比

連続的な量が分子にくる「割算」の場合

2013.3

粕谷英一　（九大・理・生物）

『実証研究者は比を愛する一統計学者は比をひどく嫌う』

『比の誤用に対する警告は無視されることが多い』

Jasienski &Bazzaz(1999,Oikos)

連続量が分子

モデルケース

正規分布

$$\boxed{\frac{\text{正規分布する変数}}{\text{正規分布する変数}}}$$

$$\boxed{\dfrac{\text{正規分布する変数}}{\text{正規分布する変数}}}$$ **比**

比を新たな変数
として

（母平均／母平均）
が知りたい

バイオアッセイ

例．ある薬は別の薬の何倍効くか

$$\boxed{\dfrac{\text{正規分布する変数}}{\text{正規分布する変数}}}$$ **比**

（母平均／母平均）
が知りたい

比を新たな変数
として

差でなく比

Fiellerの信頼区間

片方が無限大になったり
離れた区間が信頼区間になったり
中間だけ信頼区間から外れたり

(Rのパッケージmratios)

$$\boxed{\dfrac{\text{正規分布する変数}}{\text{正規分布する変数}}}$$ **比**

比を新たな変数として

（母平均／母平均）が知りたい

バイオアッセイ

例．ある薬は別の薬の何倍効くか

モデルケース　連続量　正規分布

独立
（相関なし）

$$\frac{\text{標準正規分布する変数}}{\text{標準正規分布する変数}}$$

比　標準コーシー分布

コーシー分布

Cauchy分布

コーシー

確率密度関数 $\dfrac{\gamma}{\pi\{(y-y_0)^2+\gamma^2\}}$

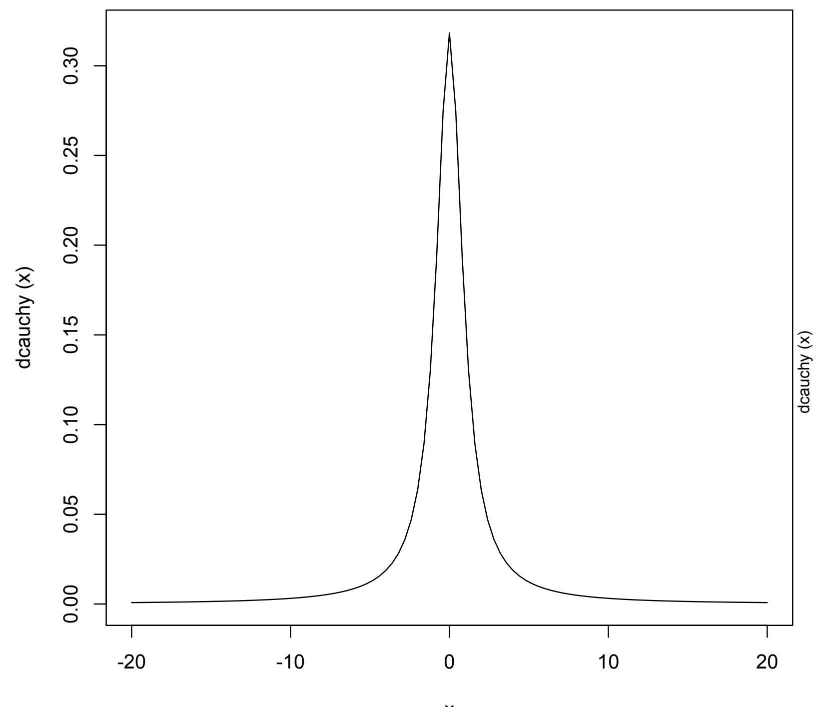
y_0 位置

γ 横の広がり

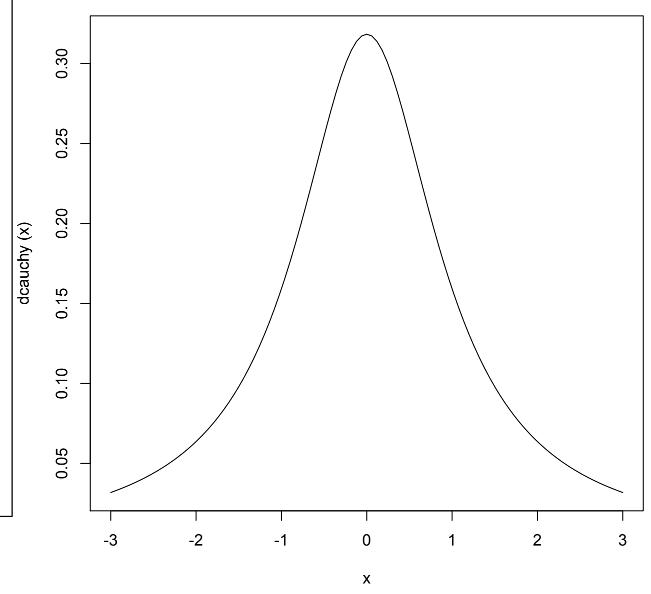
Lorenz分布
Breit-Wigner分布

標準コーシー 分布
確率密度関数 $\dfrac{1}{\pi\{y^2+1\}}$

標準コーシー 分布



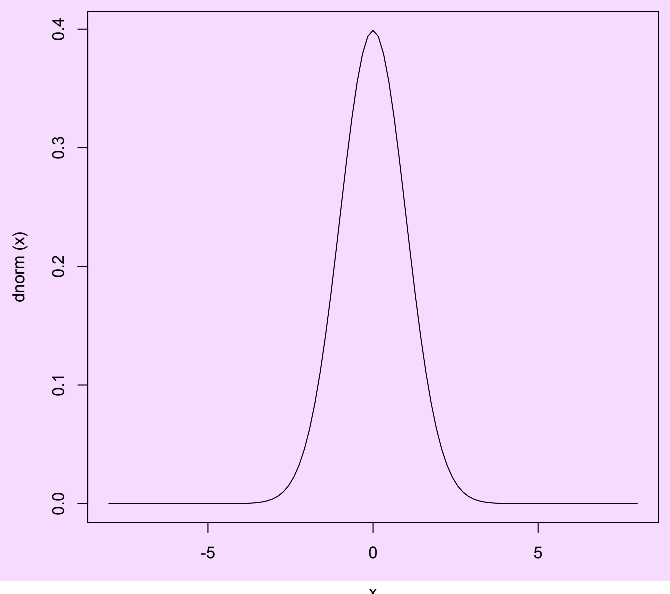
拡大図



比較：標準正規分布

$$\frac{1}{\sqrt{(2\pi)}}\, e^{\frac{-y^2}{2}}$$

確率密度関数



コーシー分布

モーメント

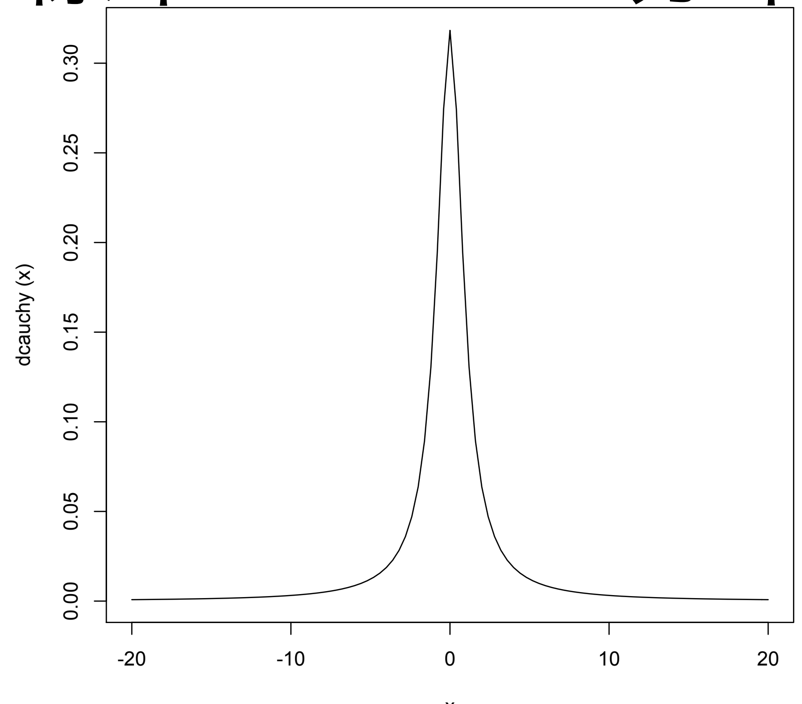
モーメント 変数のk乗の期待値→k次のモーメント

期待値（平均）	1次のモーメント	μ	λ
分散	２次のモーメントより	σ^2	λ
歪度	（非対称性）３次のモーメントより	0	$1/\sqrt{\lambda}$
尖度	（裾の重さ）４次のモーメントより	3	$1/\lambda+3$

