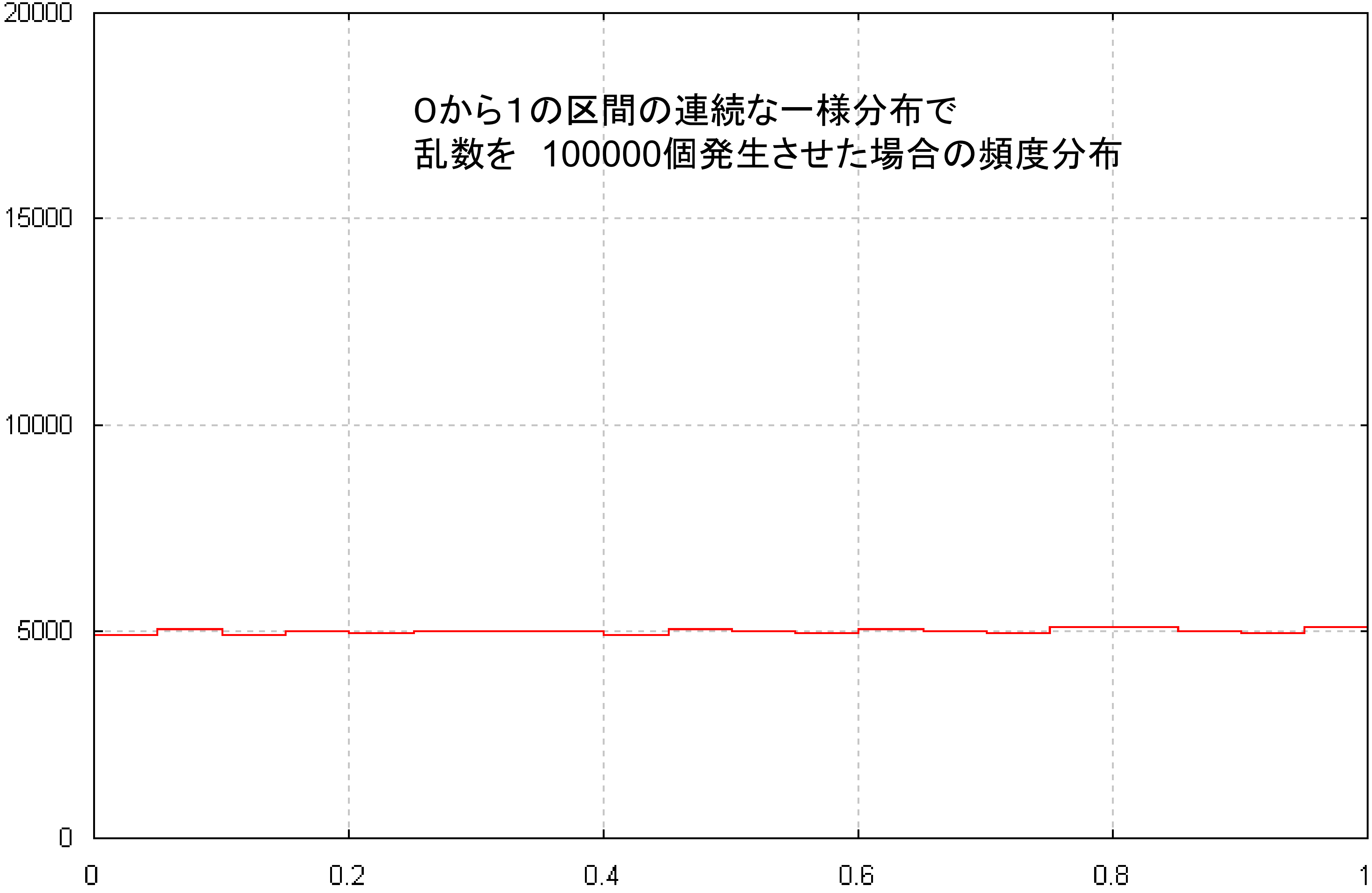
1回の試行で、ある事象の起こる確率が p
この試行を n 回足し合わせる → ベルヌイ試行

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

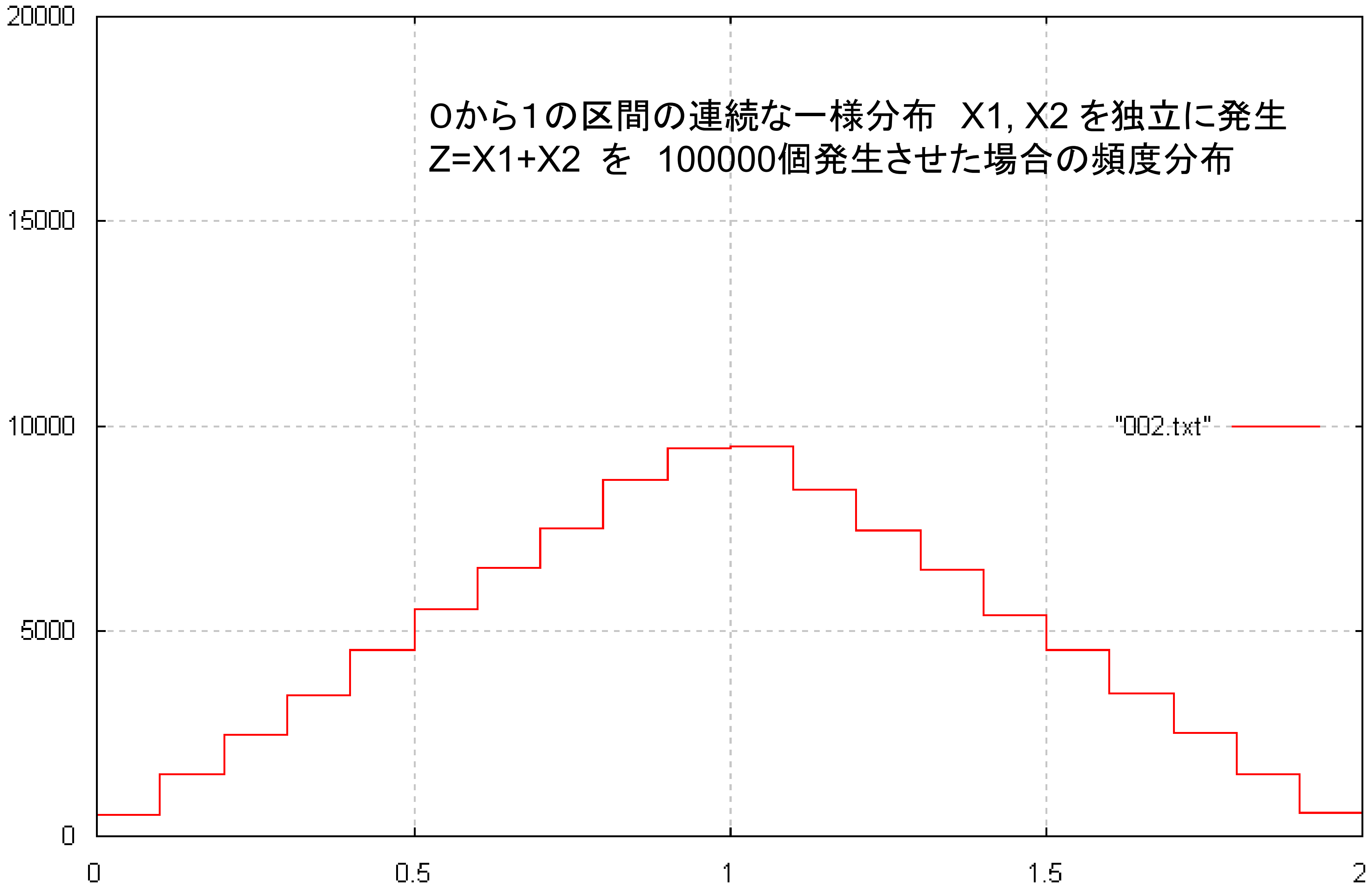
$$P(x) = {}_nC_x \; p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$



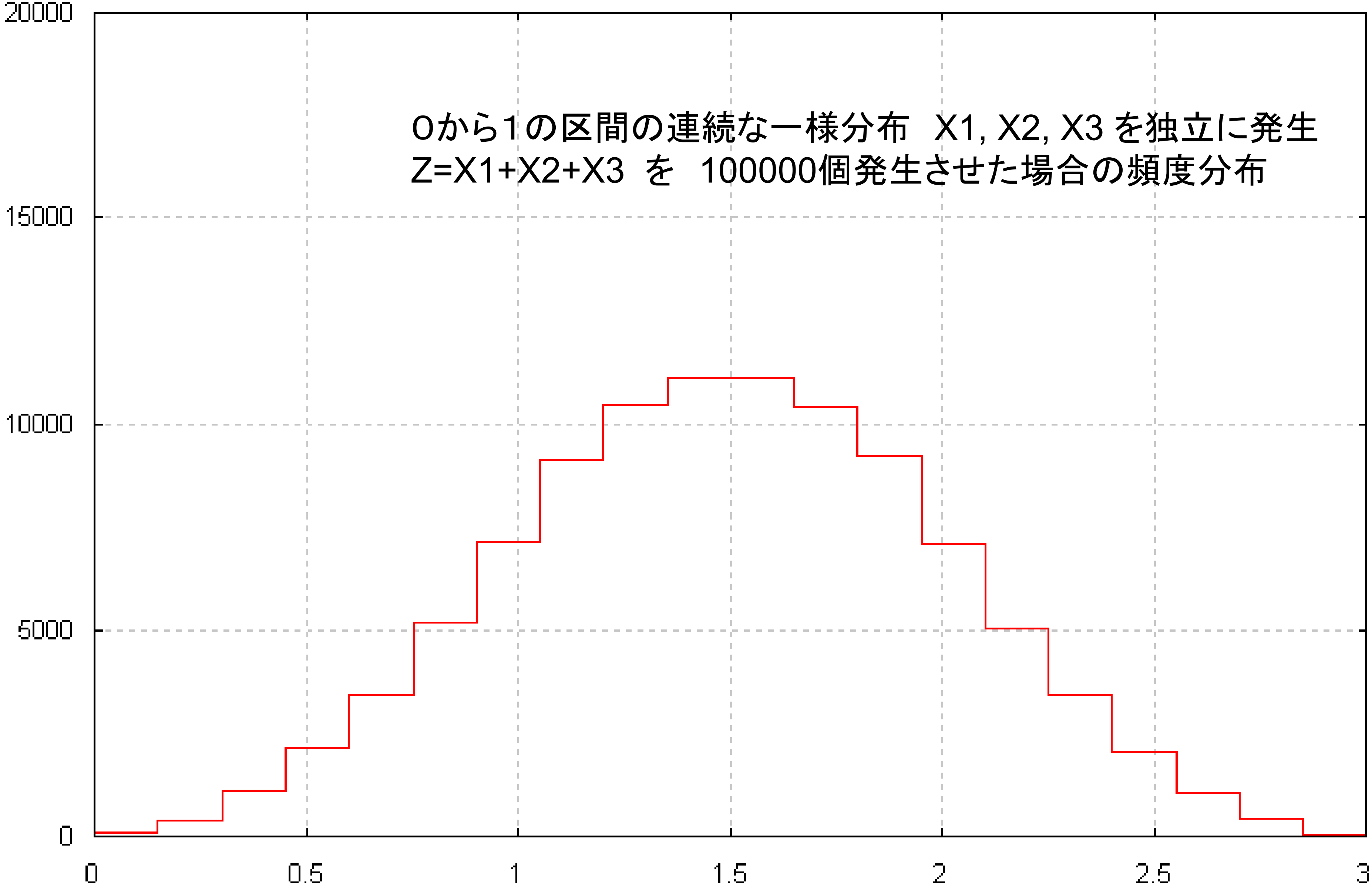
独立な確率変数の和の分布例(一様分布)



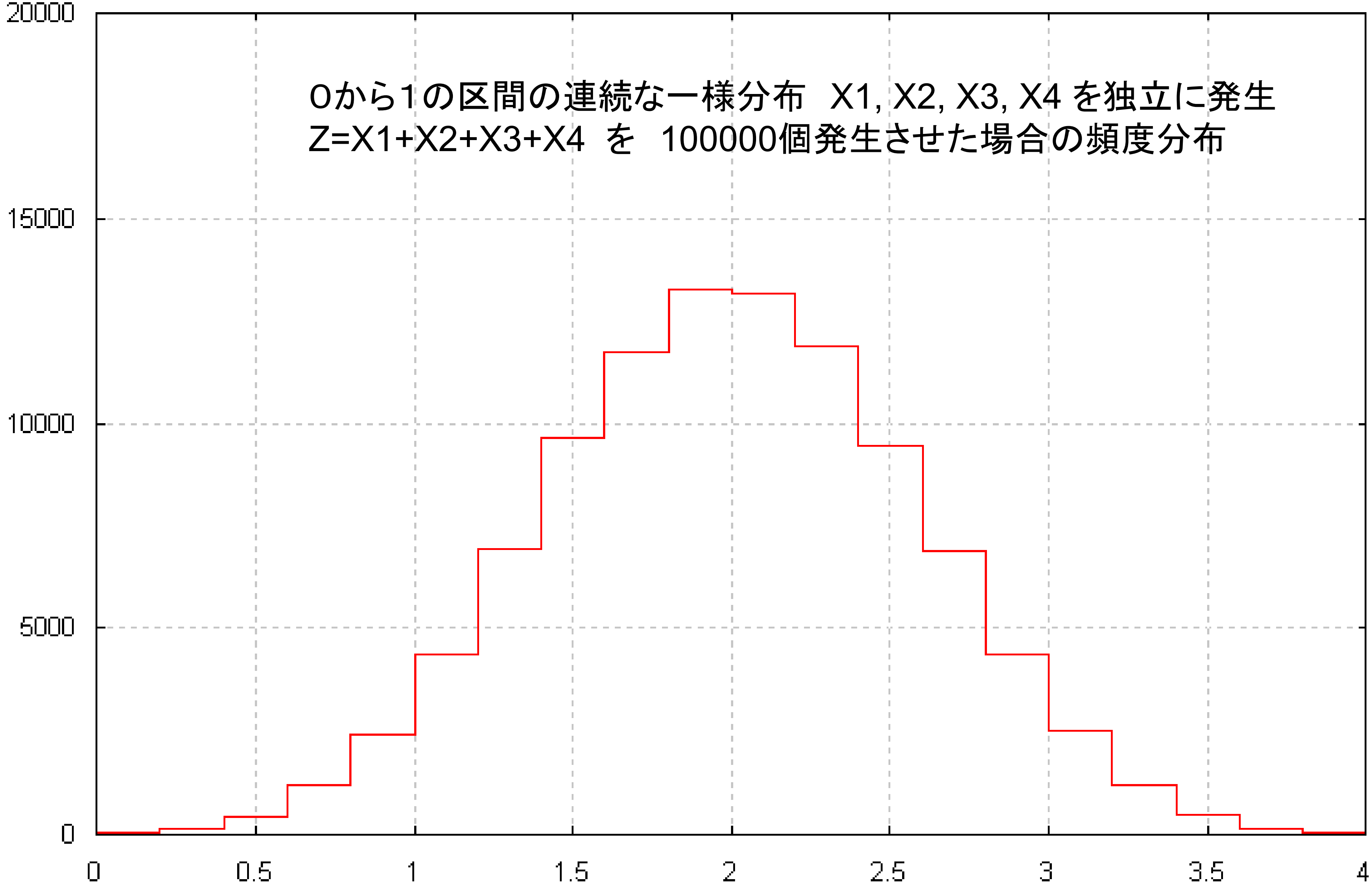
独立な確率変数の和の分布例(一様分布)