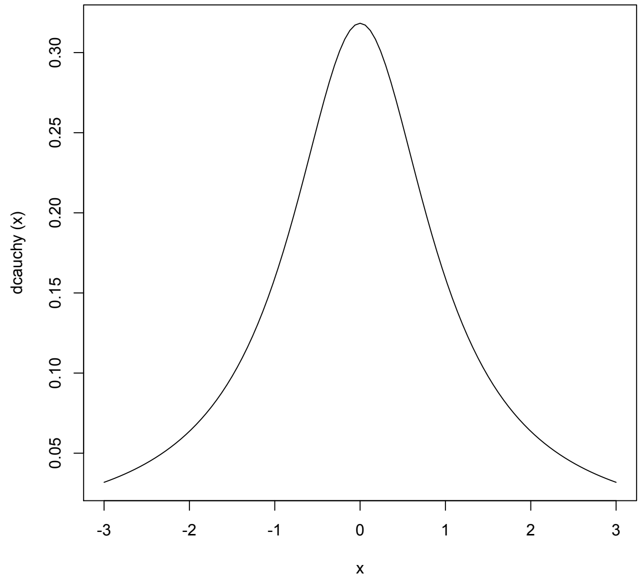
正規分布

ポアソン分布

標準コーシー 分布



拡大図



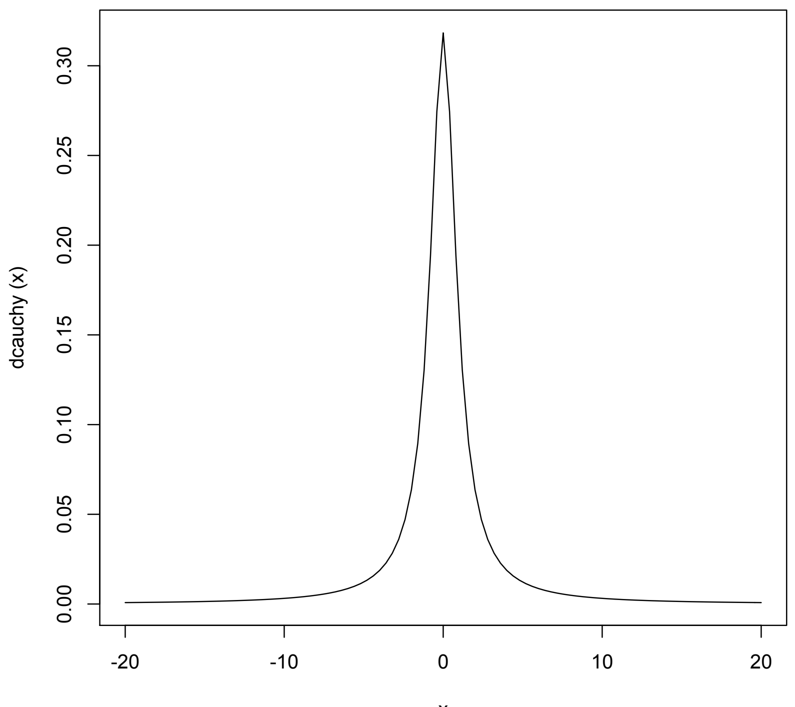
コーシー分布　　　モーメント　存在しない

期待値（平均）　存在しない

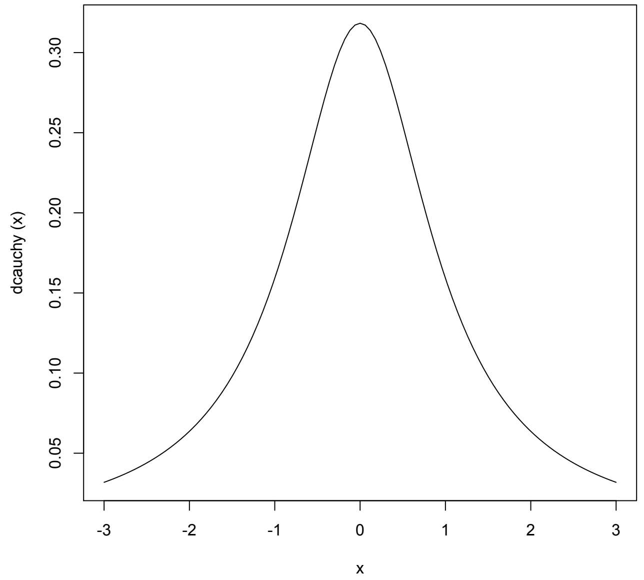
分散（正確には原点まわりの2次モーメント）　無限大

・・・

標準コーシー 分布

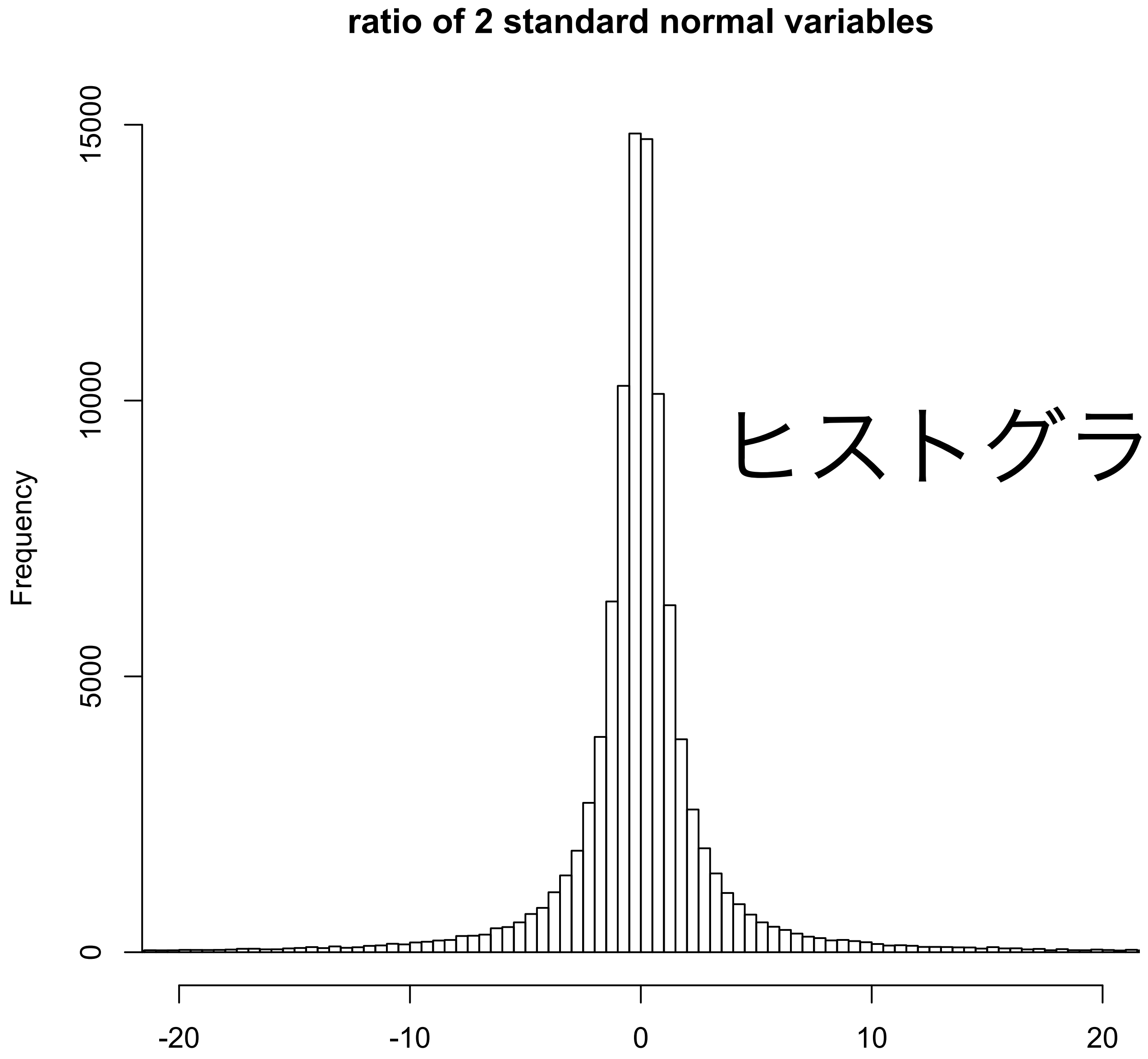


拡大図



2つの標準正規分布する変数の比　　乱数

（平均0で分散1の正規分布）

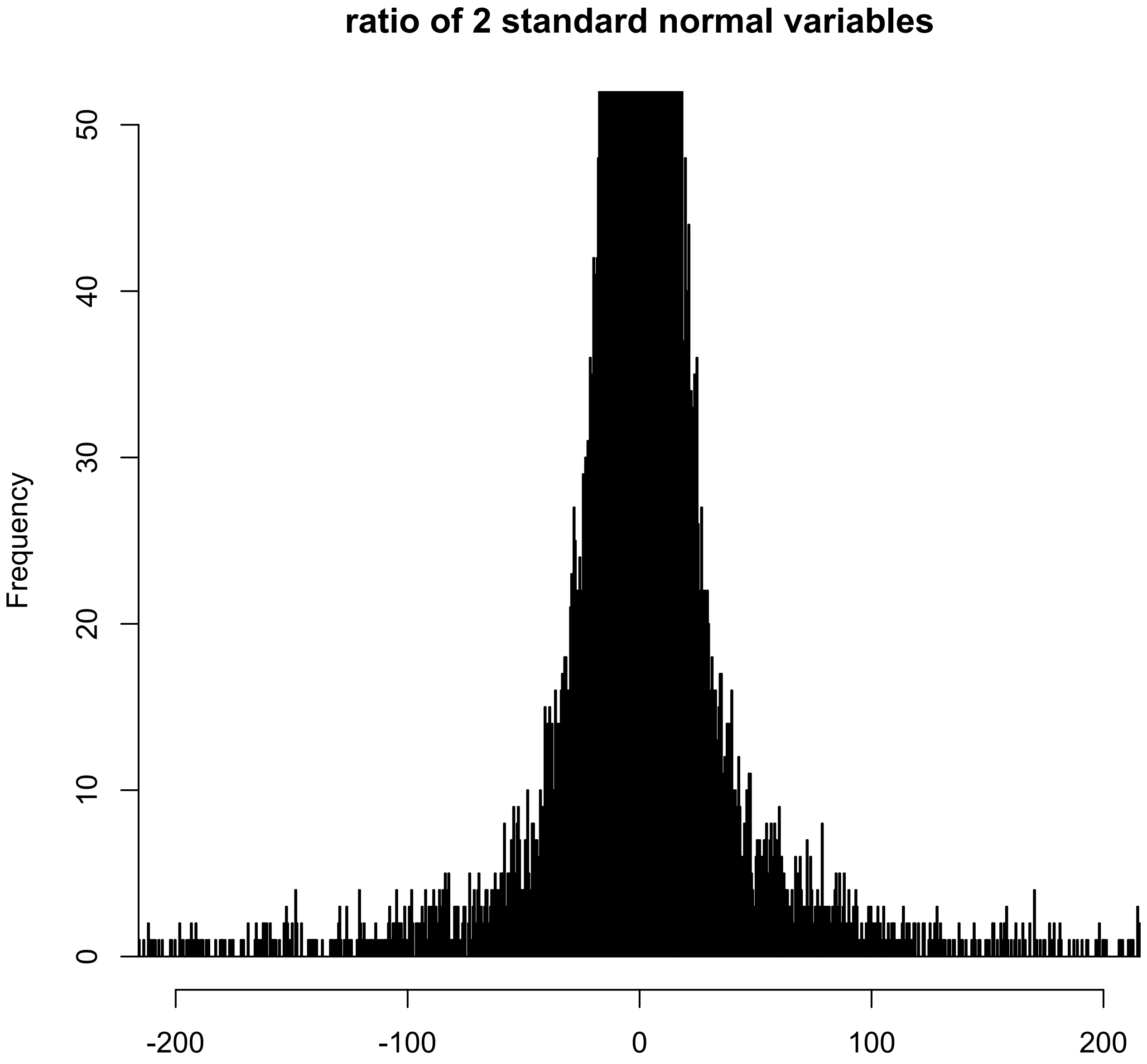


ヒストグラム

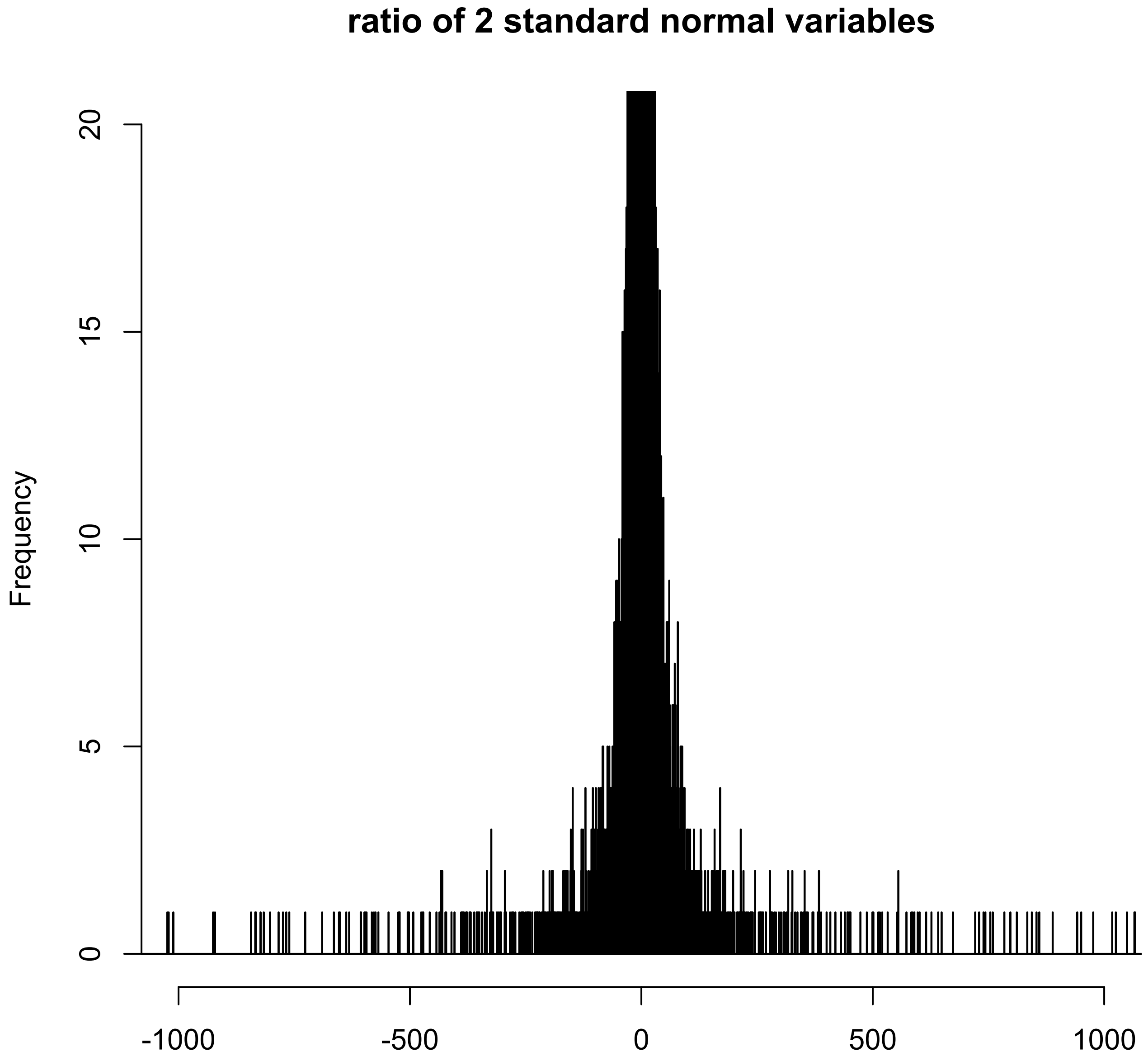
100000個

-10　　+10

拡大その１　　ヒストグラム　　拡大その2



-100　　+100

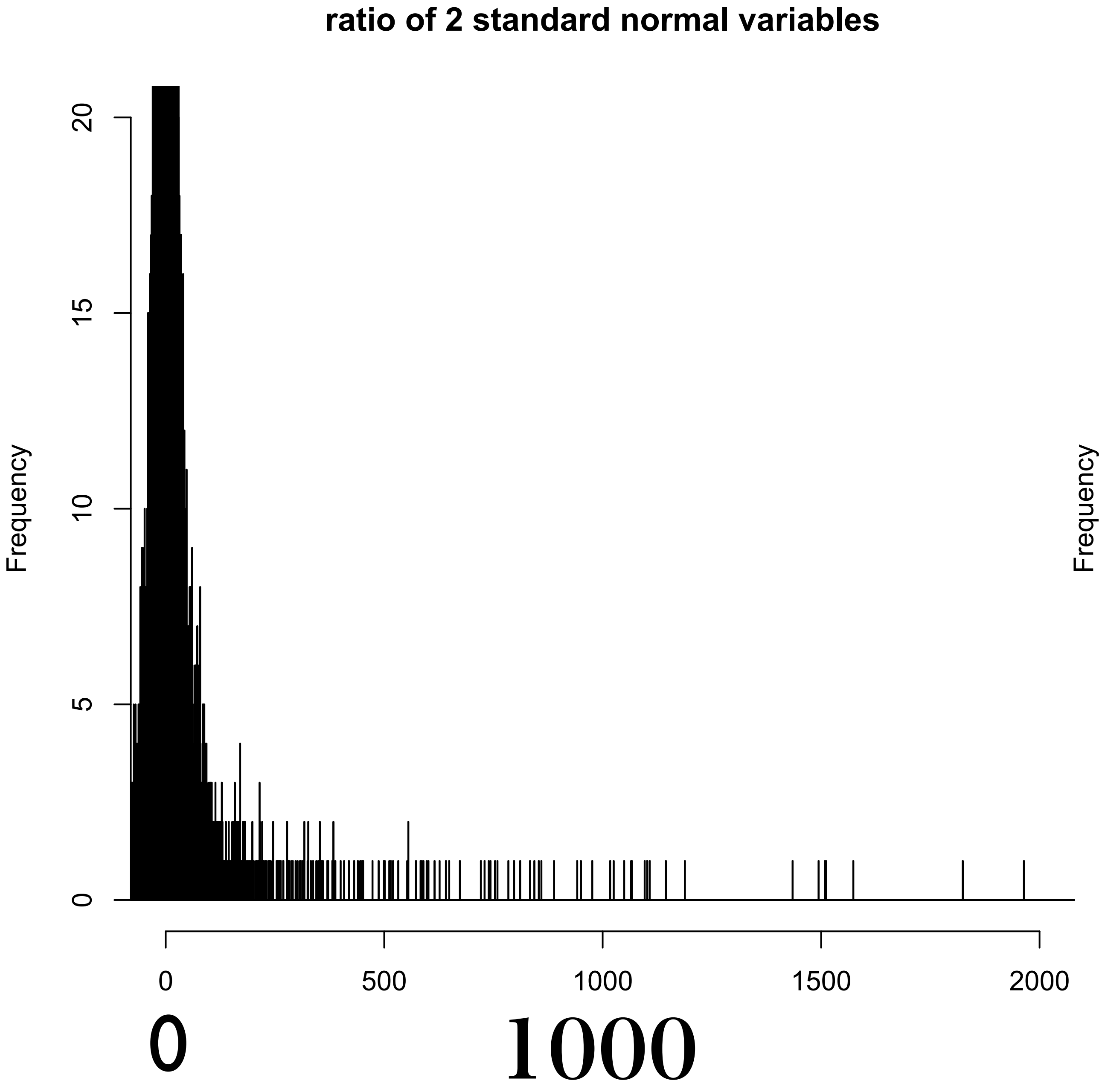


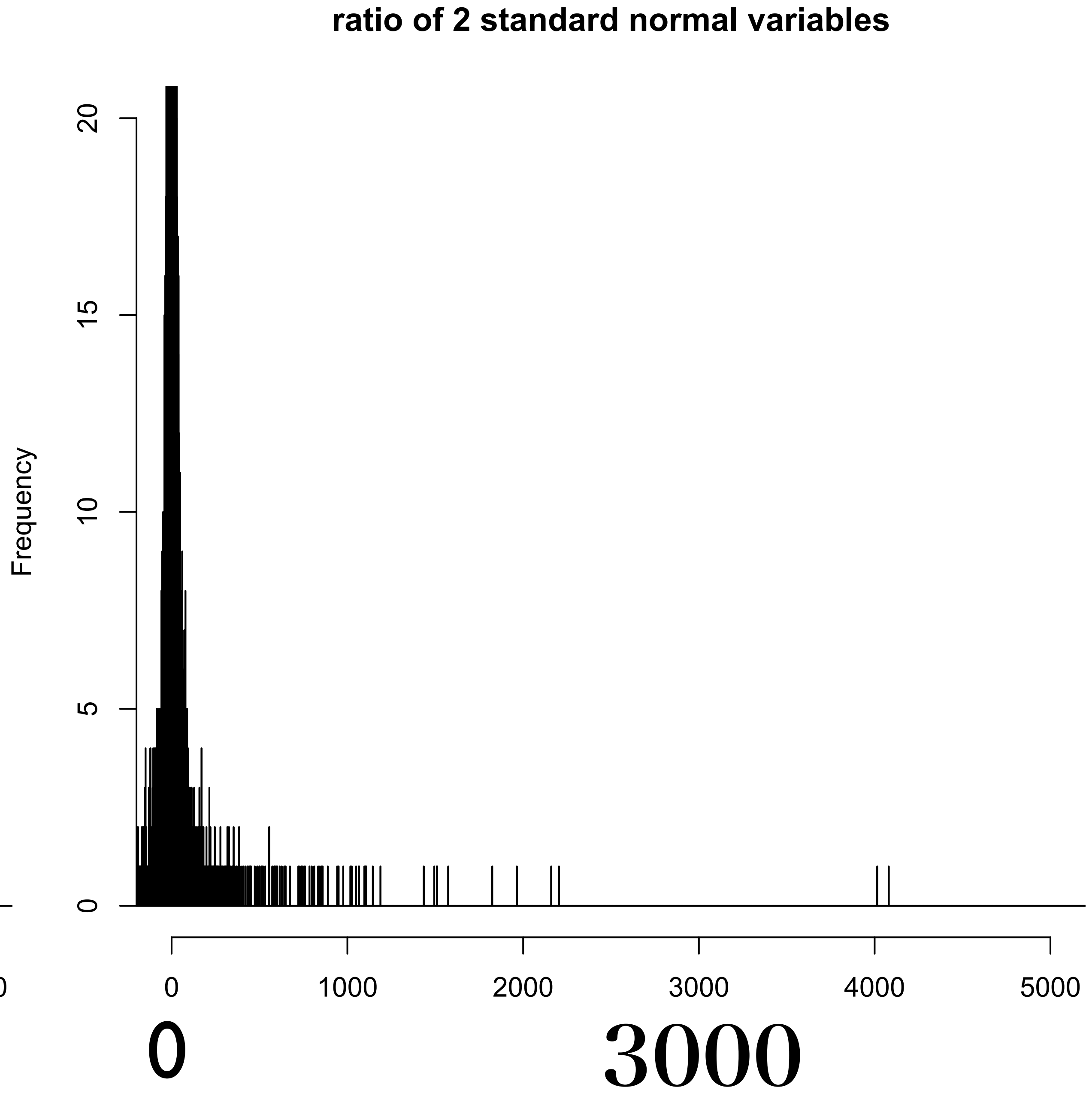
-500　　+500

100000個

右半分の拡大

ヒストグラム

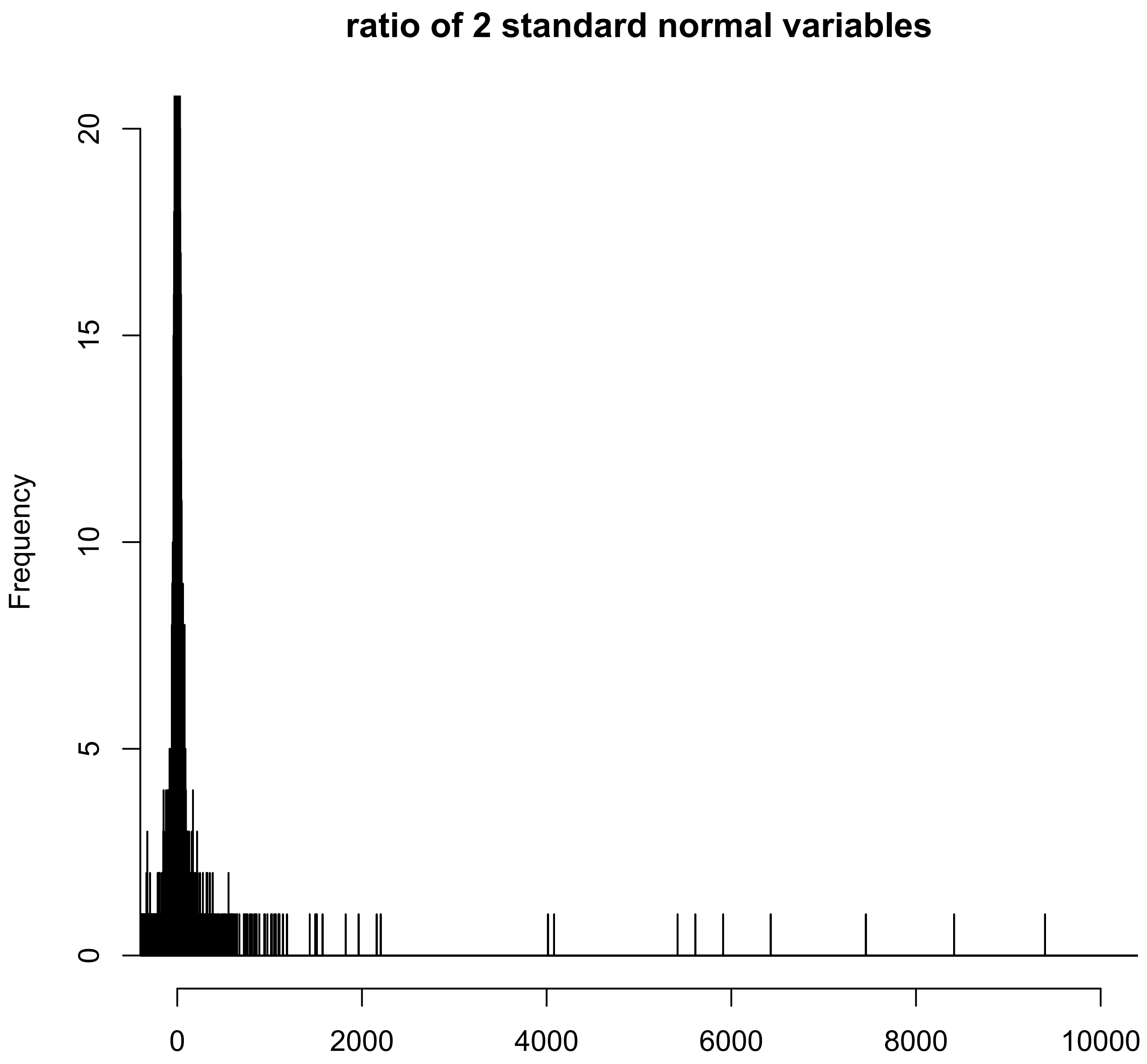


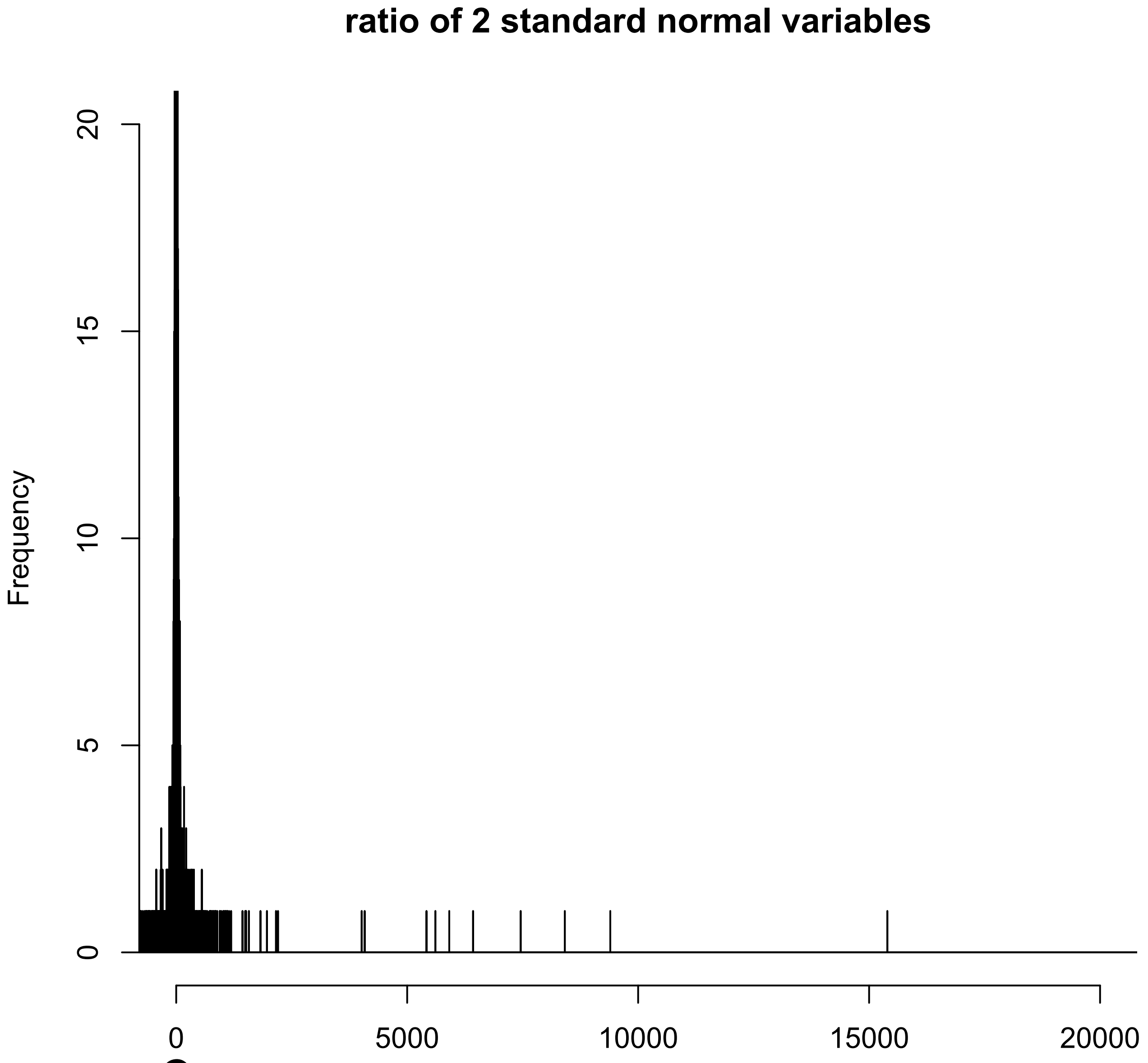


100000個

右半分のさらに拡大

ヒストグラム





0 10000

0 20000

100000個

2つの標準正規分布する変数の比

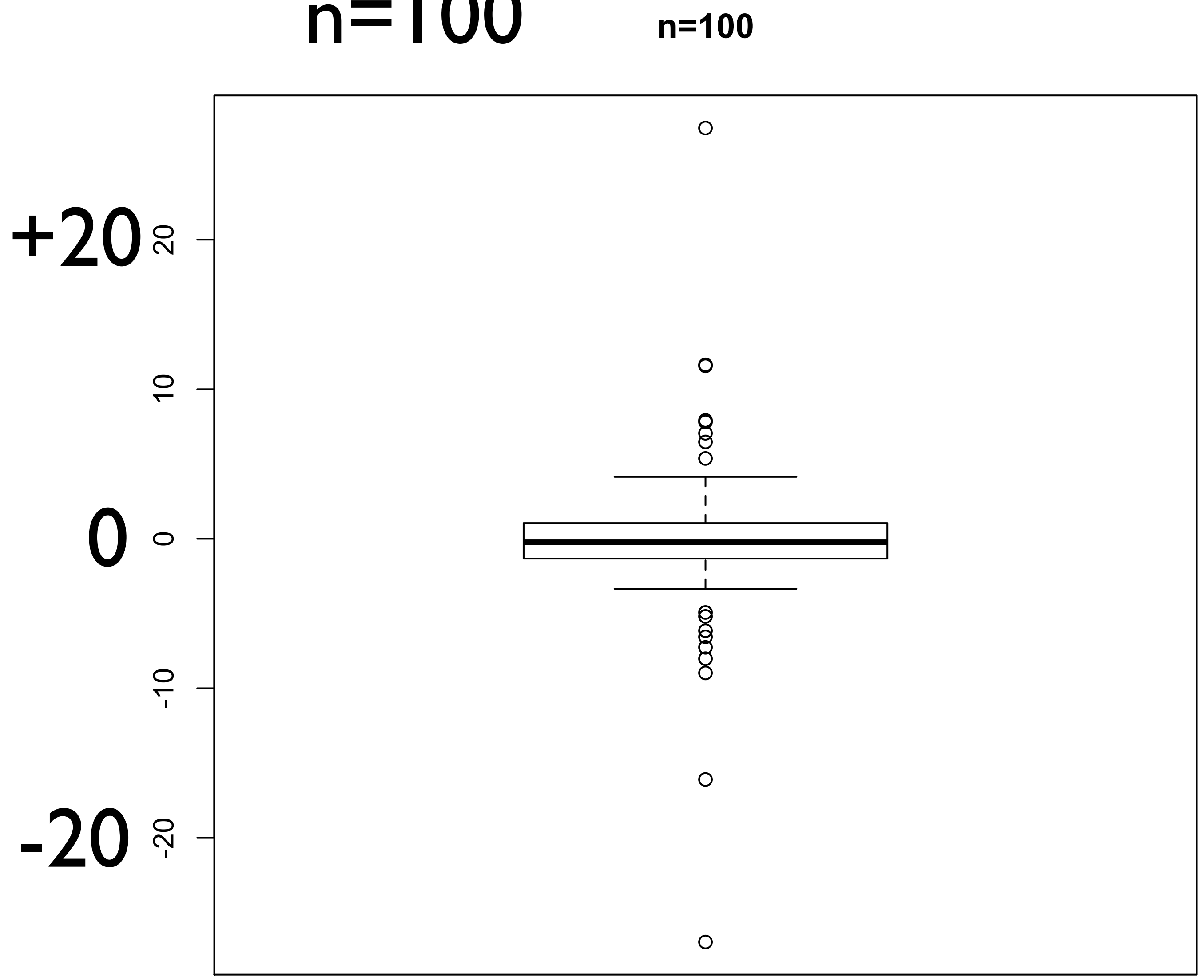
標準コーシー分布

分布のまん中から遠く離れた値が出やすい

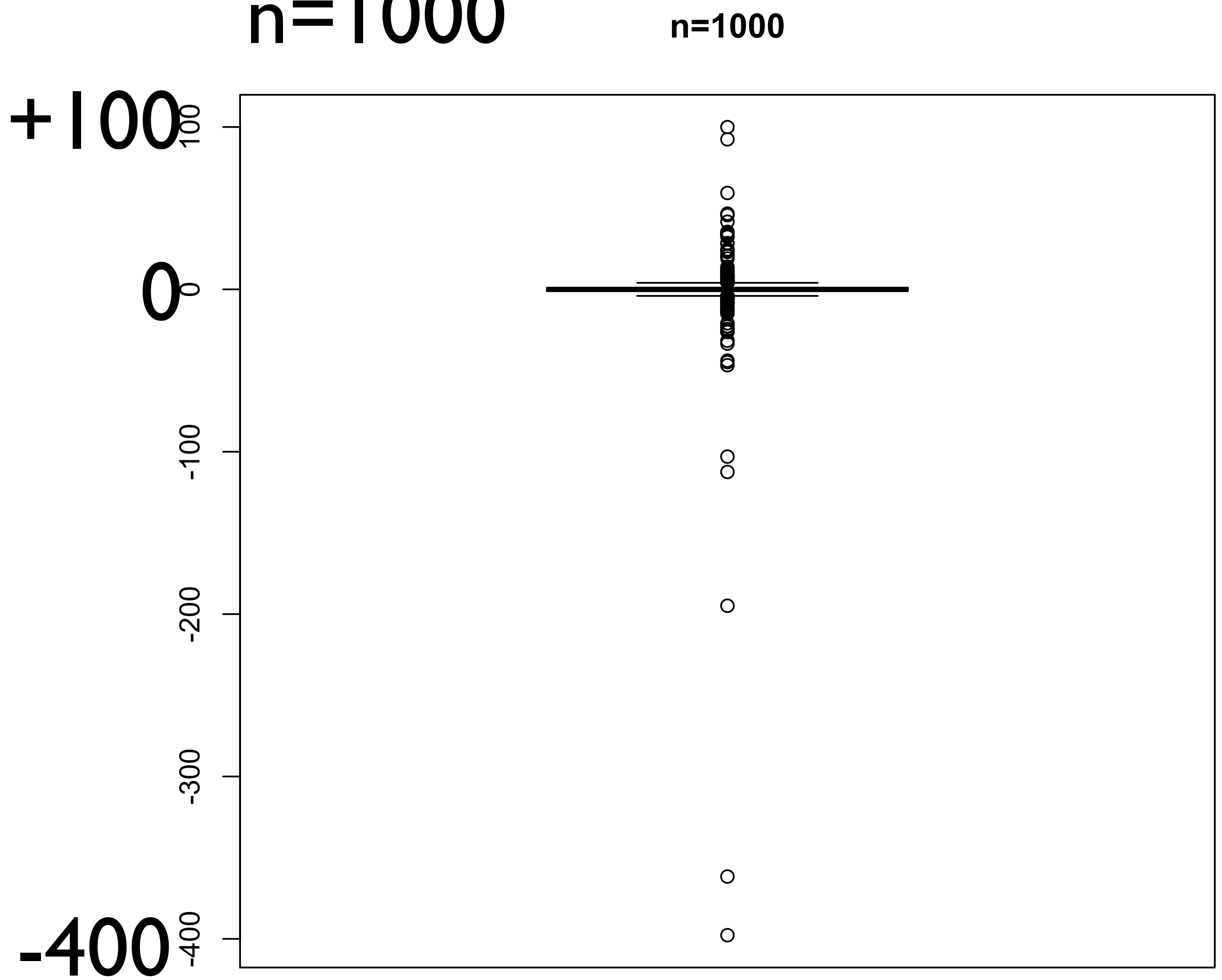
分布の裾が重い

箱ひげ図

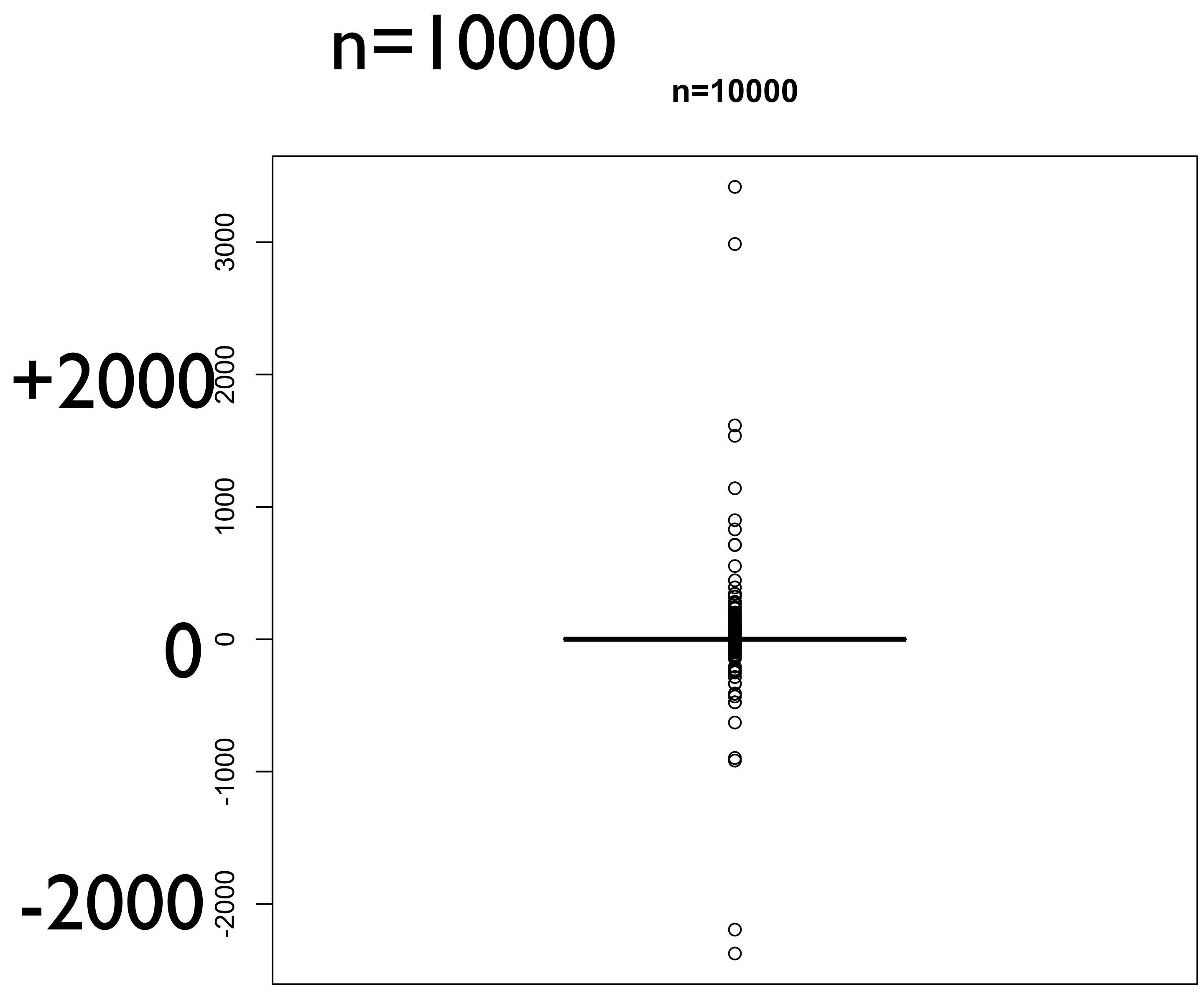
n=100

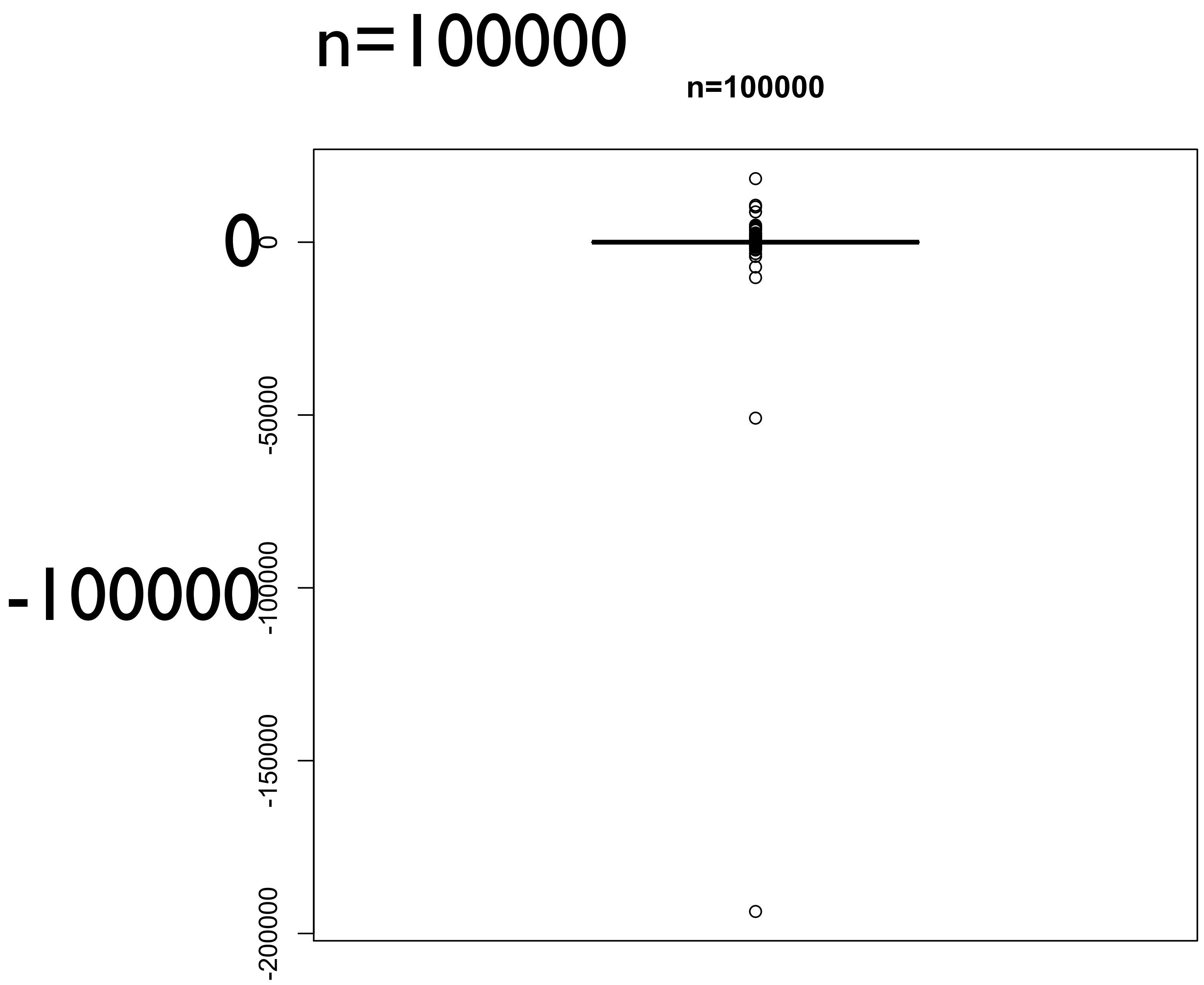


n=1000