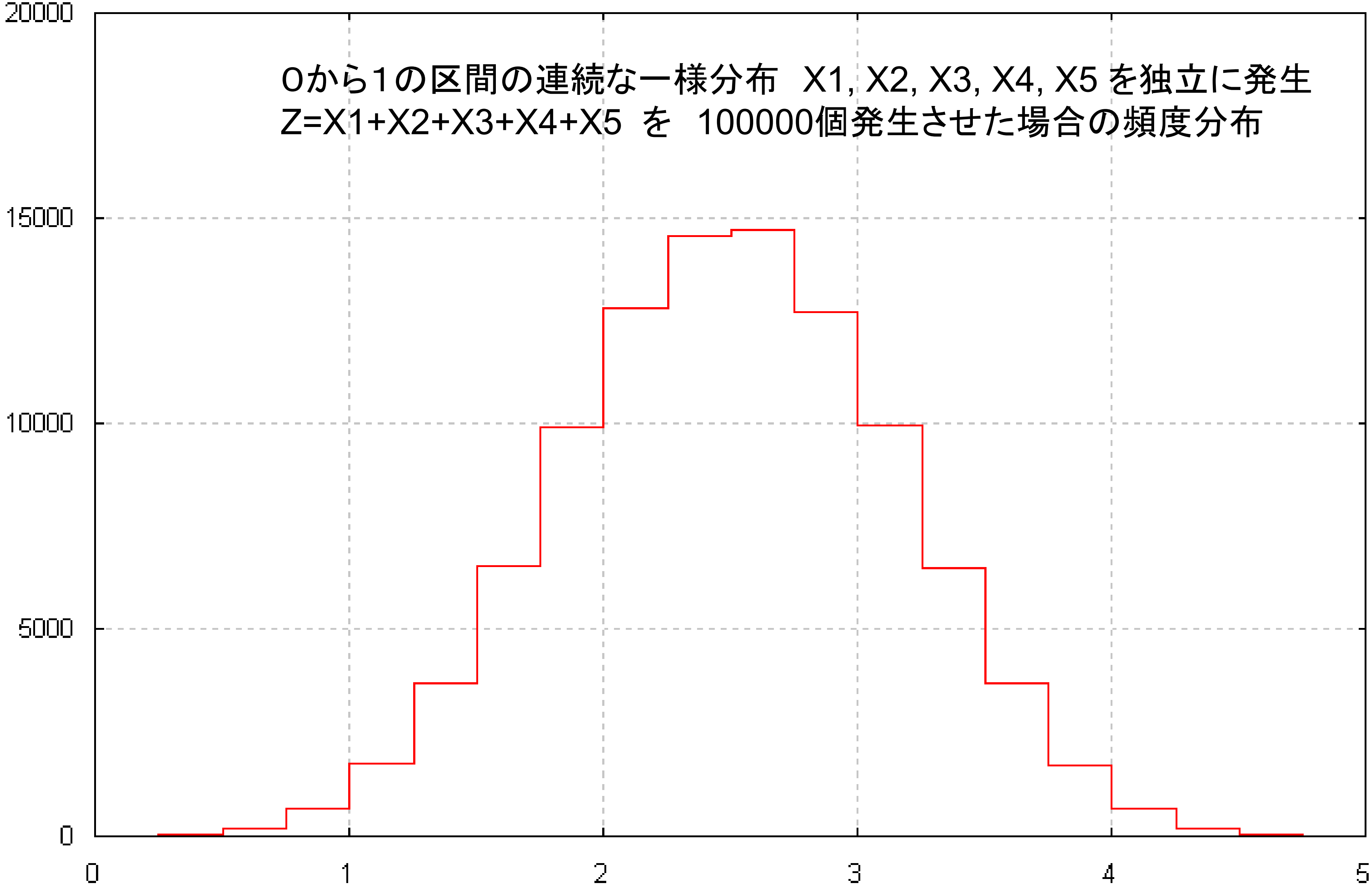
独立な確率変数の和の分布例（一様分布）



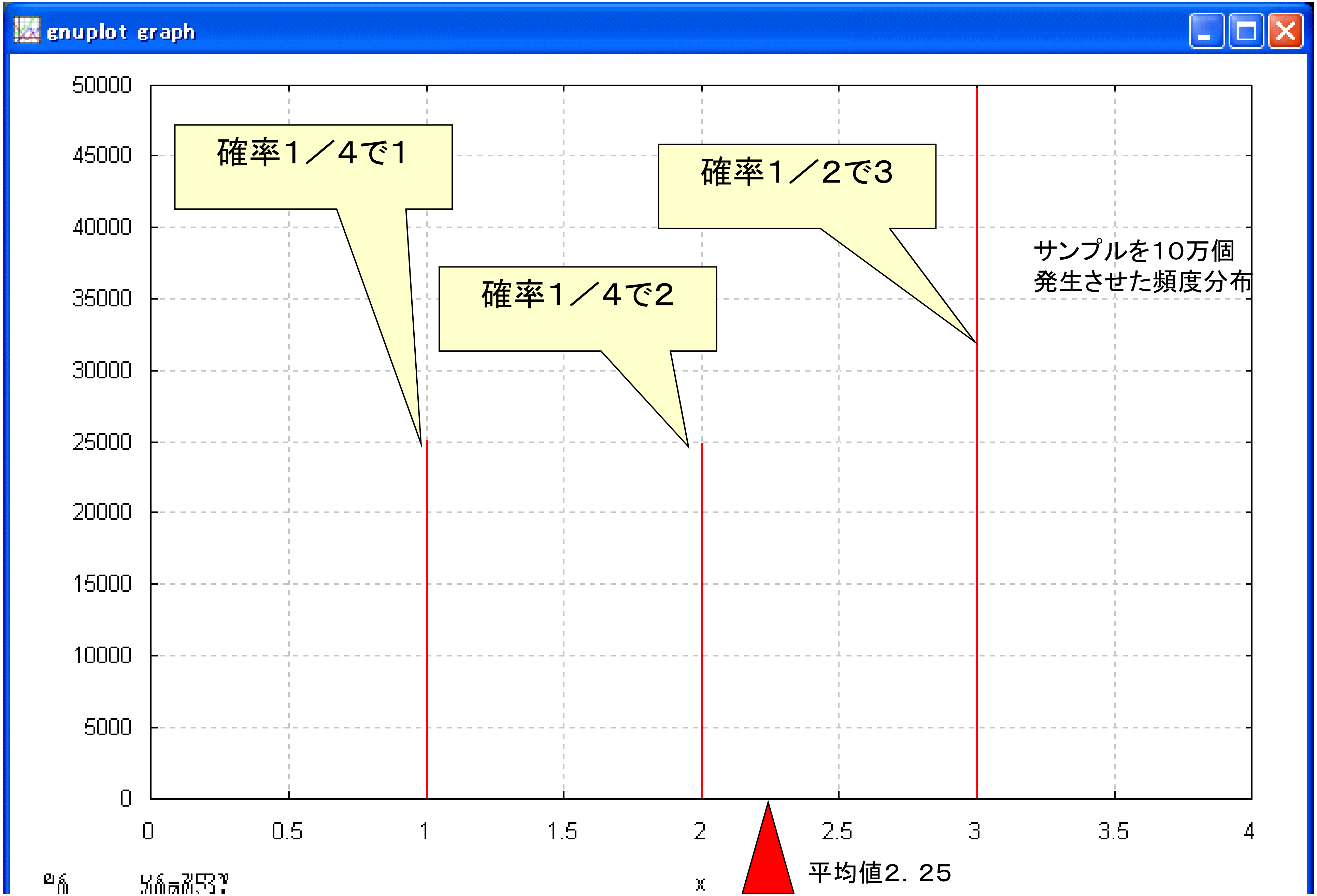
独立な確率変数の和の分布例(一様分布)



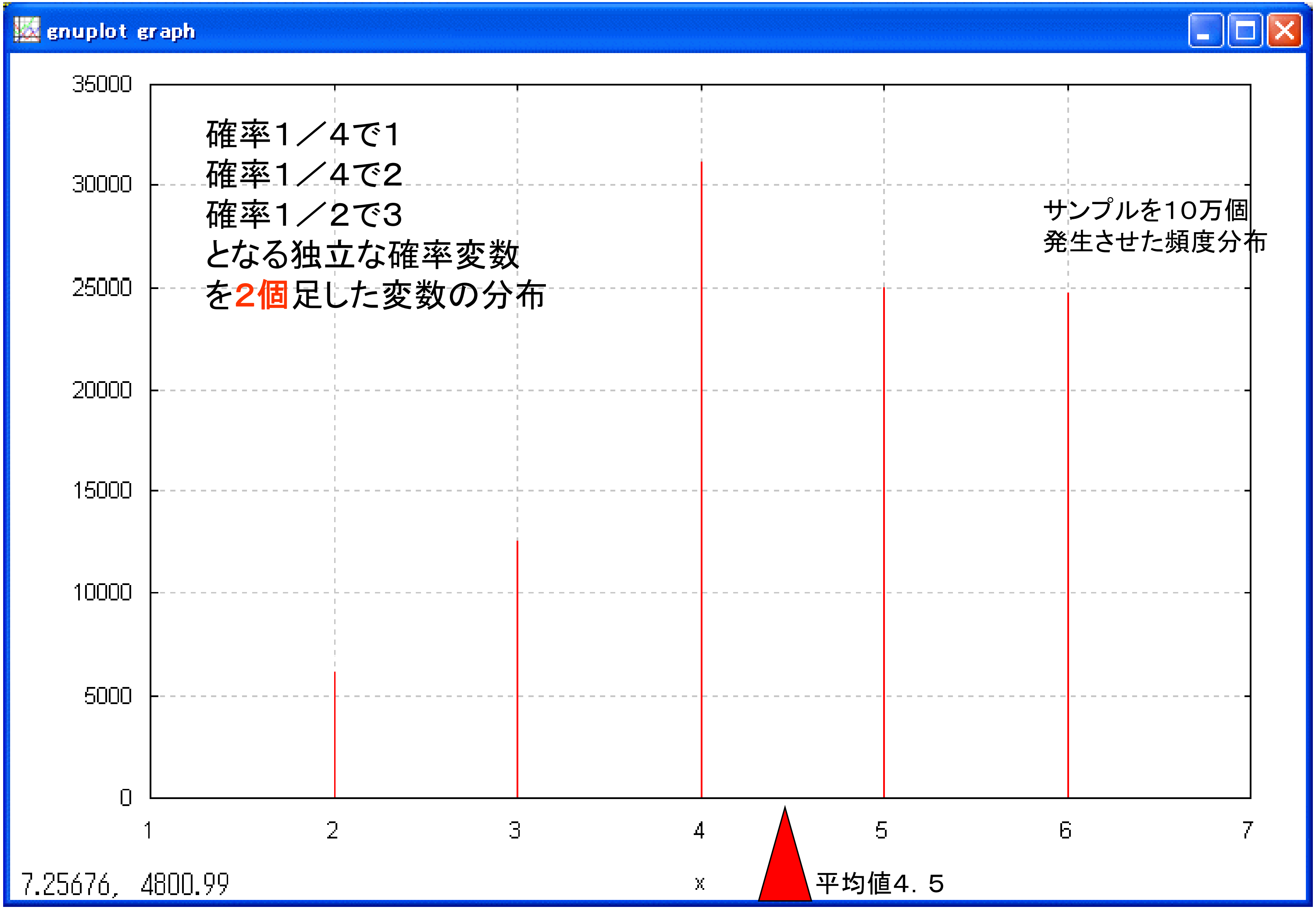
独立な確率変数の和の分布例(一様分布)