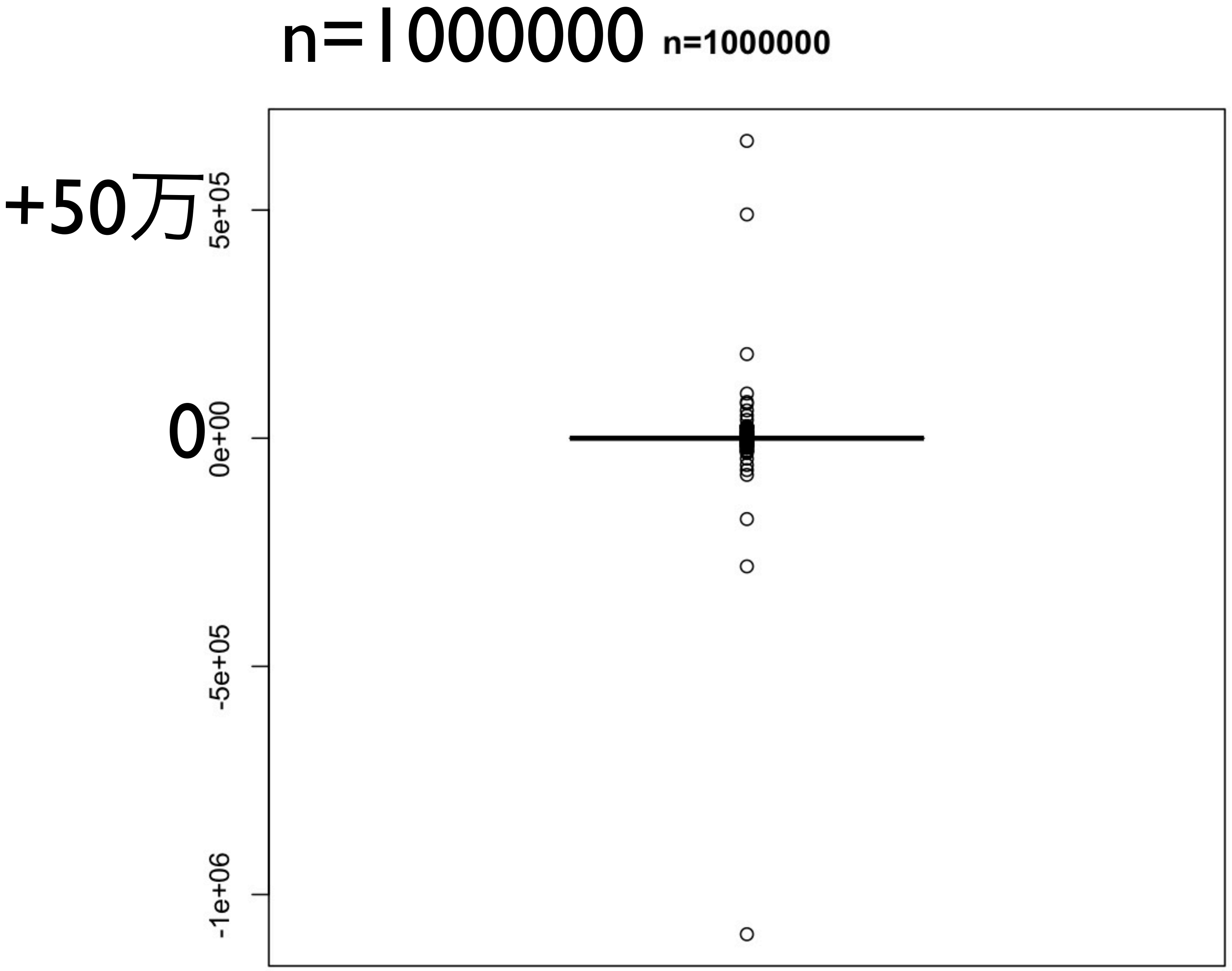


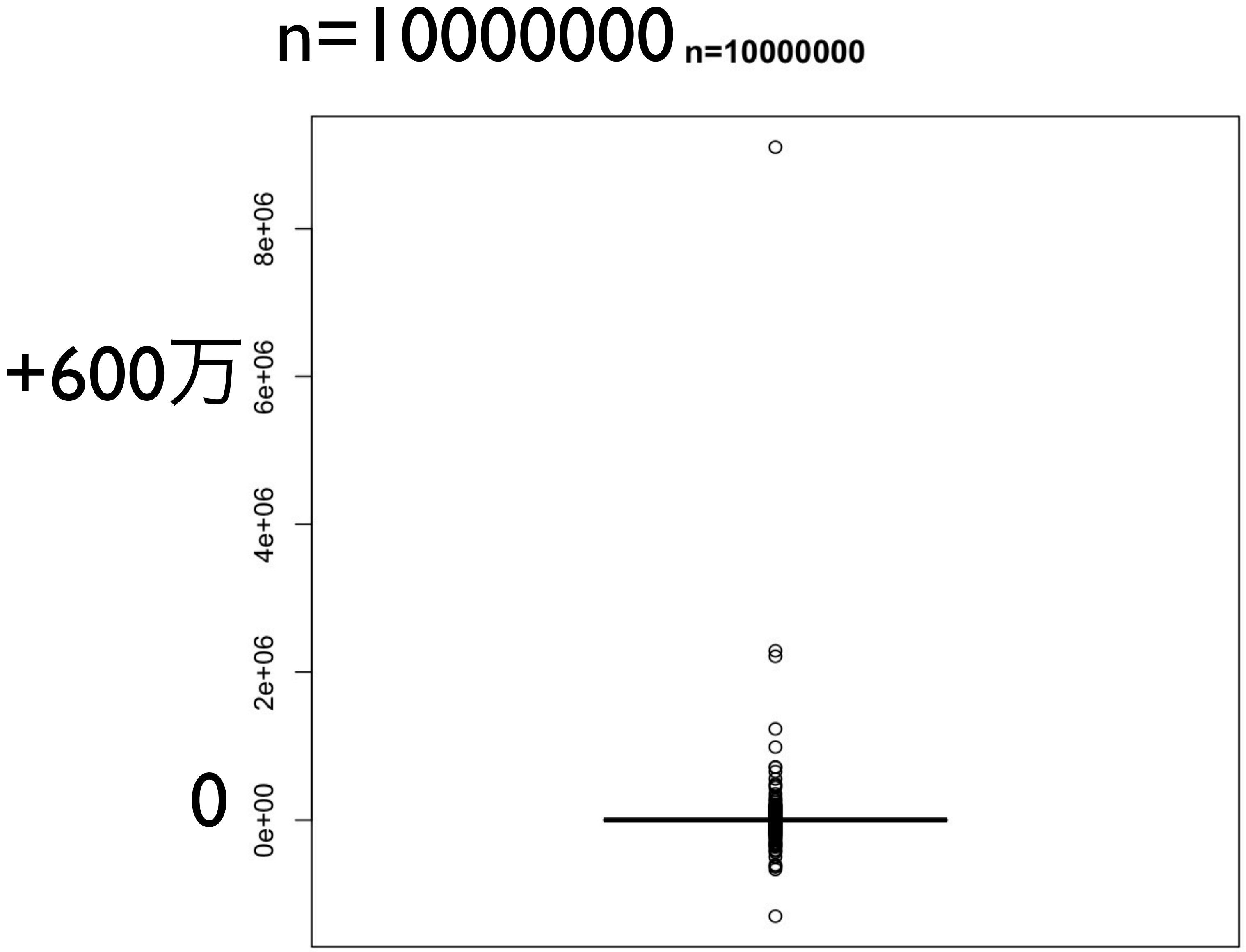
箱ひげ図


n=10000
n=10000
+2000
0
-2000
3000
2000
1000
0
-1000
-2000


n=100000
n=100000
0
-100000
0
-50000
-100000
-150000
-200000

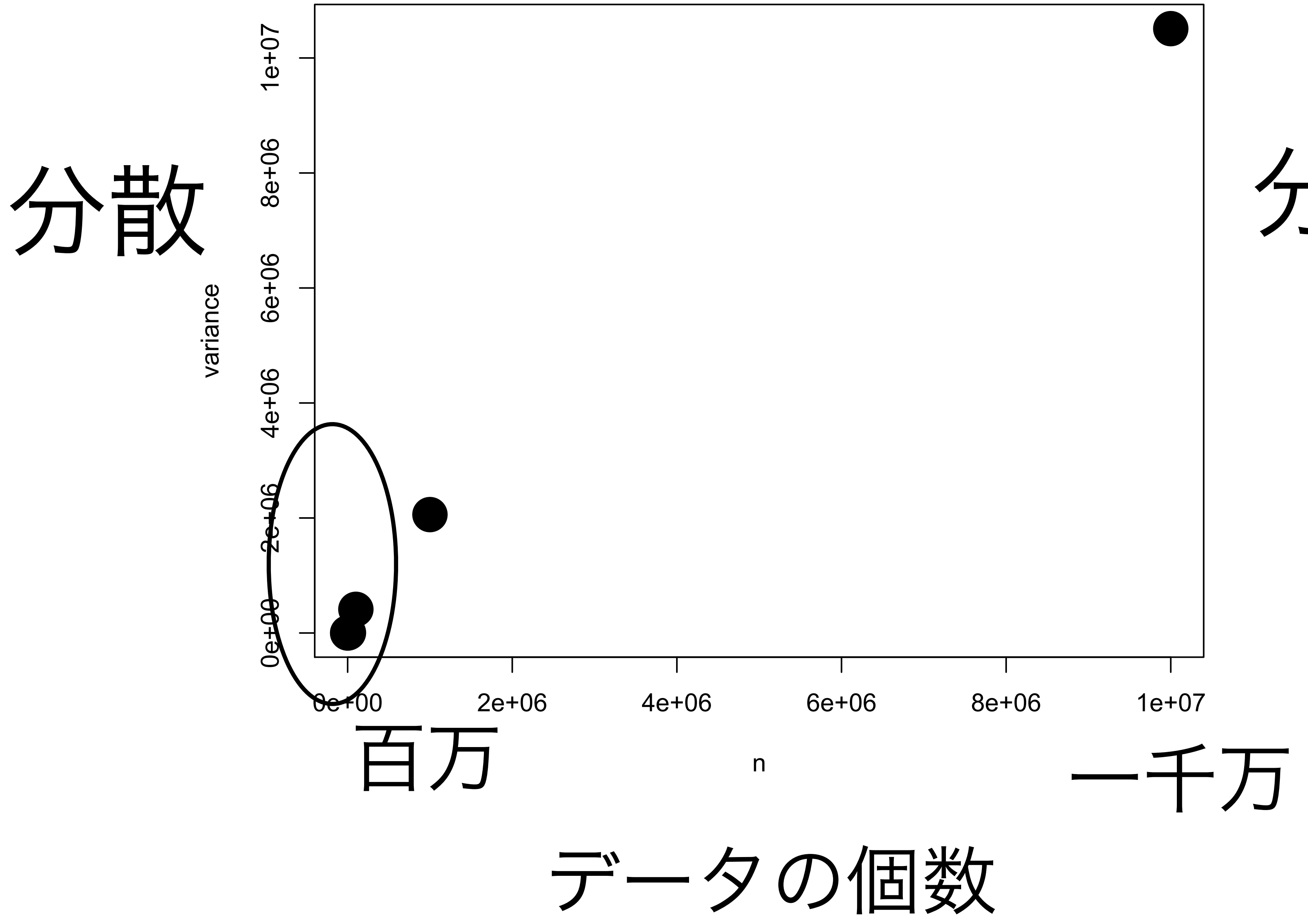
箱ひげ図


n=1000000
n=1000000
+50万
0
5e+05
0e+00
-5e+05
-1e+06

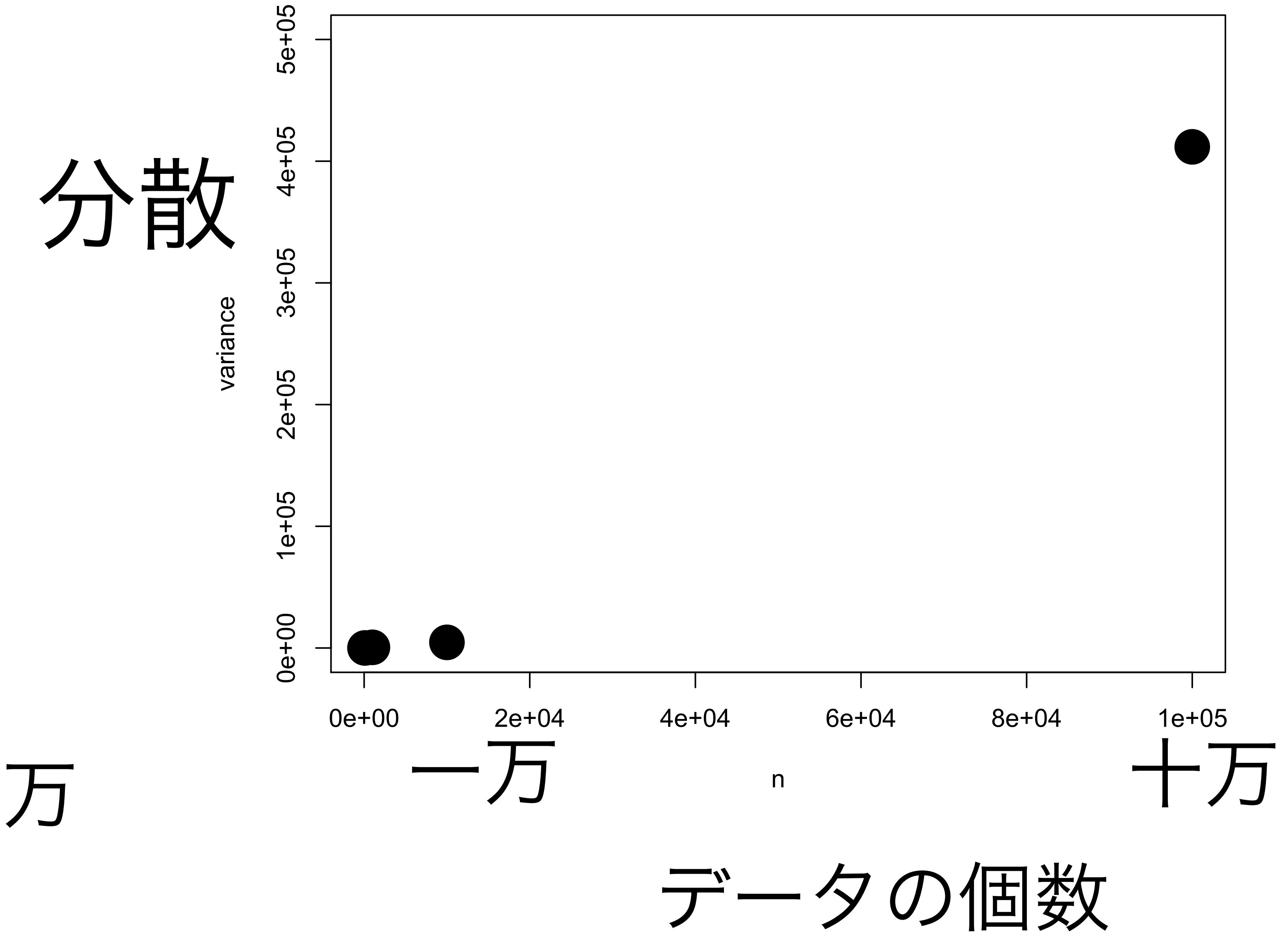

n=10000000
n=10000000
+600万
0
8e+06
6e+06
4e+06
2e+06
0e+00

分散

２つの標準正規分布の比



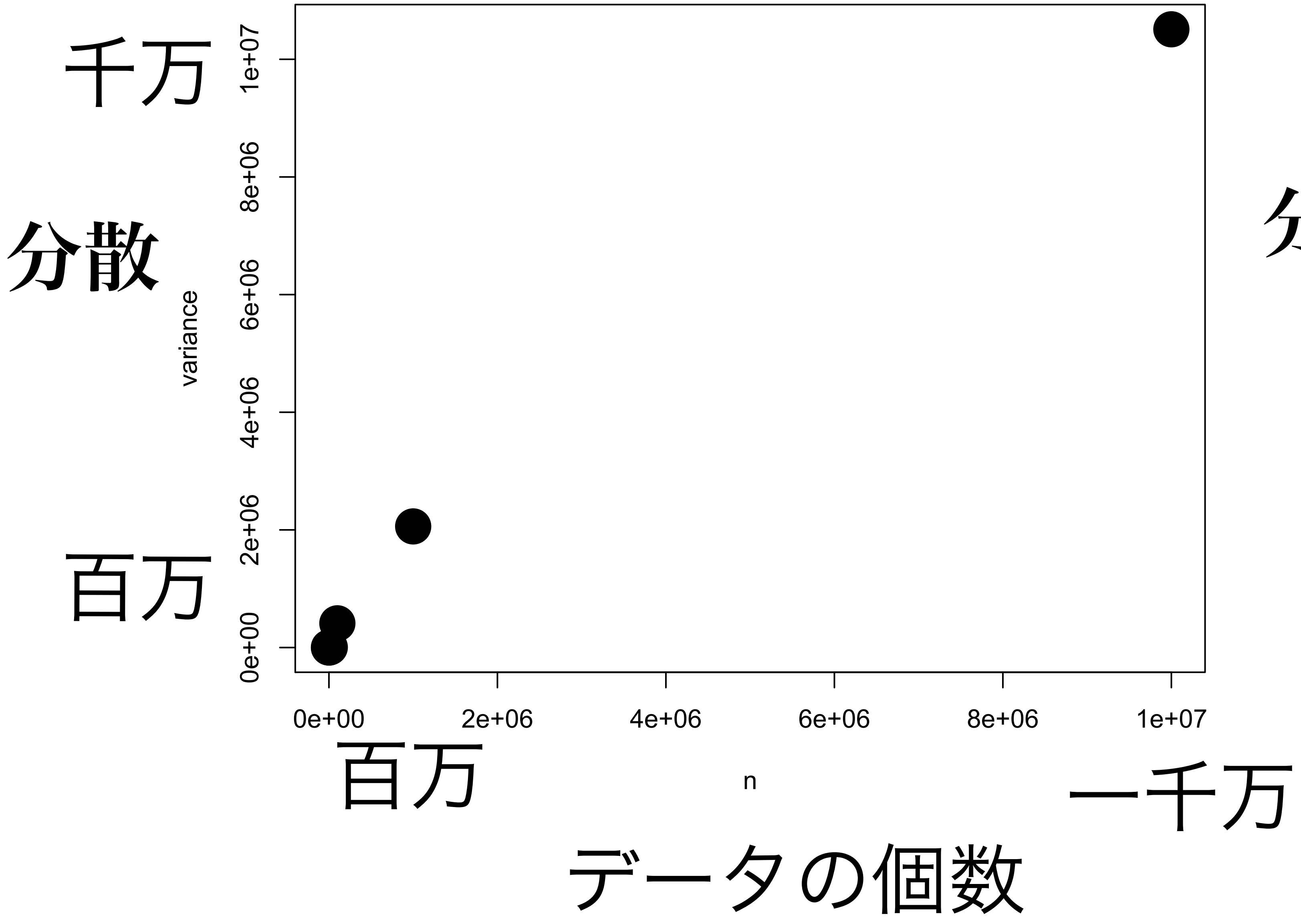
拡大図



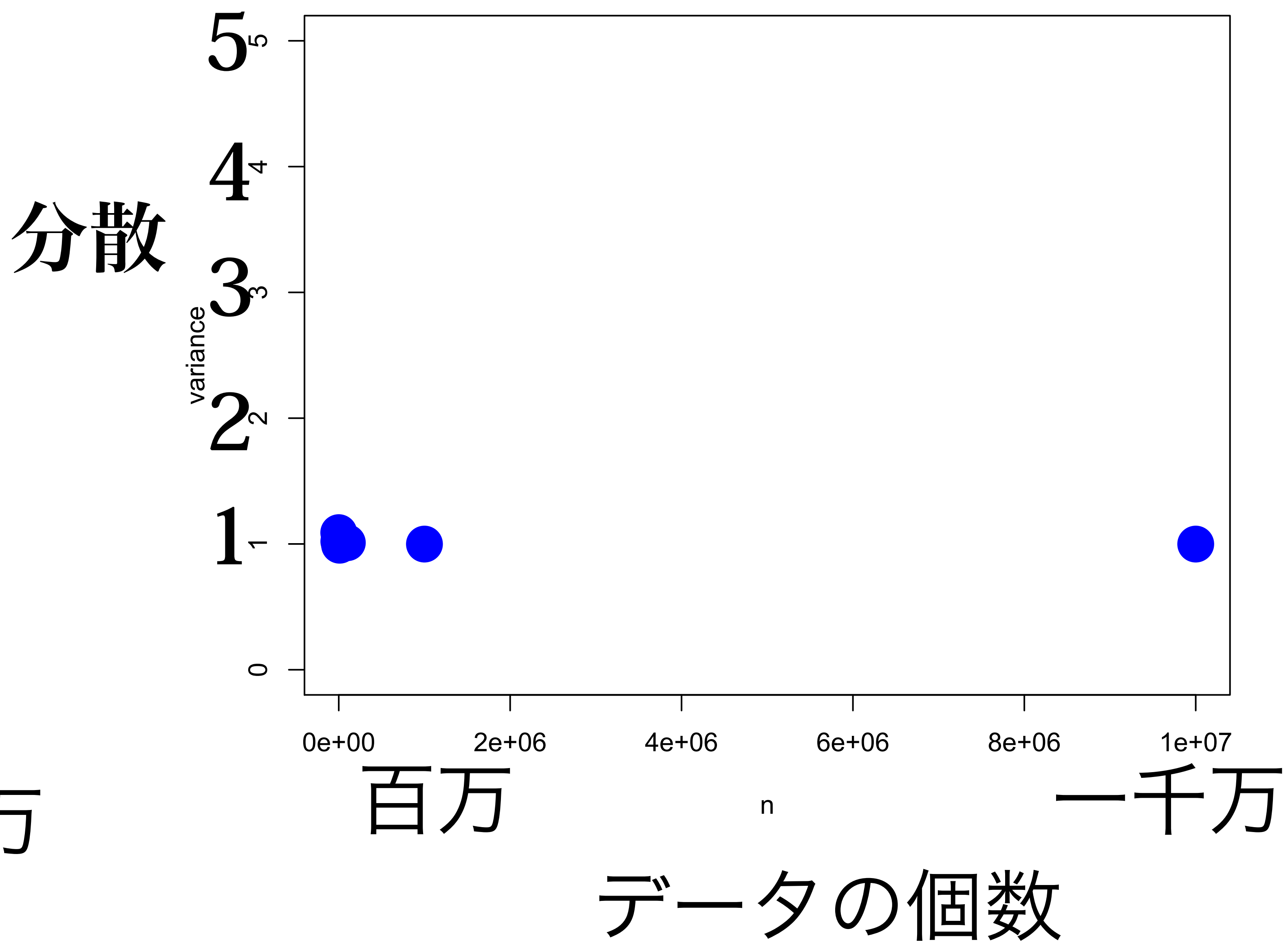
約40万倍

分散

２つの標準正規分布の比



標準正規分布



2つの標準正規分布する変数の比

標準コーシー分布

データの個数が大きくなると、

分散が大きくなる

サンプルサイズnのサンプル平均の分布

普通は
nが大きくなれば、母平均のまわりにより集まる

推定精度が改善

コーシー分布

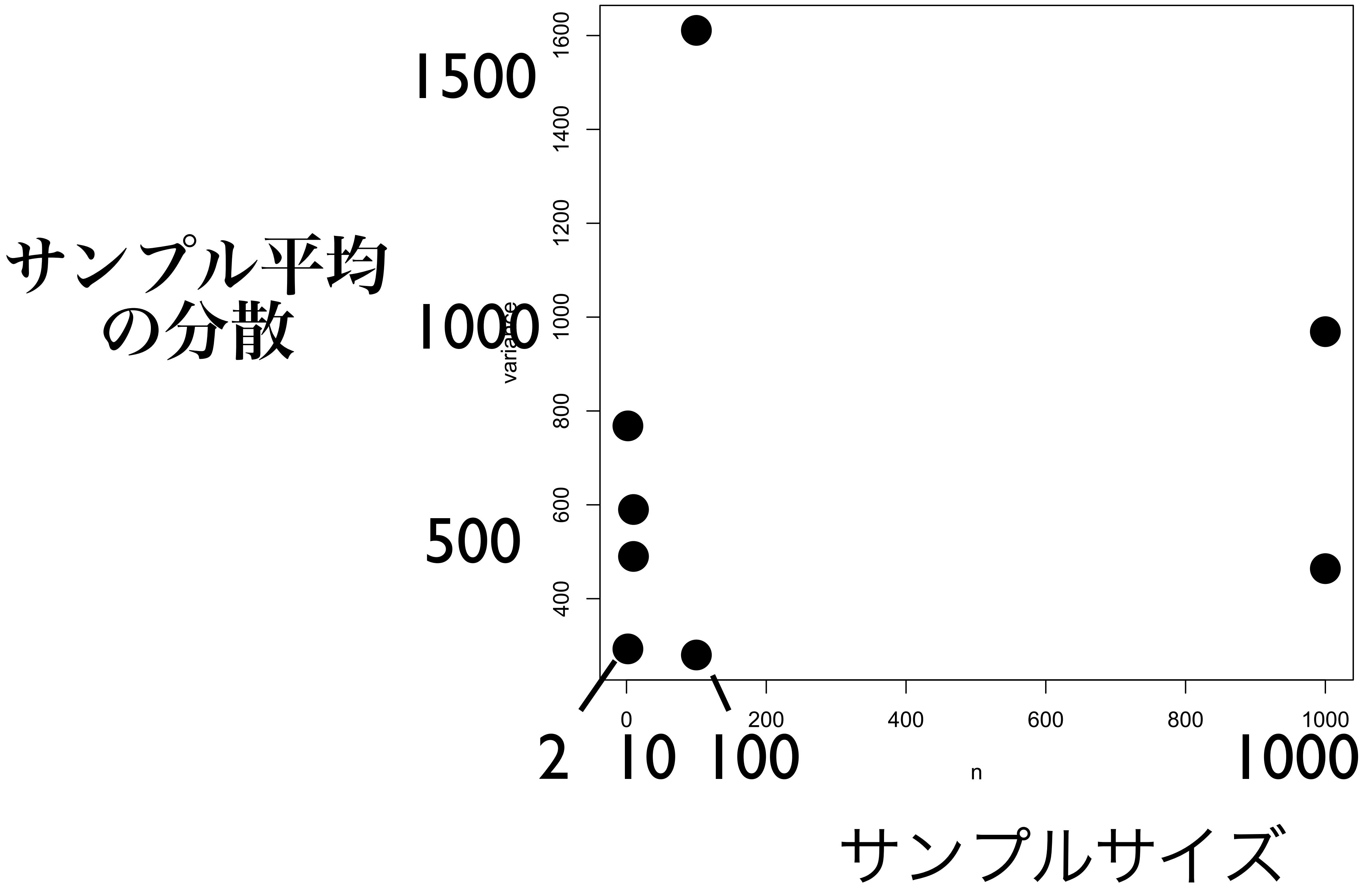
サンプルサイズnのサンプル平均の分布

もとのコーシー分布

推定値としての良さは、

サンプルサイズを増やしても改善しない

サンプルの平均値のばらつき（分散）



サイズnのサンプル1000個の間の分散

各サンプルサイズ2セット

正規分布する変数の比ーー般の場合

正規分布変数（平均θ_1分散σ_1^2） ／ 正規分布変数（平均θ_2分散σ_2^2）

相関 ρ

5パラメーター

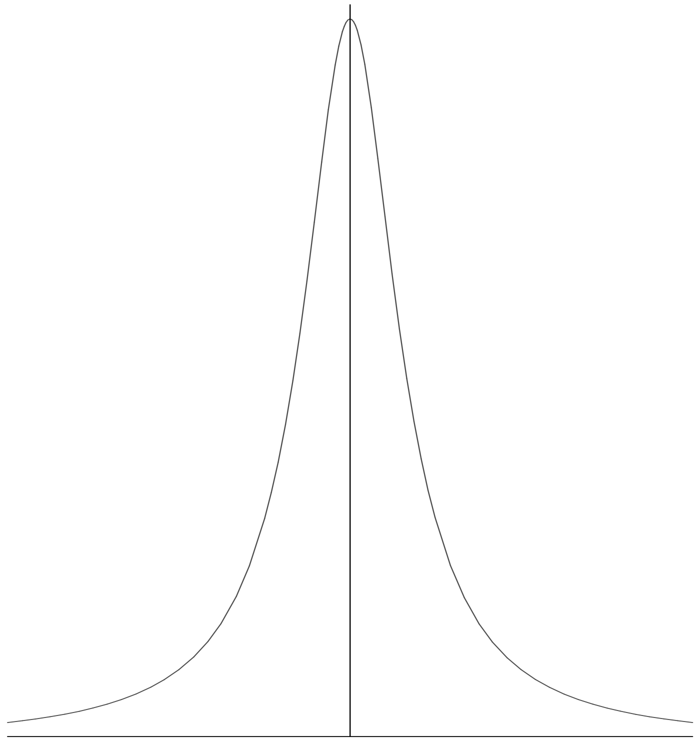
Geary(1930), Fieller(1932)