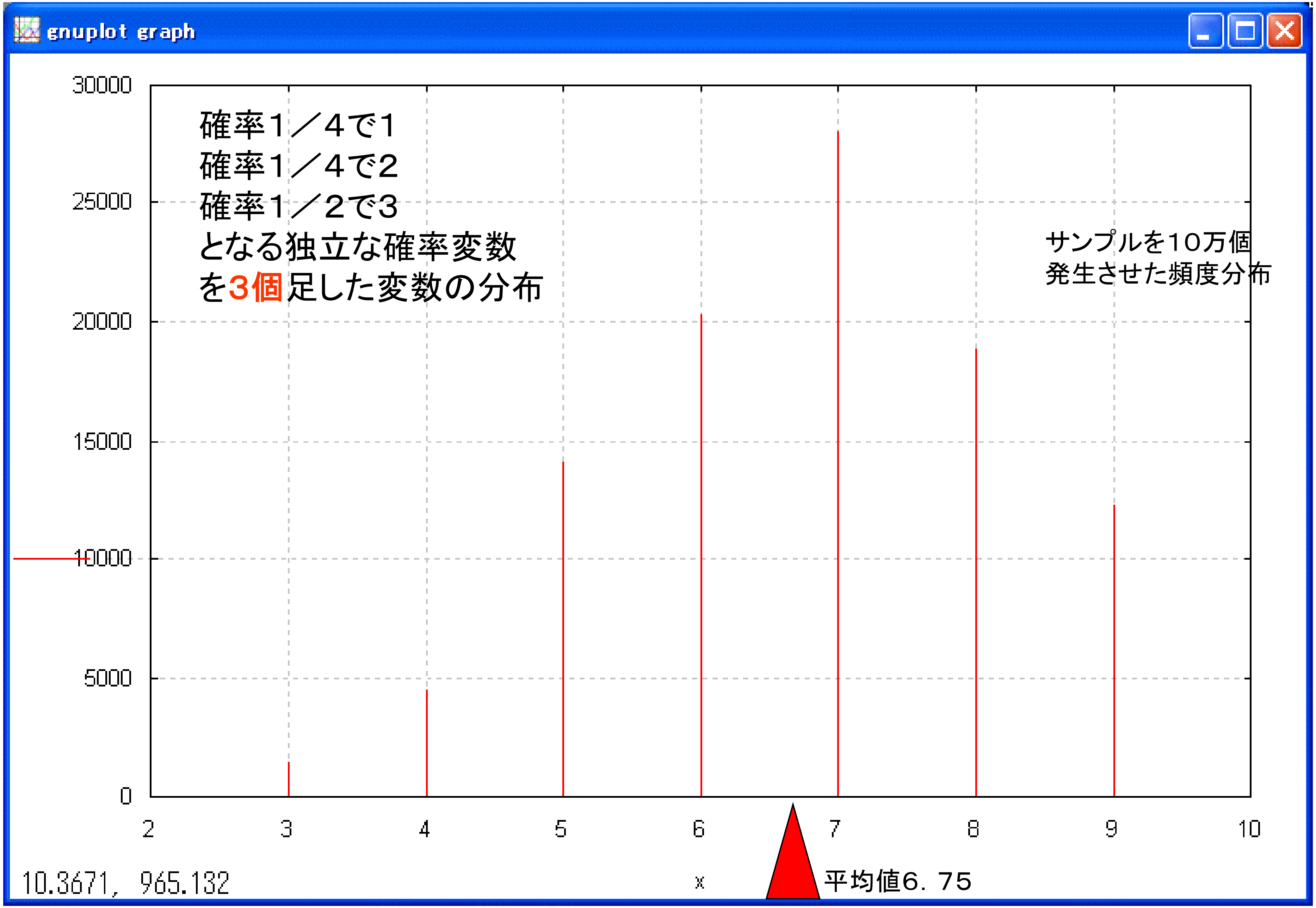
独立な確率変数の和の分布例（偏った離散分布）



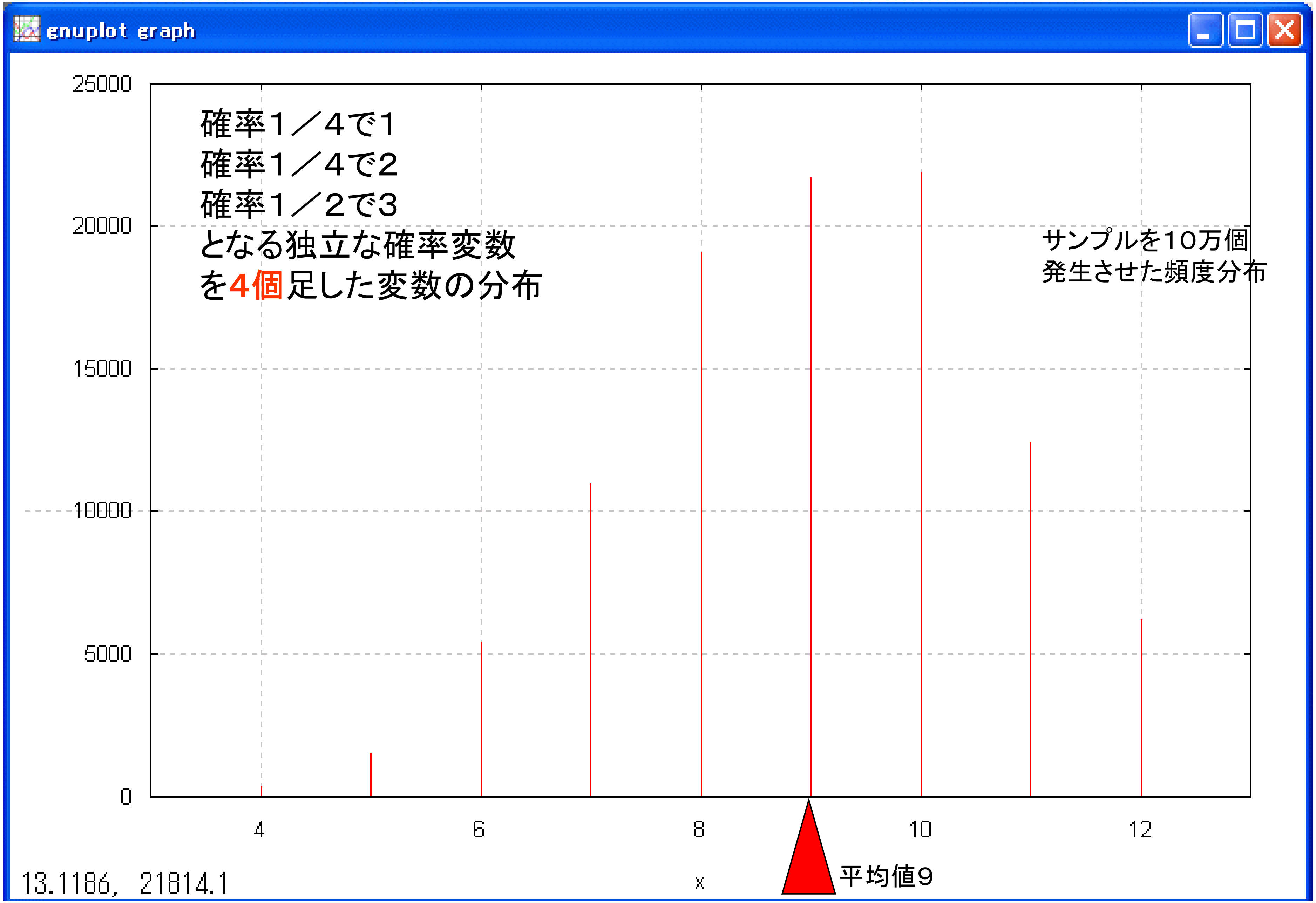
独立な確率変数の和の分布例（偏った離散分布）



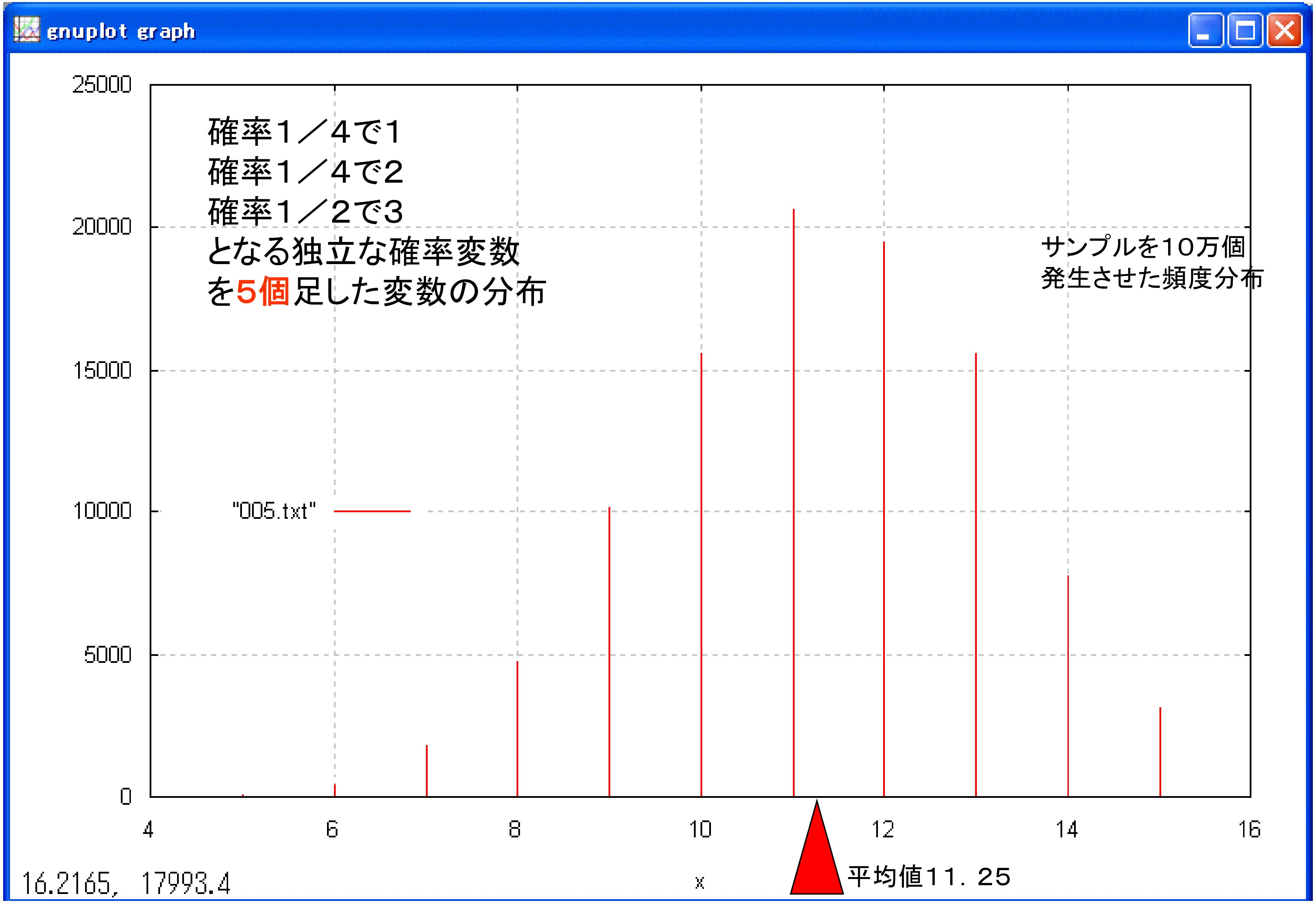
独立な確率変数の和の分布例（偏った離散分布）



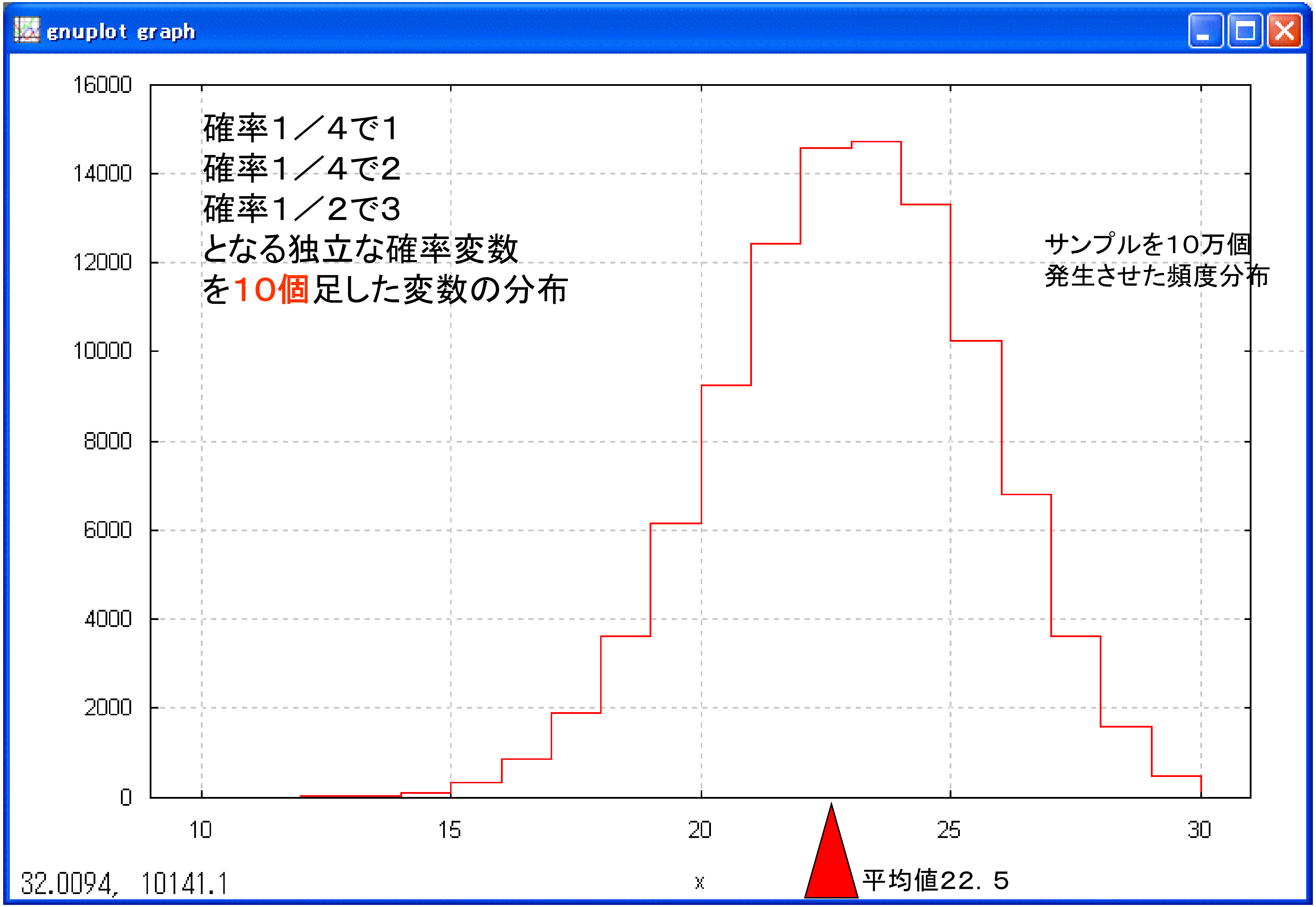
独立な確率変数の和の分布例（偏った離散分布）



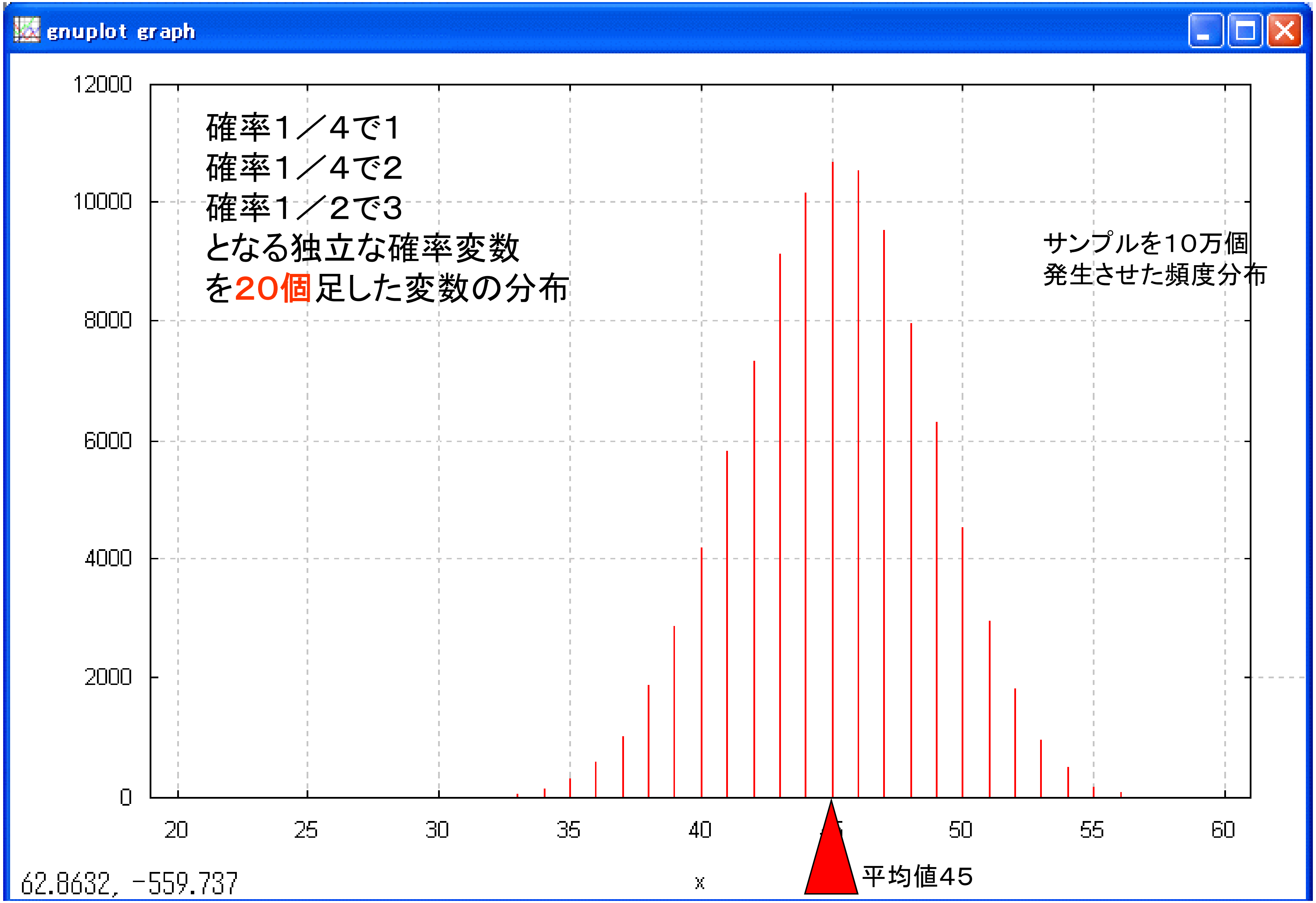
独立な確率変数の和の分布例（偏った離散分布）



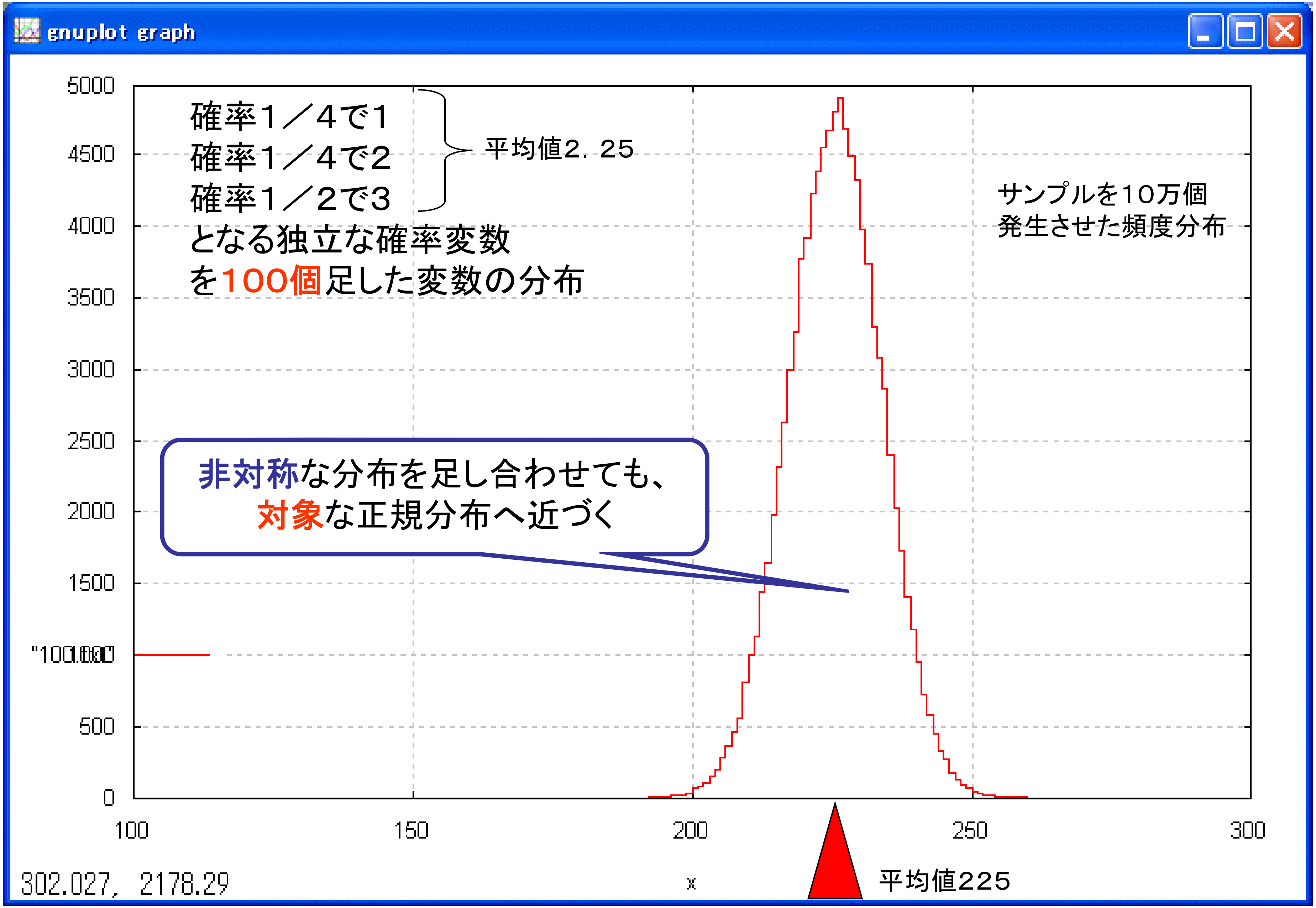
独立な確率変数の和の分布例（偏った離散分布）



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

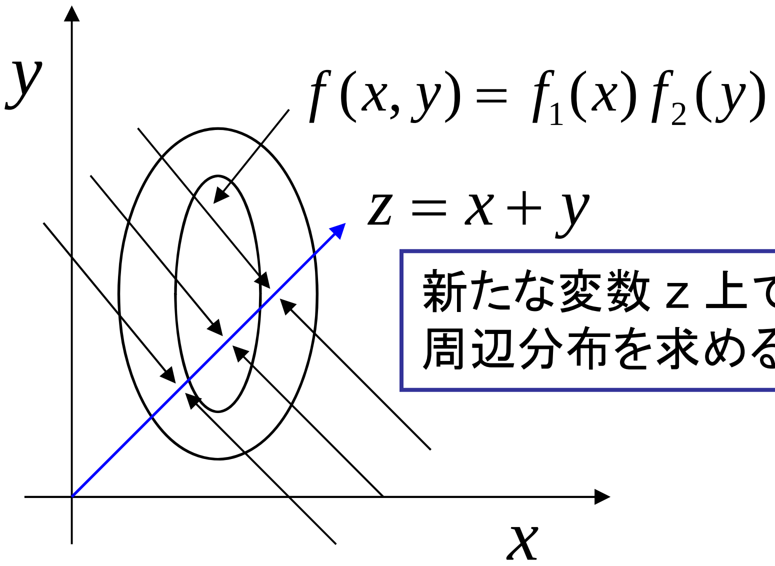
$$E\{x+y\}=E\{x\}+E\{y\}$$

→ 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

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

→ 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？



新たな変数 z 上での周辺分布を求める