確率密度関数

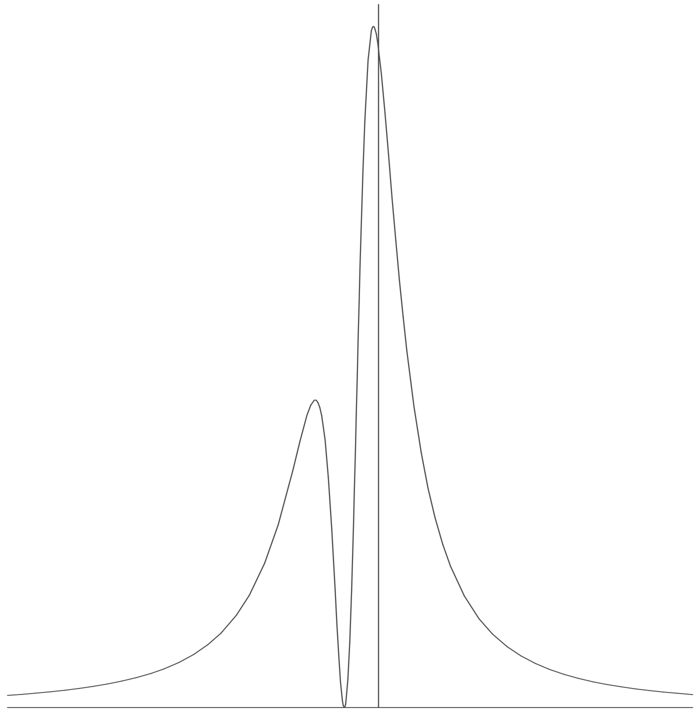
Hinkley(1969) Biometrika,56:635-639参照

標準正規分布と標準正規分布の比では１山型だが、

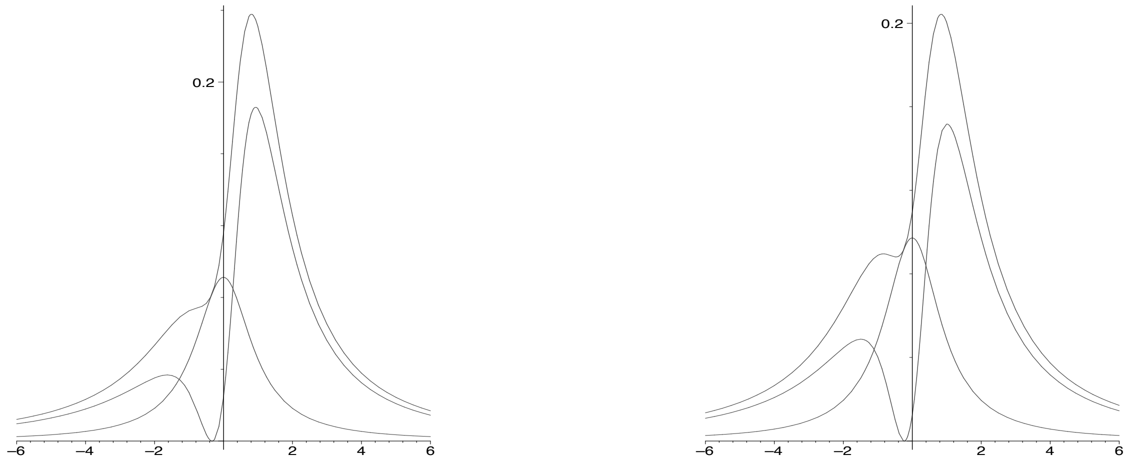
２山型になることもある



1山型の例



2山型の例



1山型と2山型の例

Marsaglia(2000)より

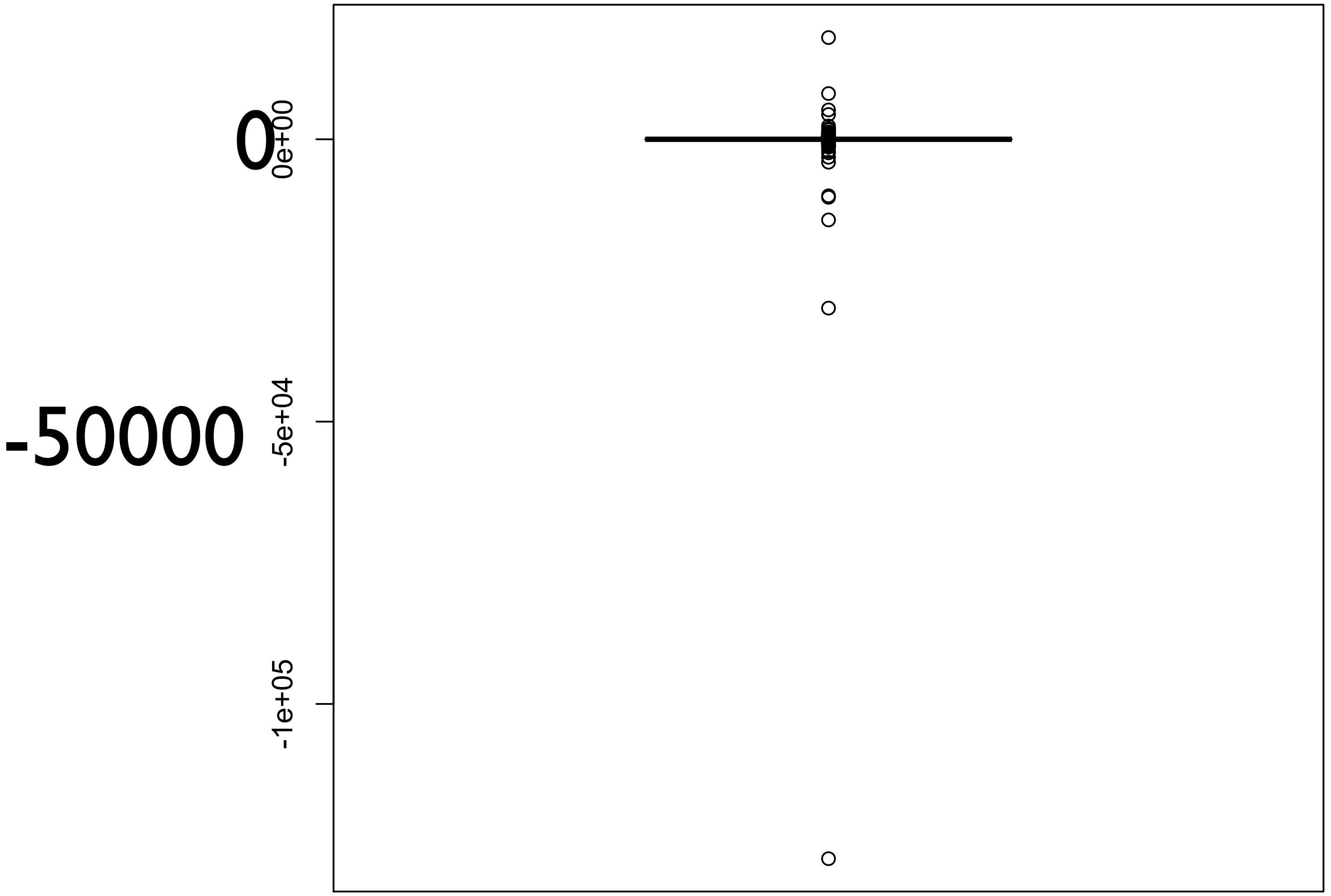
例

$$\frac{\text{平均2、分散1の正規分布する変数}}{\text{平均5、分散4の正規分布する変数}}$$ 独立（相関なし）

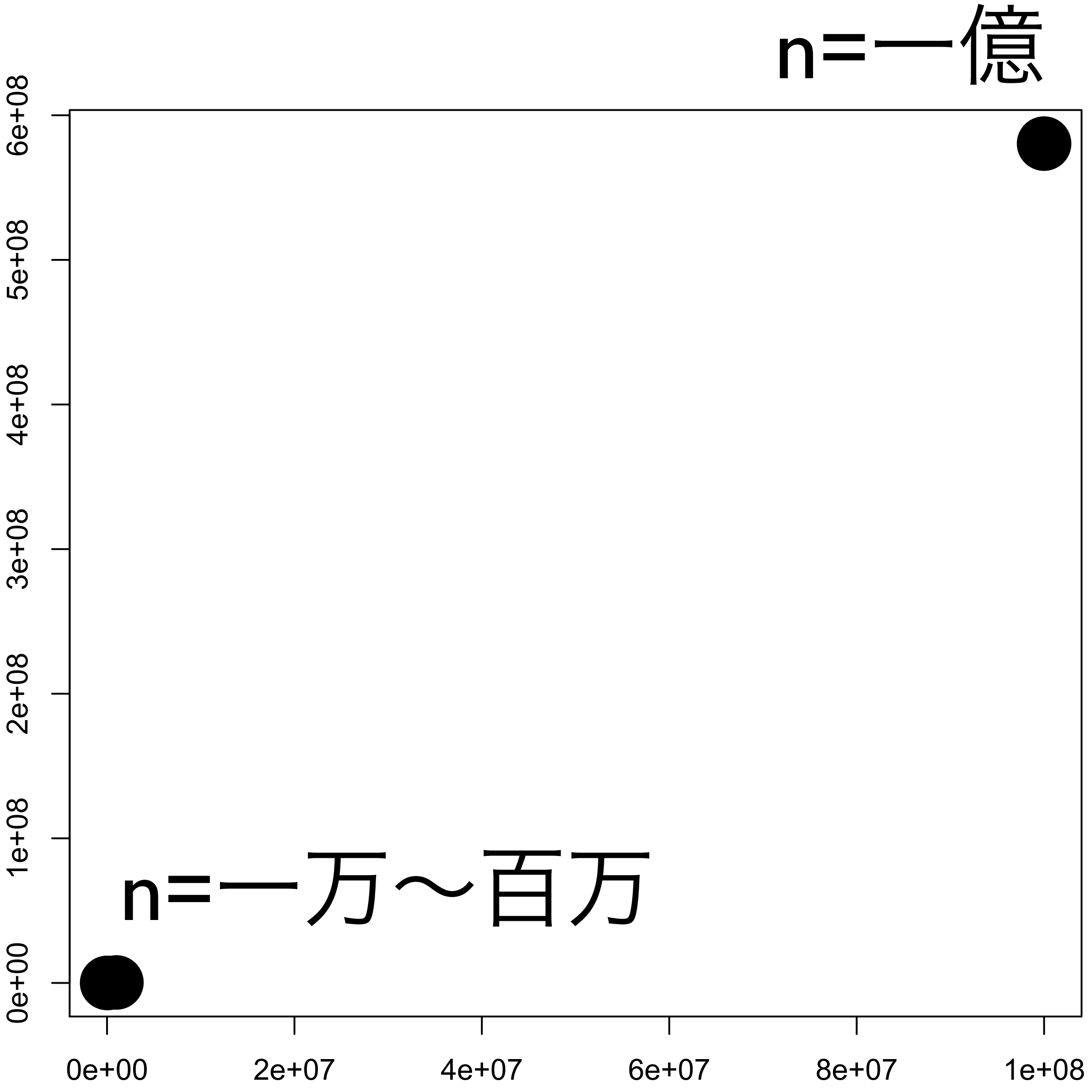
n=100000での例

’箱’部分は横棒に見える

箱ひげ図

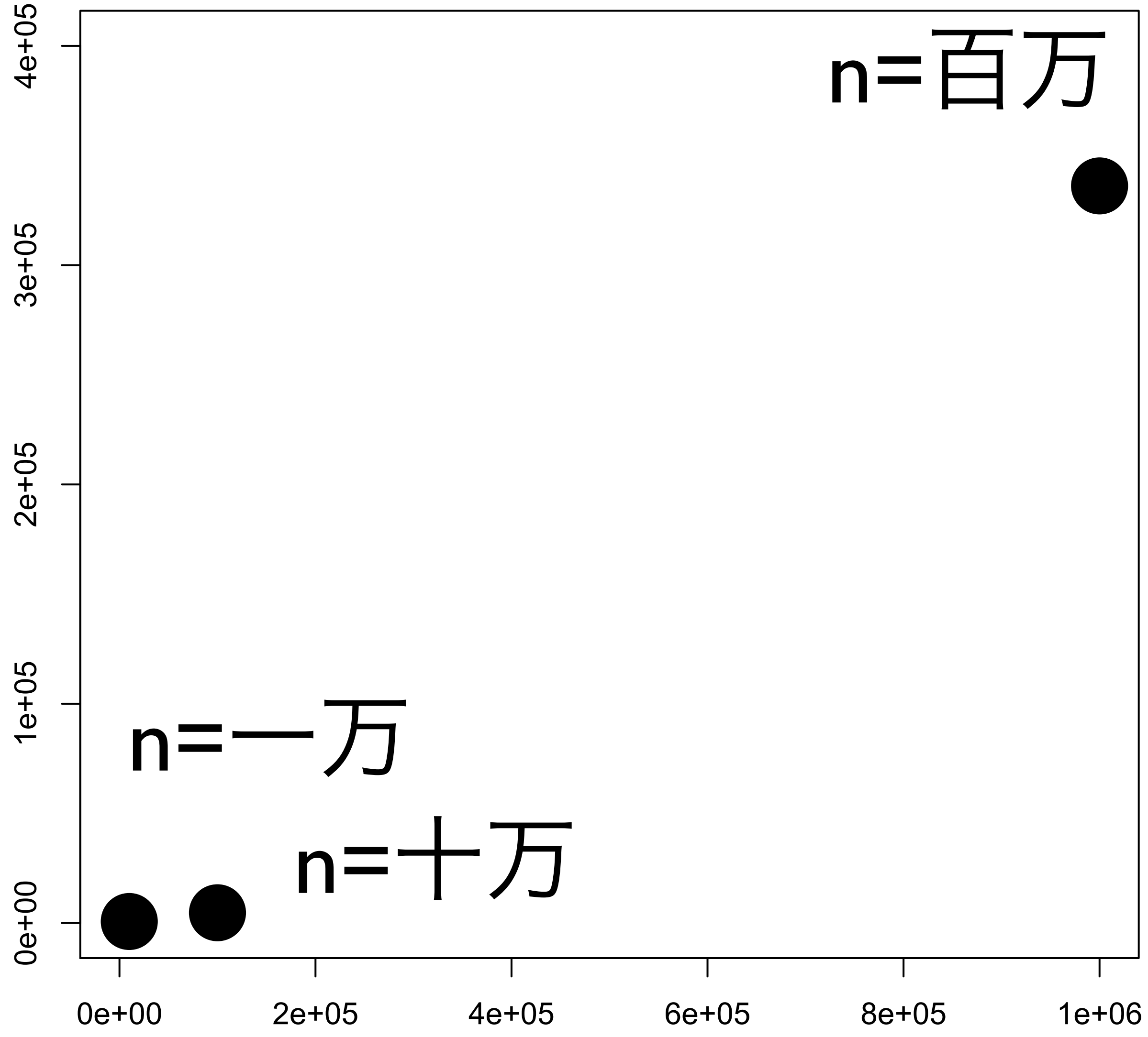


分散



拡大

分散



約80万倍

例

X 平均３で分散1の正規分布

Z 平均３で分散1の正規分布

$$\frac{\text{正規分布する変数X}}{\text{正規分布する変数X＋正規分布する変数Z}}$$