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 $\begin{cases} f_1(x) = N(\mu_1, \sigma_1^2) \\ f_2(x) = N(\mu_2, \sigma_2^2) \end{cases}$

に従うとき、これらの和 $Z = X_1 + X_2$ の確率密度関数は

$$f(z) = \underbrace{\int_{-\infty}^{\infty} f_1(x) f_2(z-x)dx}_{\text{たたみこみ積分}} = \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} e^{-\frac{(z-x-\mu_1)^2}{2\sigma_1^2}-\frac{(x-\mu_2)^2}{2\sigma_2^2}}dx$$

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

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

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

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

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

$x_1, x_{2,} \cdots, x_n, \cdots$ 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 $\boxed{\text{任意の分布}}$ に従うとき、

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

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

分散

標準偏差 $\dfrac{\sigma}{\sqrt{N}}$

「1回の計測の誤差」＝±2σ

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

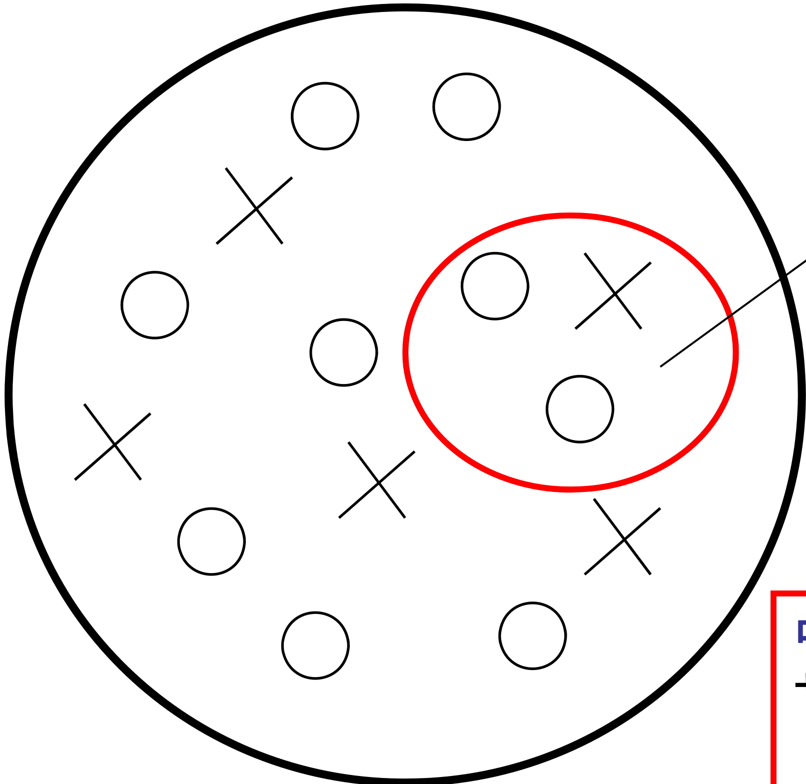
母比率の区間推定（大標本の場合）

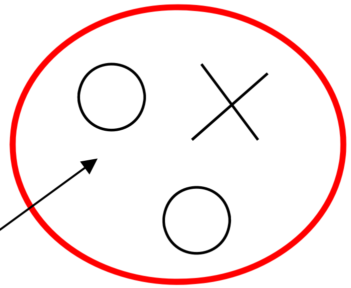
調べる対象となる全体の集団：
母集団（population）

標本（sample）
$x_1, x_2, \cdots, x_n$

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 $\boxed{\quad}$ $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ の

期待値は $\boxed{\quad}$ 分散 $\boxed{\quad}$

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 $\boxed{\quad}$ 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 $\boxed{\quad}$ 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