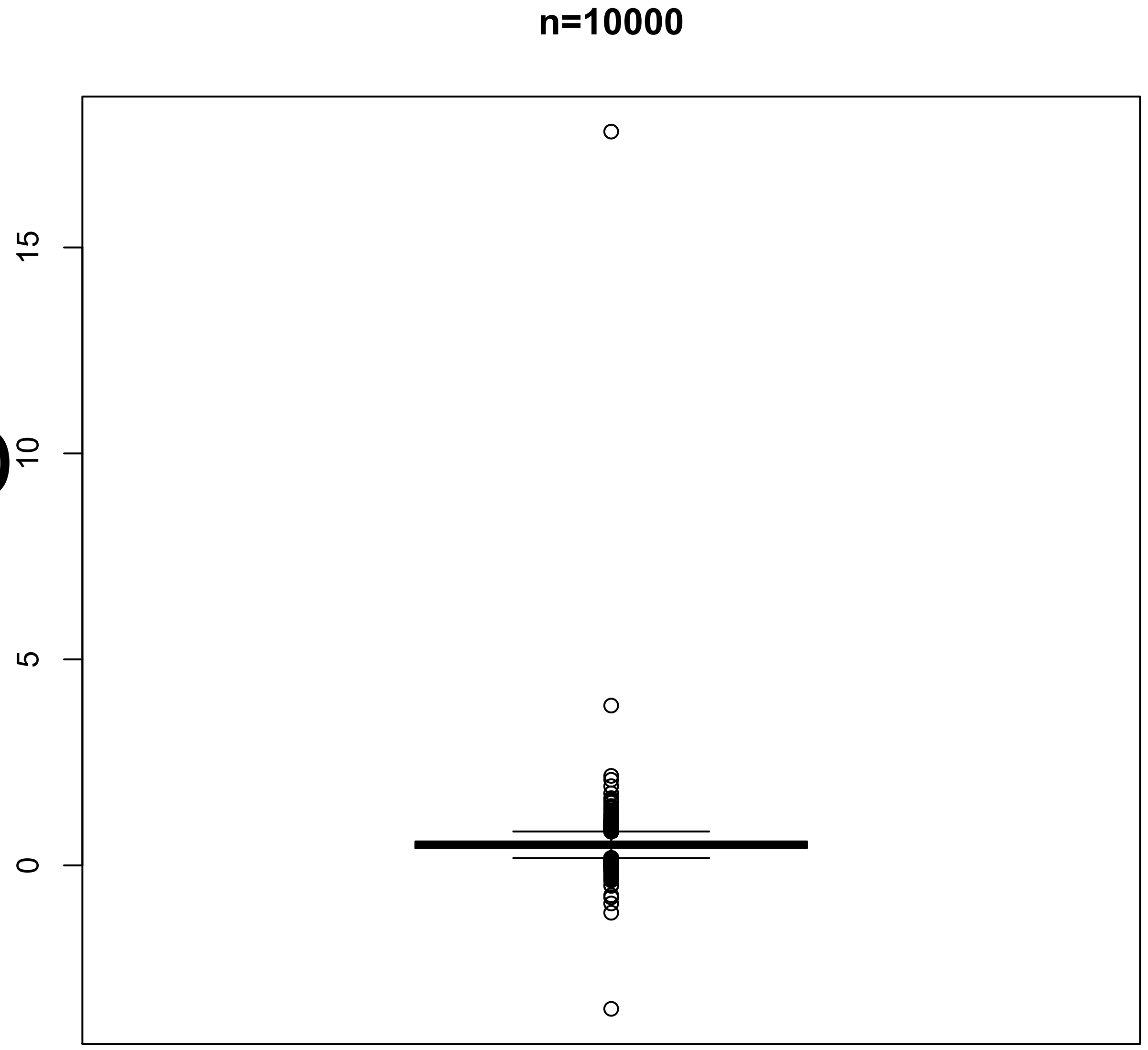
独立（相関なし）

たとえば

全体は２つの構成部分からなる一全体のうちその片方の構成部分の割合

をイメージしてみた

箱ひげ図



10

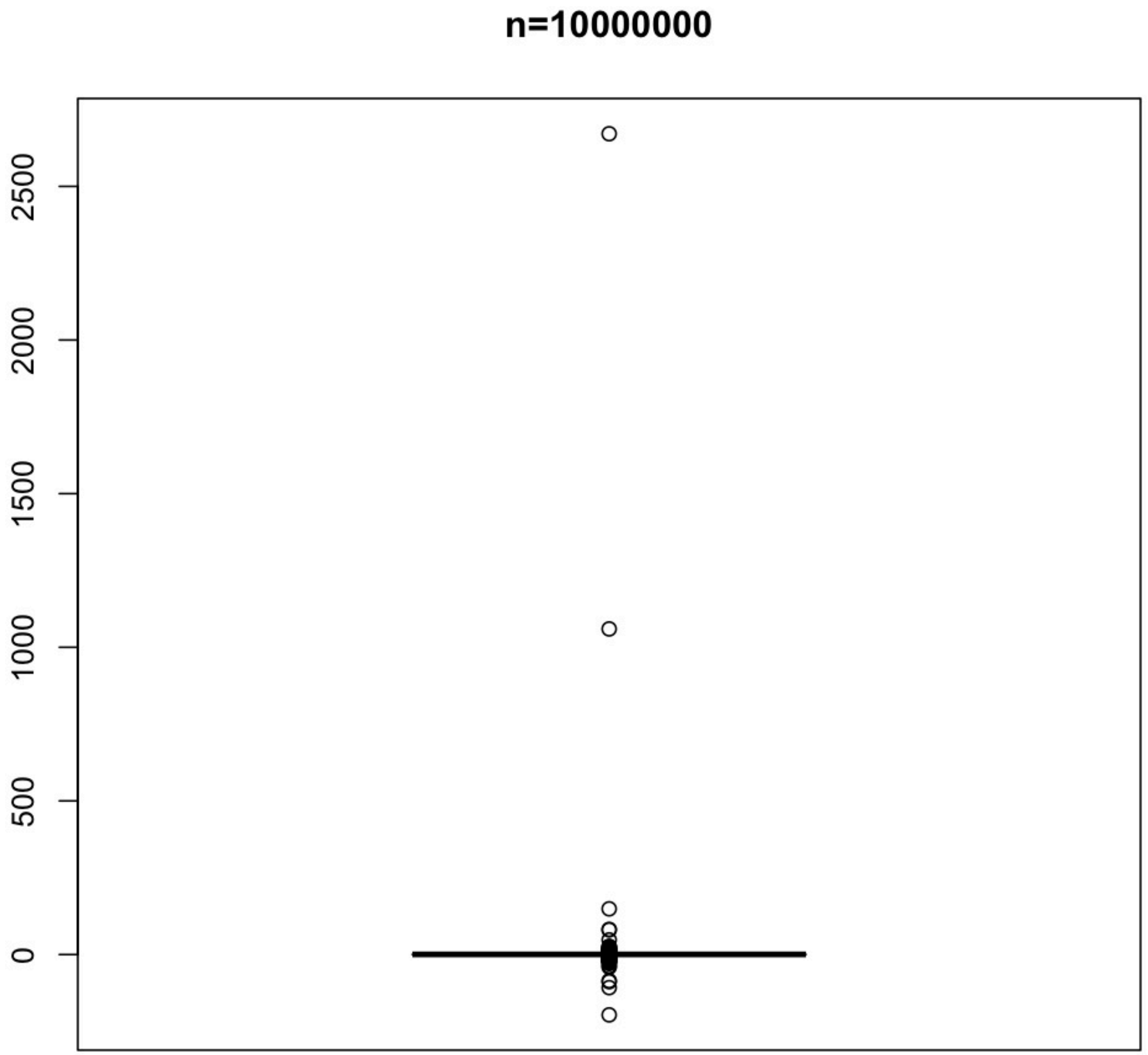
0

n=10000

箱ひげ図

1000

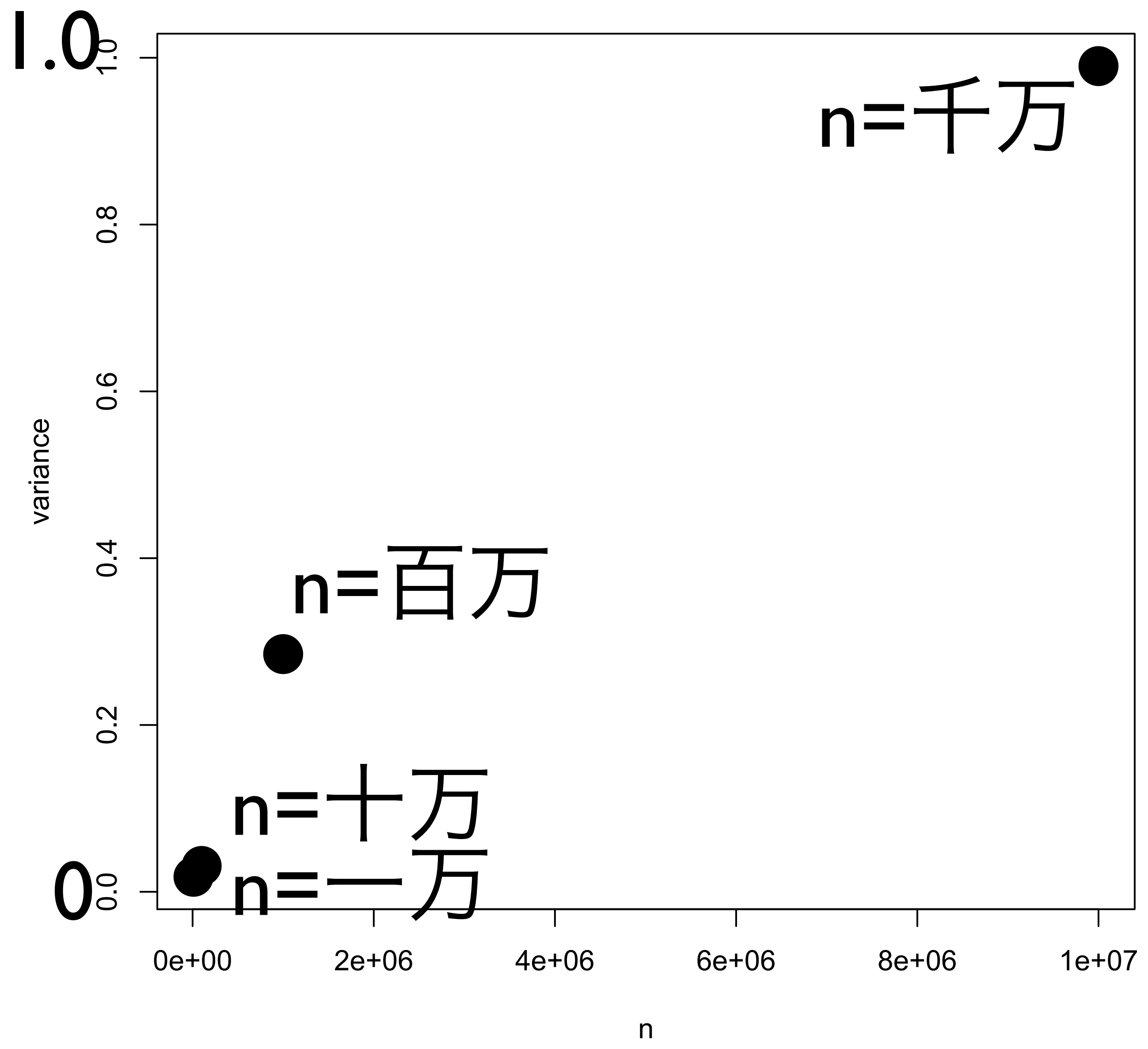
0



n=10000000

全体は２つの構成部分からなる
全体のうちその片方の構成部分の割合
をイメージしてみた

分散



サンプルサイズ

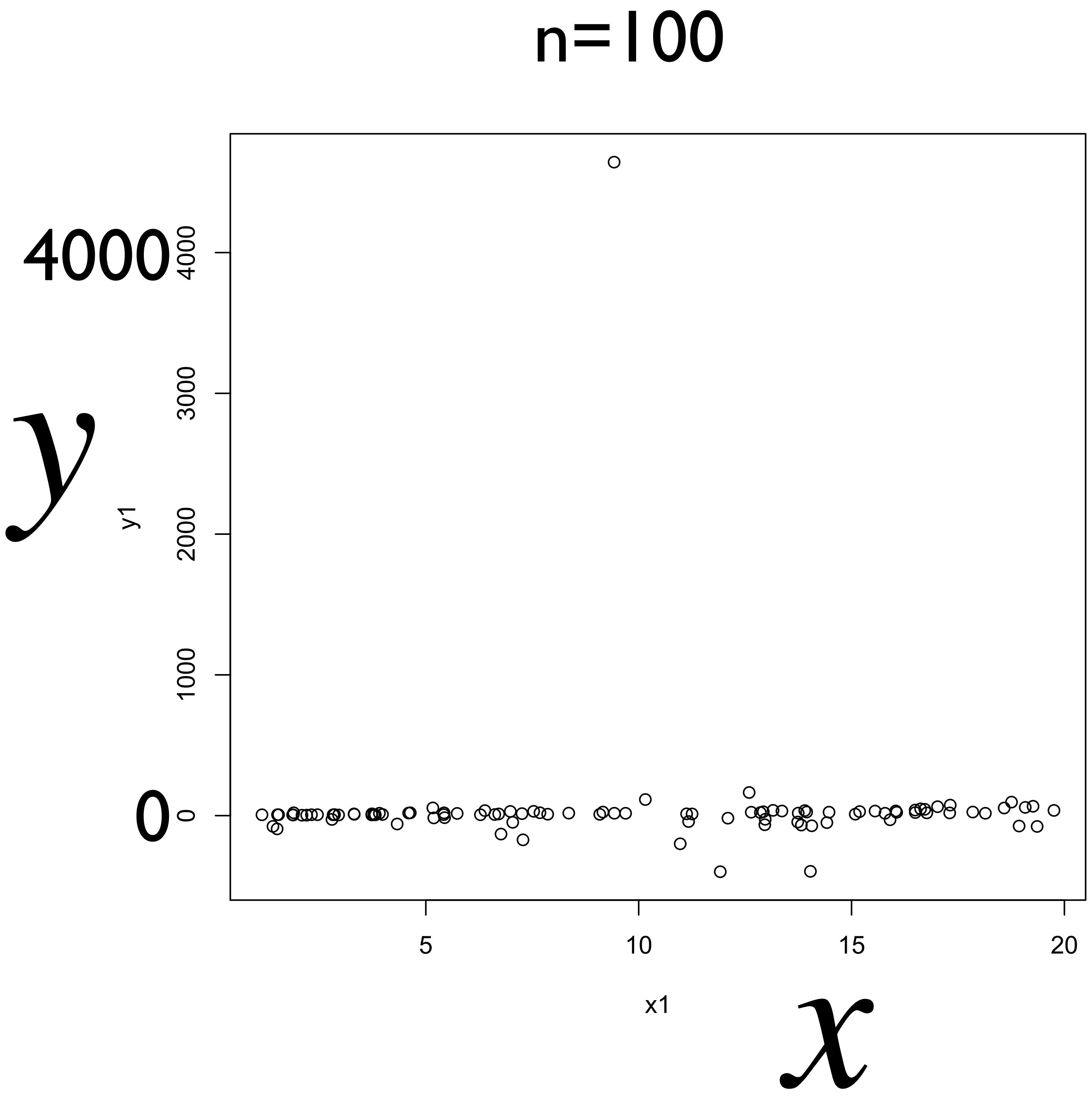
二項分布する変数なら
n=１千万のとき、
標準偏差は約0.00016

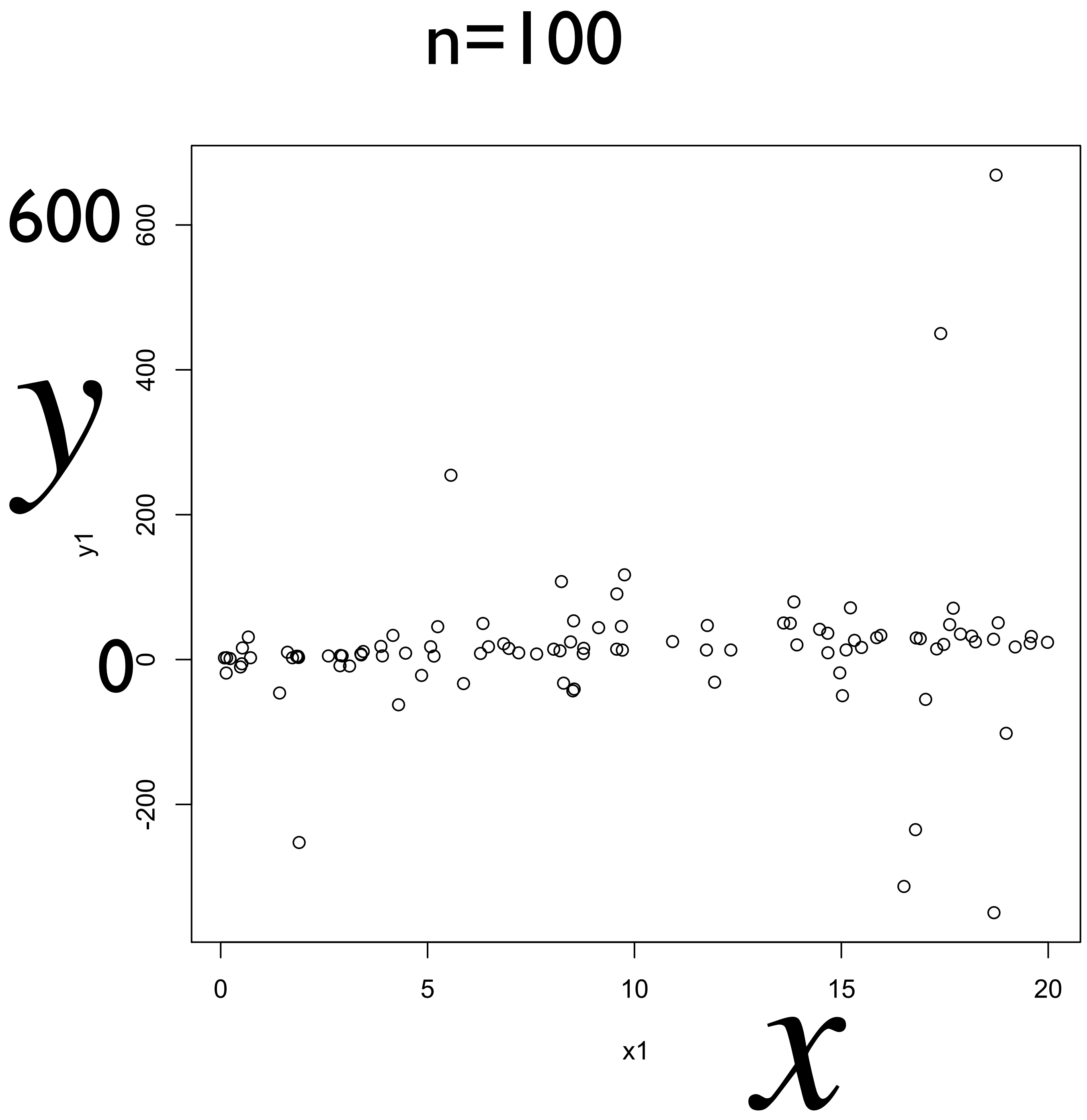
回帰　目的変数が正規分布の比

説明変数 x　0~20の一様乱数

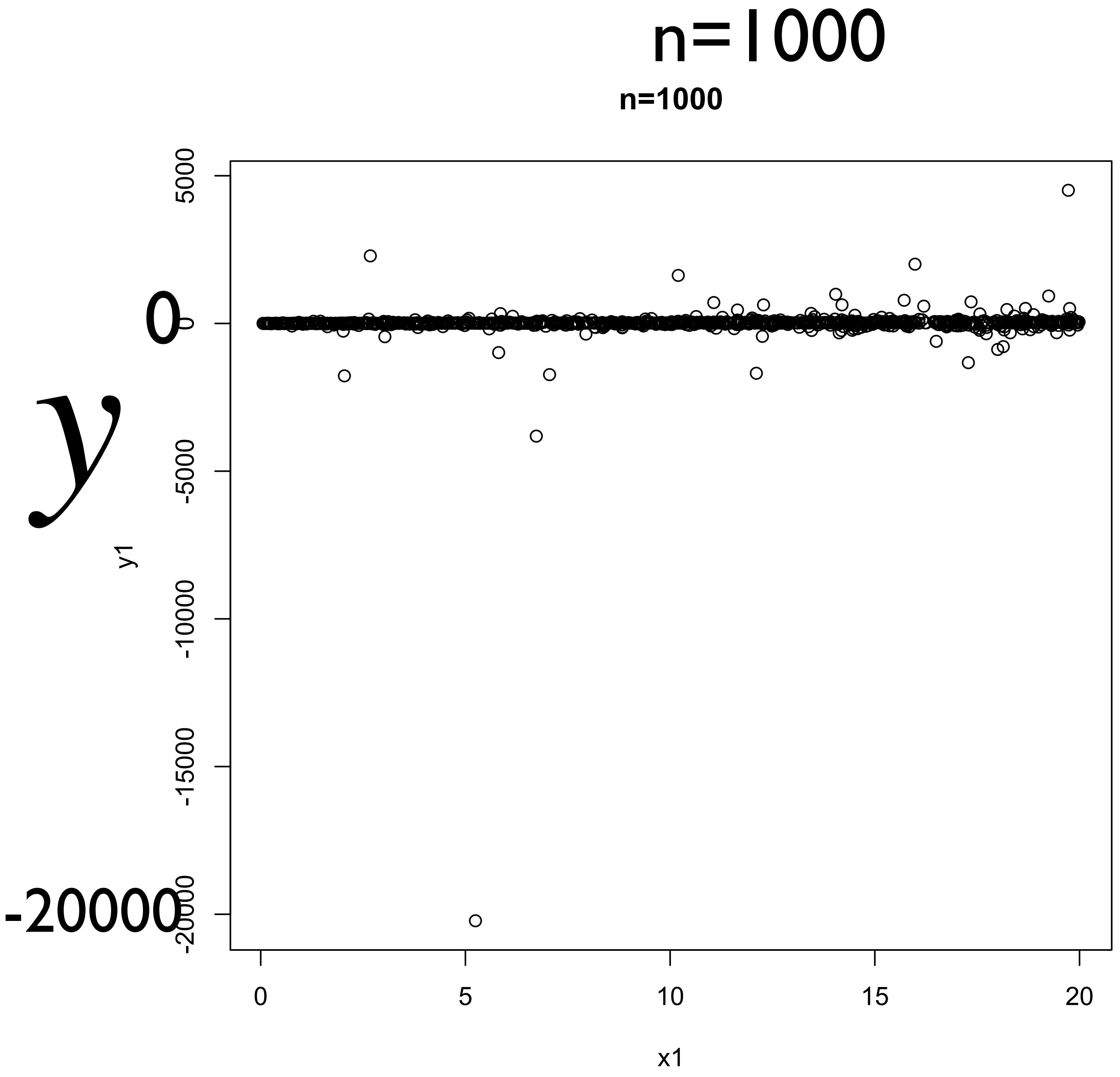
目的変数 y　比 $$\frac{\text{平均}2x+3.5\text{、分散}1\text{の正規分布}}{\text{平均}1\text{、分散}1\text{の正規分布}}$$