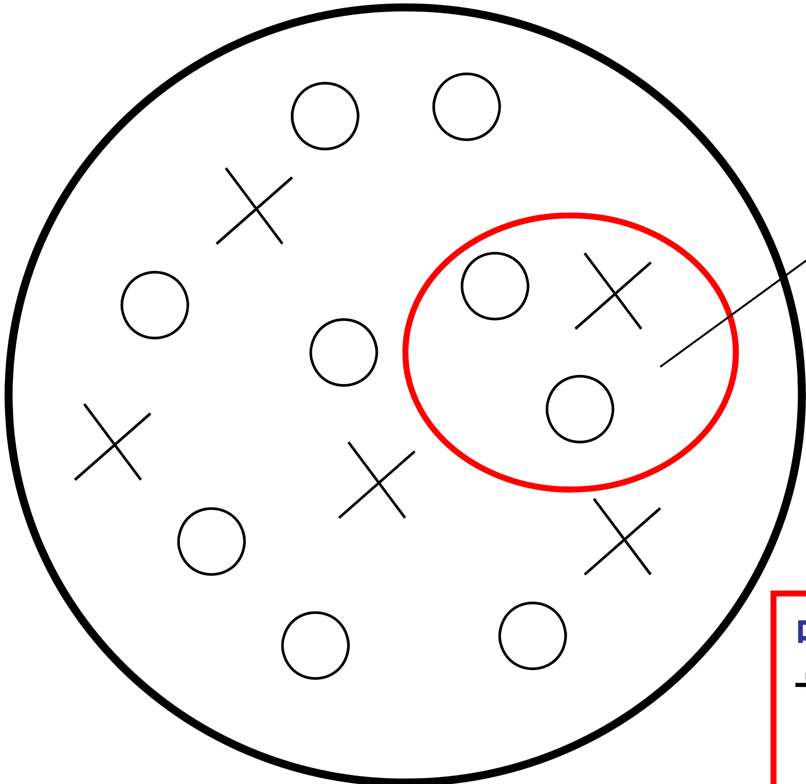
母比率の区間推定（大標本の場合）

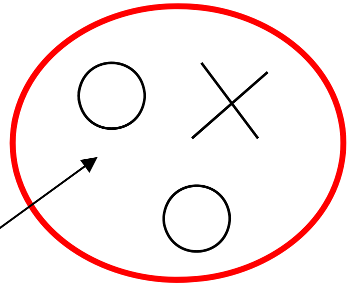
調べる対象となる全体の集団：
母集団（population）

標本（sample）
$x_1, x_2, \cdots, x_n$

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 $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ の

標本比率

期待値は $\boxed{p}$ 分散 $\boxed{\dfrac{p(1-p)}{n}}$

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 $\boxed{N(p, \dfrac{p(1-p)}{n})}$ 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 母比率 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

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

標準化

$\boxed{}$ 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 $z\left(\dfrac{\alpha}{2}\right)$ とおくと、

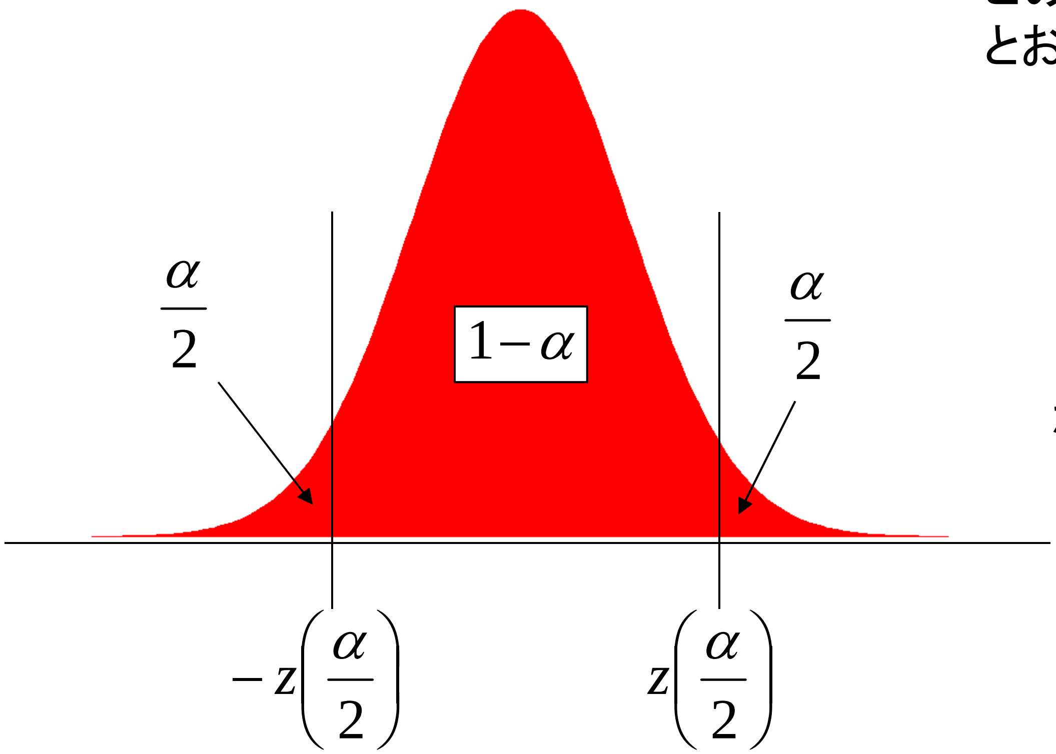
$$P\left(-z\left(\frac{\alpha}{2}\right) \le \boxed{} \le z\left(\frac{\alpha}{2}\right)\right) = 1-\alpha$$

が成り立つ。 この不等式

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

は、以下のように書き換えられる:

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

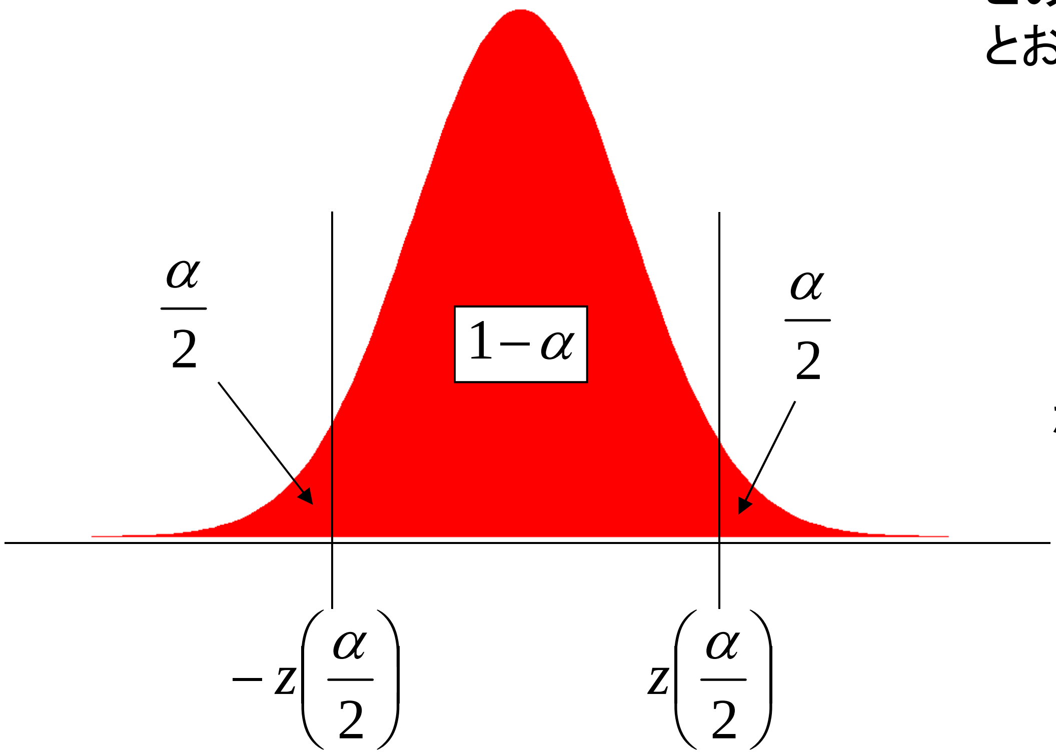


標本平均 $\bar{x}=\dfrac{x_1+x_2+\cdots+x_n}{n}$ の分布は、期待値 p 標準偏差 $\sqrt{\dfrac{p(1-p)}{n}}$ の正規分布

標準化

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

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 $z\left(\dfrac{\alpha}{2}\right)$ とおくと、

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

が成り立つ。 この不等式

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

は、以下のように書き換えられる:

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

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

この式は以下のように変形できる:

$$\le 0$$

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

$$p_1 \le p \le p_2$$

ただし

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

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

標本数 n が大きいとき、

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

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

この式は以下のように変形できる:

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

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

$$p_1 \le p \le p_2$$

ただし

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

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

標本数 n が大きいとき、

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

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

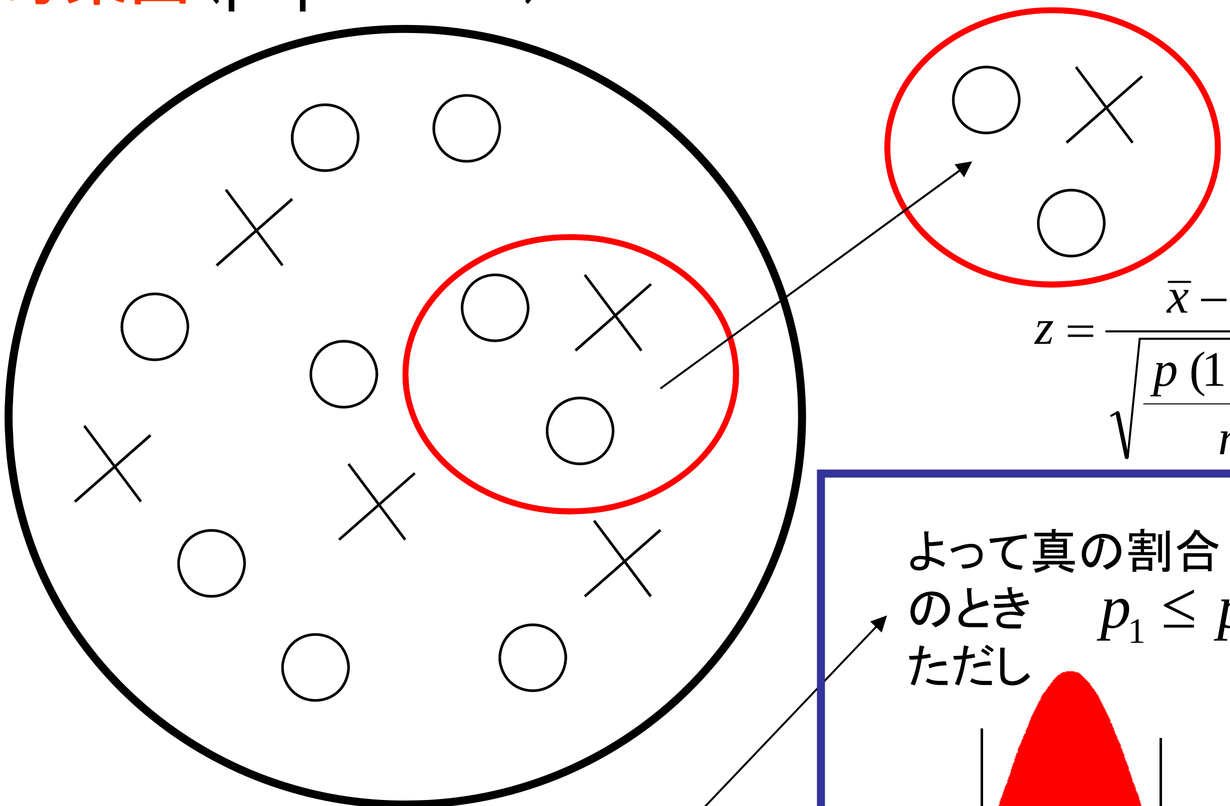
$$p_2=\bar{x}+z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

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)
$x_1, x_2, \cdots, x_n$

ただしnは大きな数



標本平均(標本比率)

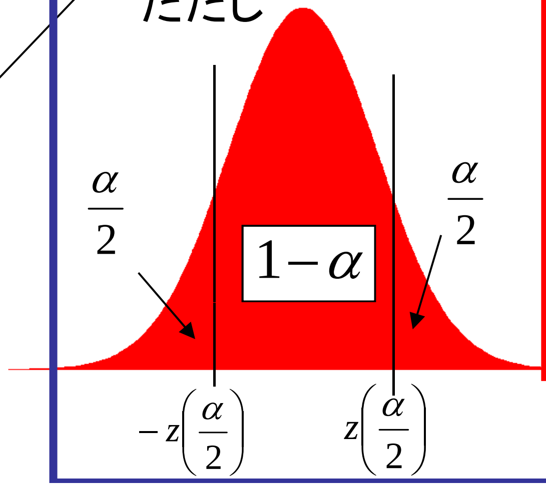
$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$z = \frac{\bar{x} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$$ 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　$p_1 \le p \le p_2$
ただし

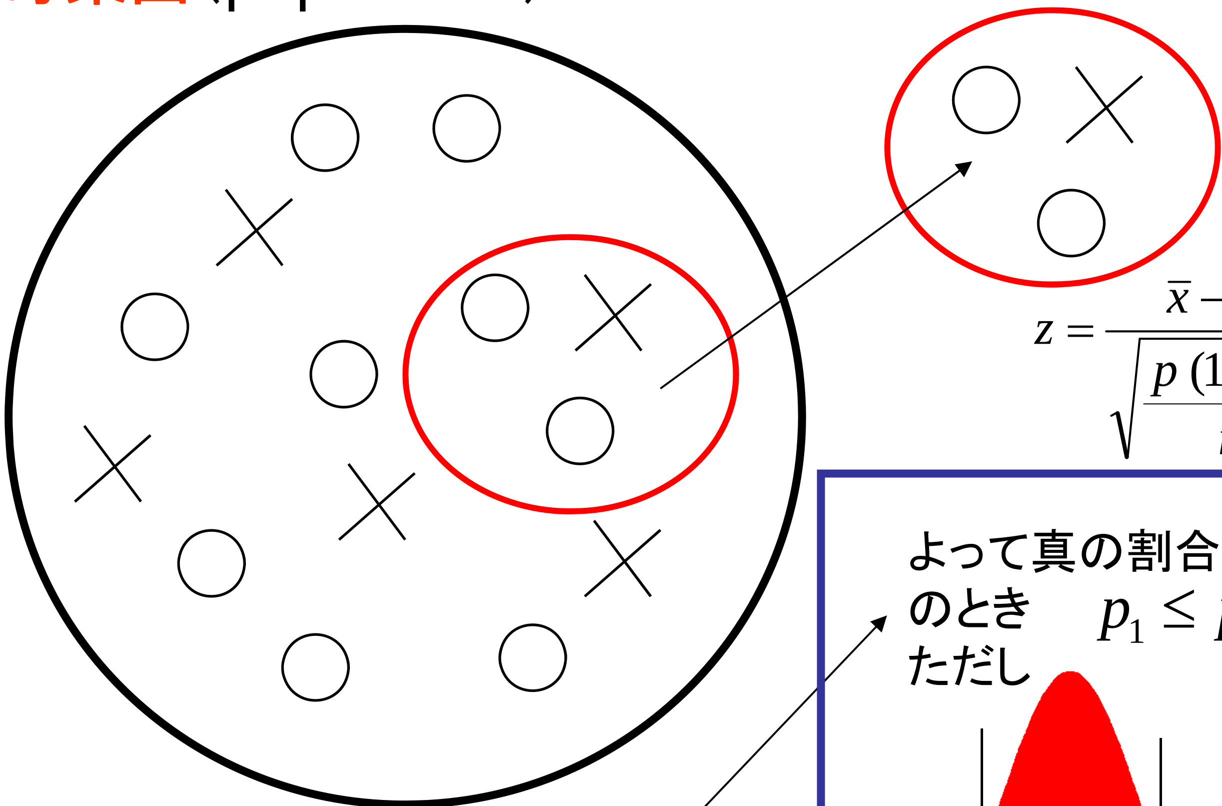


母比率の大標本区間推定(まとめ)

調べる対象となる全体の集団:
母集団(population)

標本(sample)
$x_1, x_2, \cdots, x_n$

ただしnは大きな数



標本平均(標本比率)

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

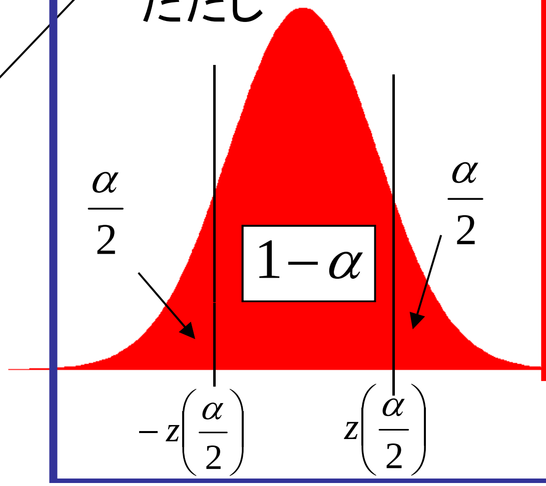
$$z = \frac{\bar{x} - p}{\sqrt{\dfrac{p\,(1-p)}{n}}}$$

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 $p_1 \le p \le p_2$
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。

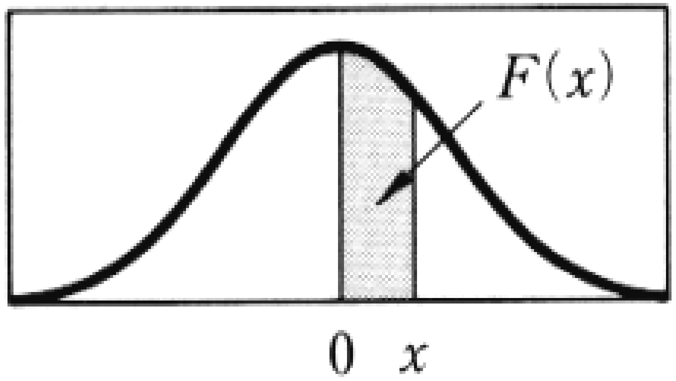
【練習問題】

（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。

$$\bar{x} = \frac{500}{2500} = \frac{1}{5}$$

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

正規分布表



カゲの部分の確率 $F(x)$ を示す。

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998

【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。

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

【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。

$$\bar{x} = \frac{500}{2500} = \frac{1}{5}$$

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

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

【演習問題】　2017.05.09　学籍番号　　　　　氏名

［問］　母比率 p を標本比率 $\bar{x}=\dfrac{x_1+x_2+\cdots+x_n}{n}$ 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 $\bar{x}(1-\bar{x})\le 0.25$ 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 $\overline{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n}$ から信頼係数95％で推定したい。

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

$\overline{x}(1-\overline{x})$ は0と1を通り上に凸な2次関数で、$\overline{x} = \dfrac{1}{2}$ 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、$$2 \times z\left(\frac{\alpha}{2}\right)\sqrt{\frac{\overline{x}(1-\overline{x})}{n}} \leq 0.04$$

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 $z\left(\dfrac{\alpha}{2}\right) = 1.96$