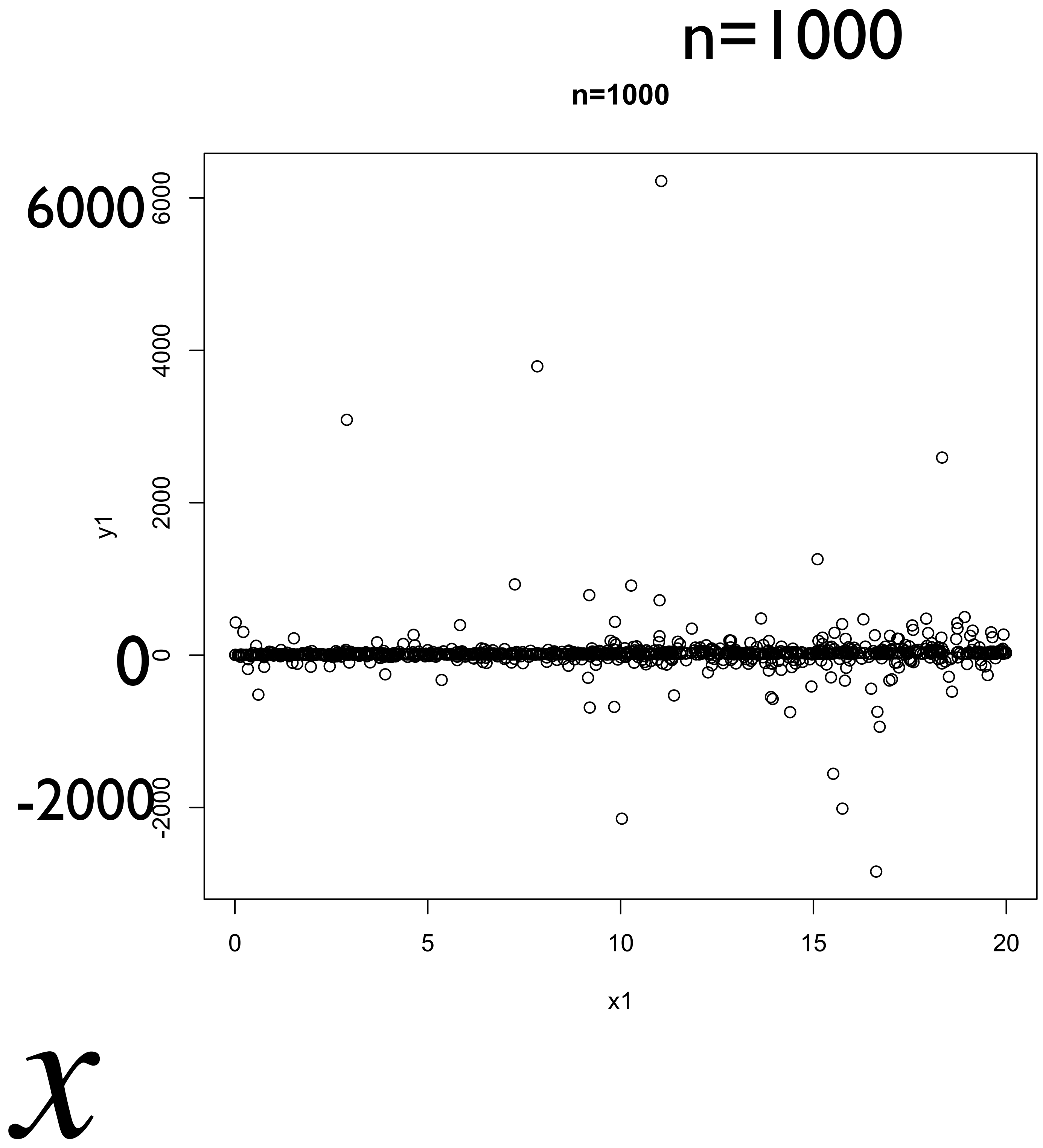
乱数で生成したデータの例



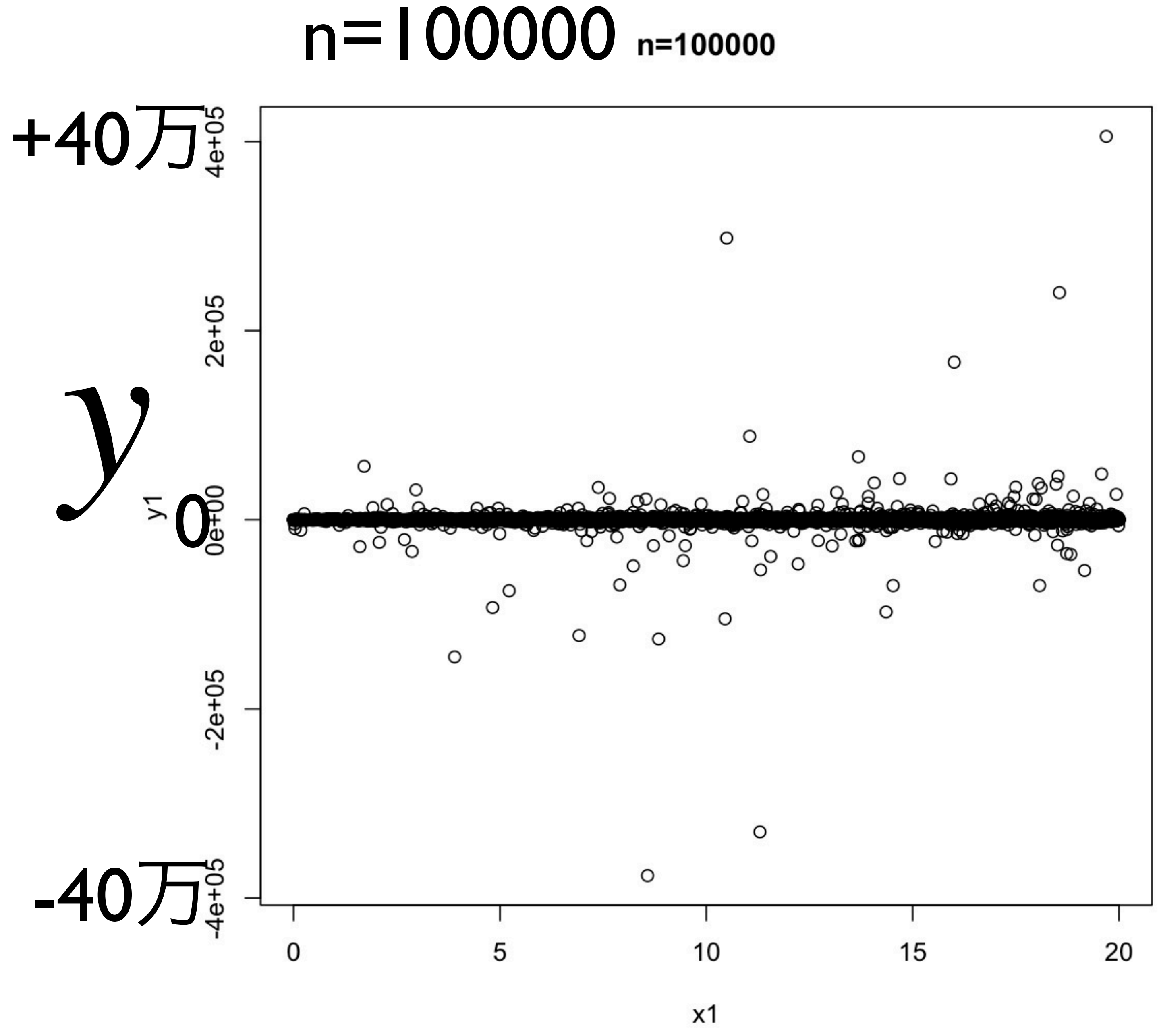


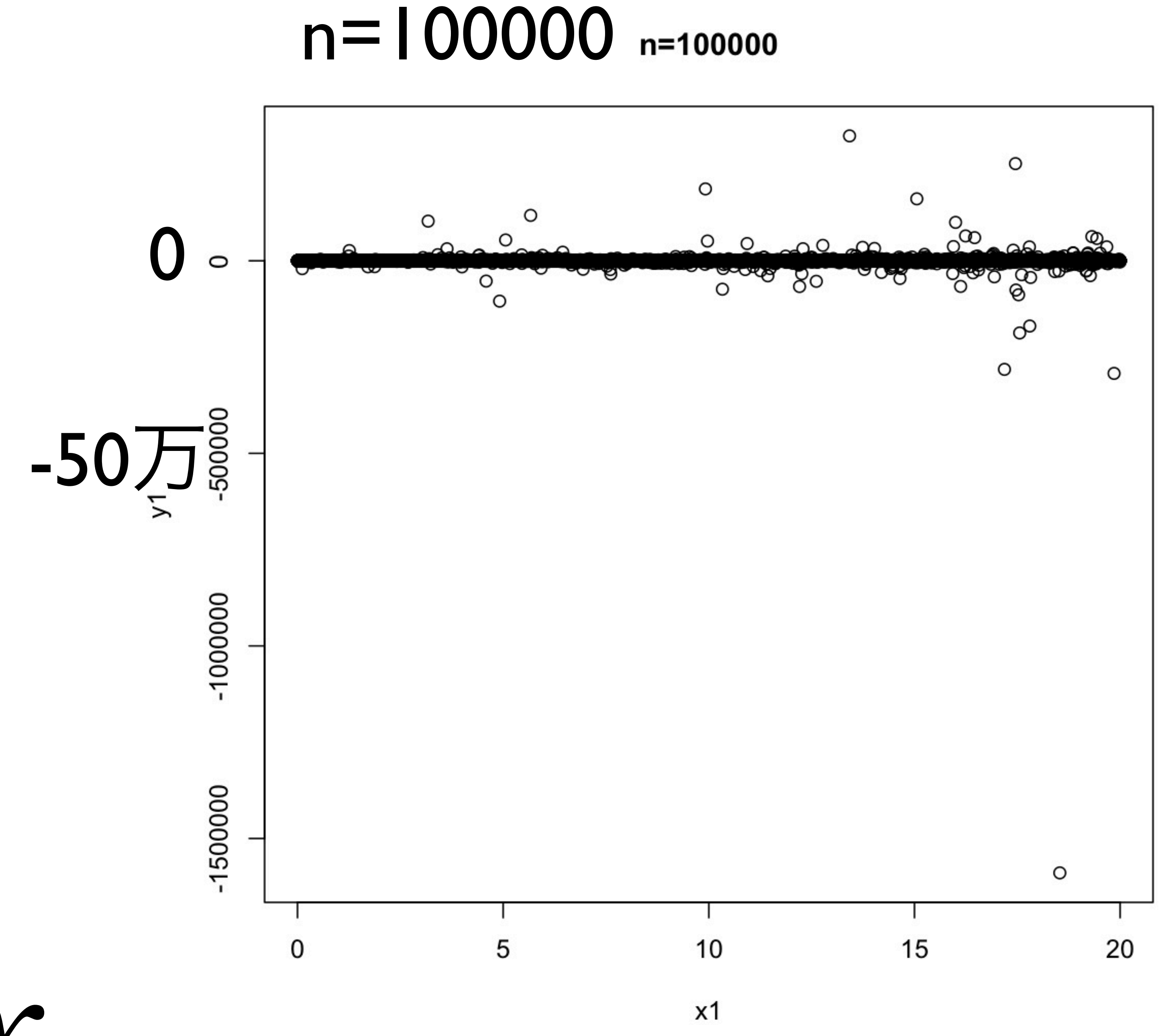
乱数で生成したデータの例





乱数で生成したデータの例



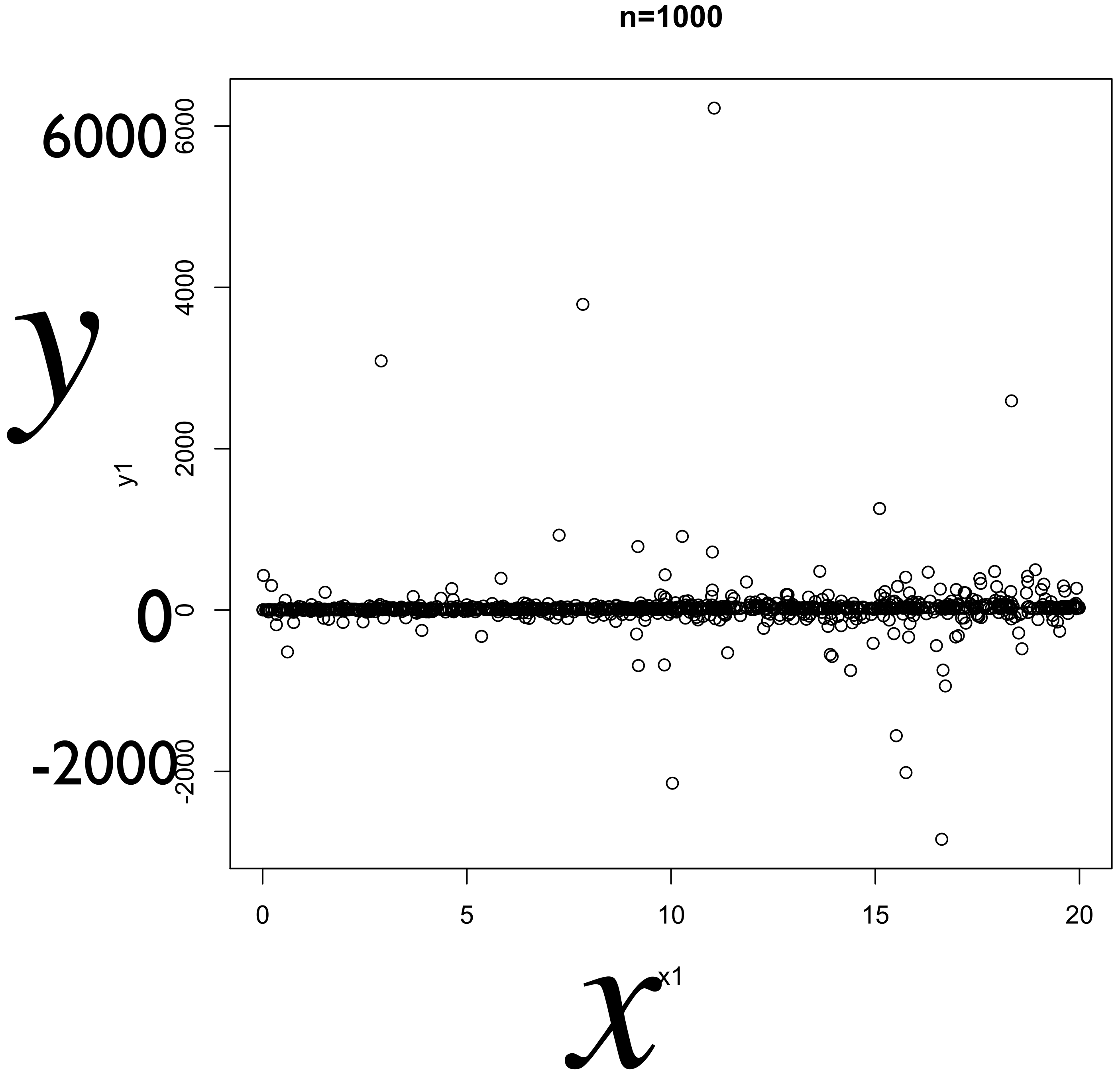


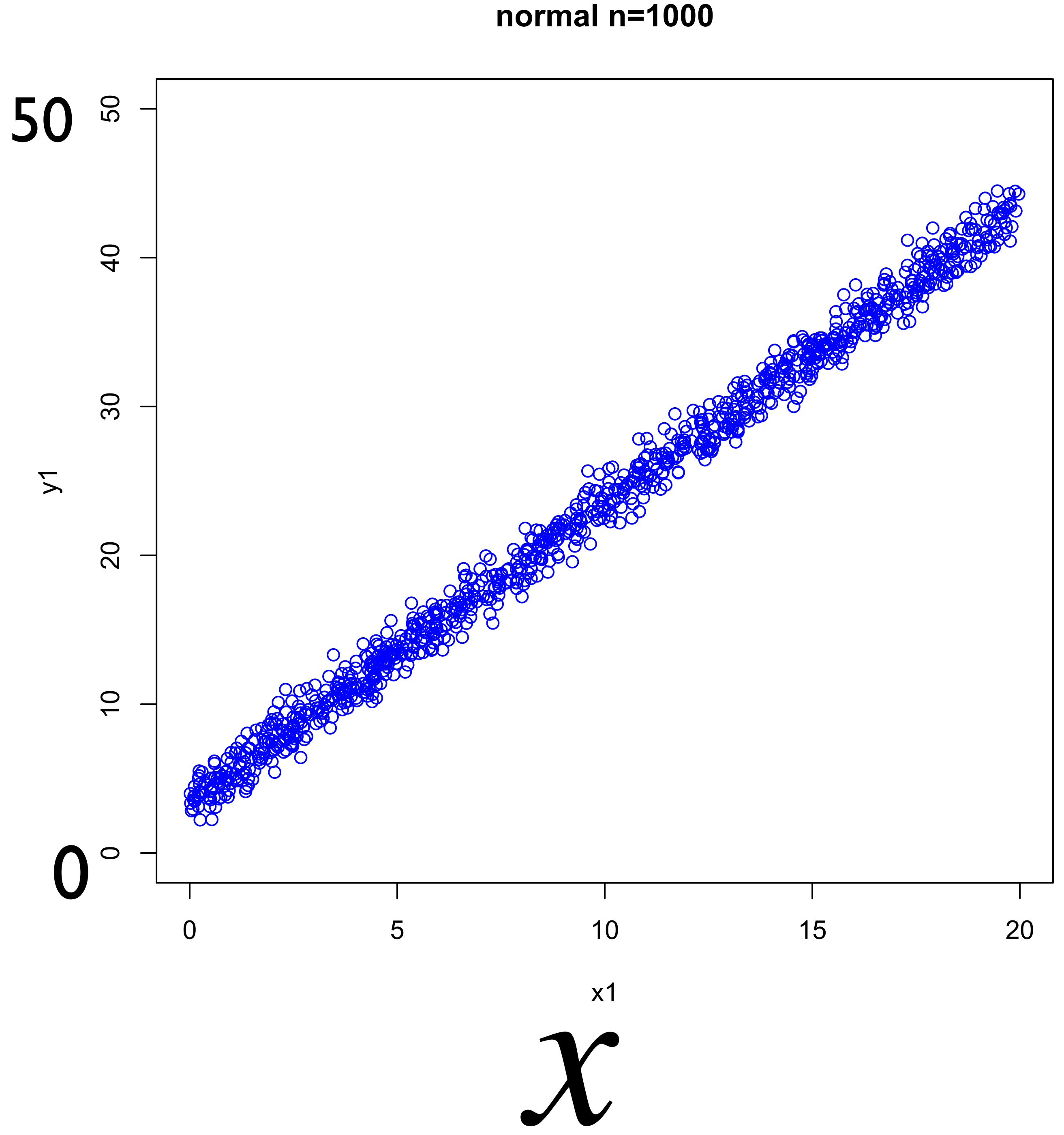
x

乱数で生成したデータの例　n=1000

y　$\dfrac{\text{平均2x+3.5、分散1の正規分布}}{\text{平均1、分散1の正規分布}}$

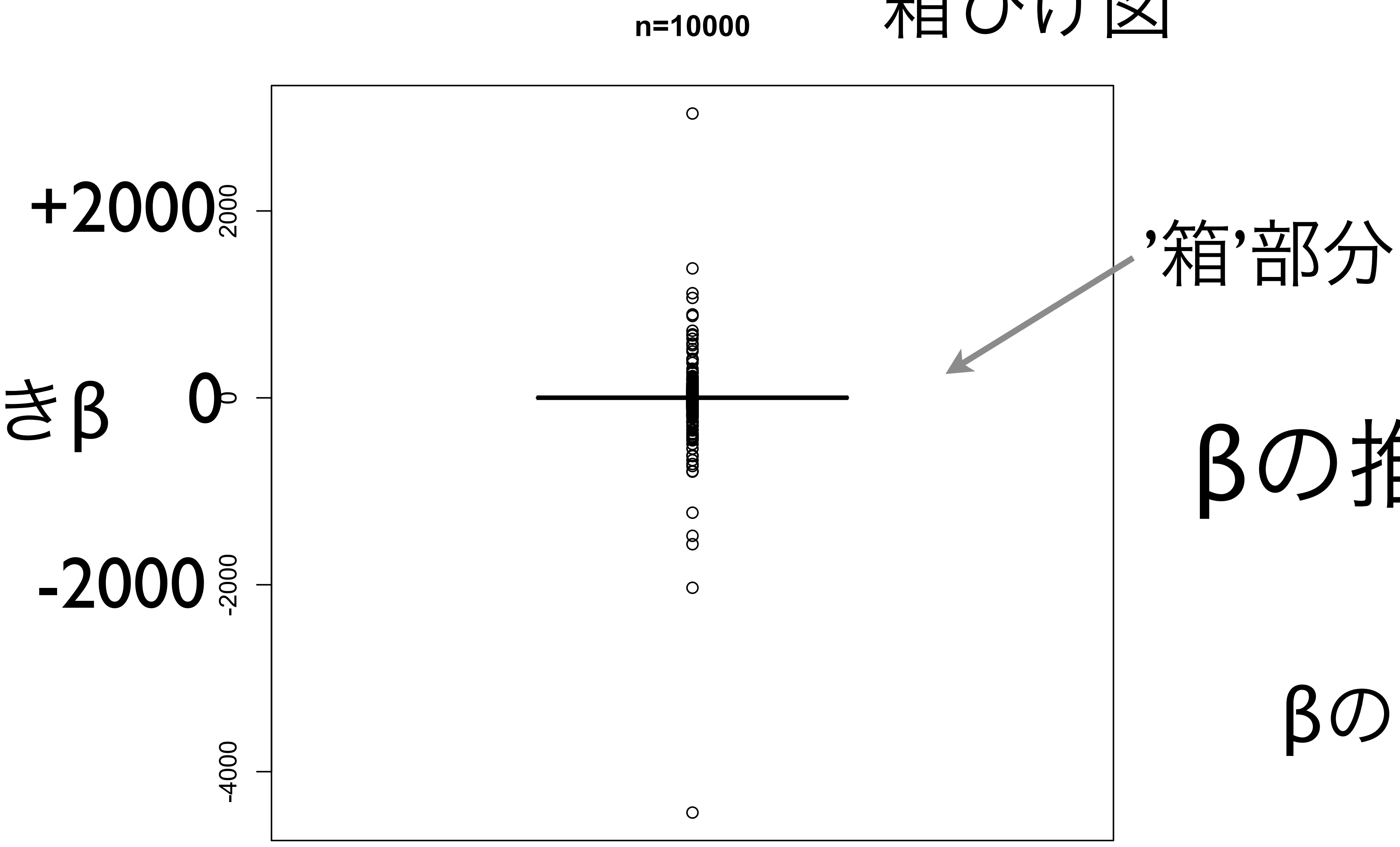
y　平均2x+3.5、分散1の正規分布





傾きβ

箱ひげ図



βの推定値の平均　0.674

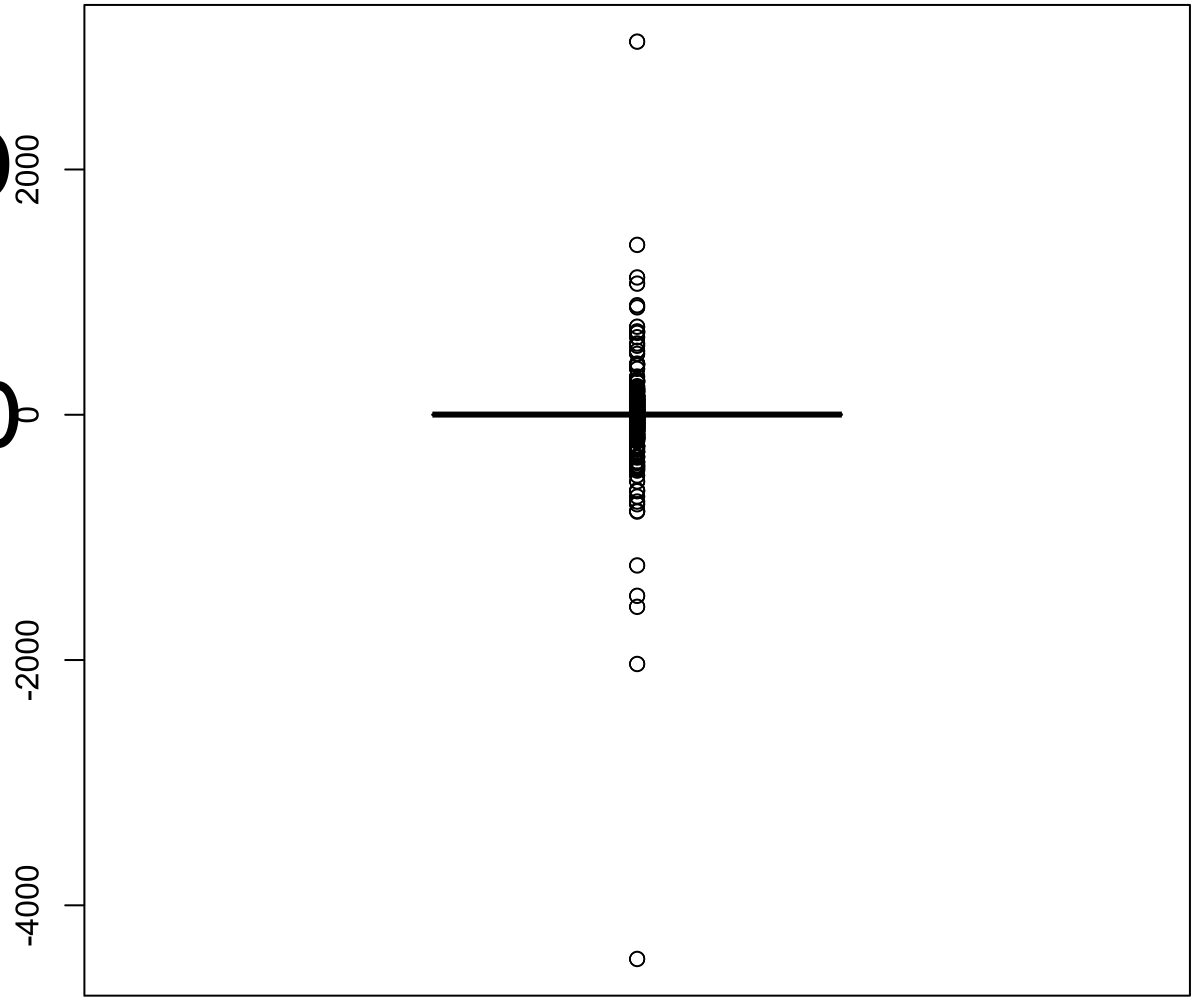
βの推定値の標準偏差　77.76

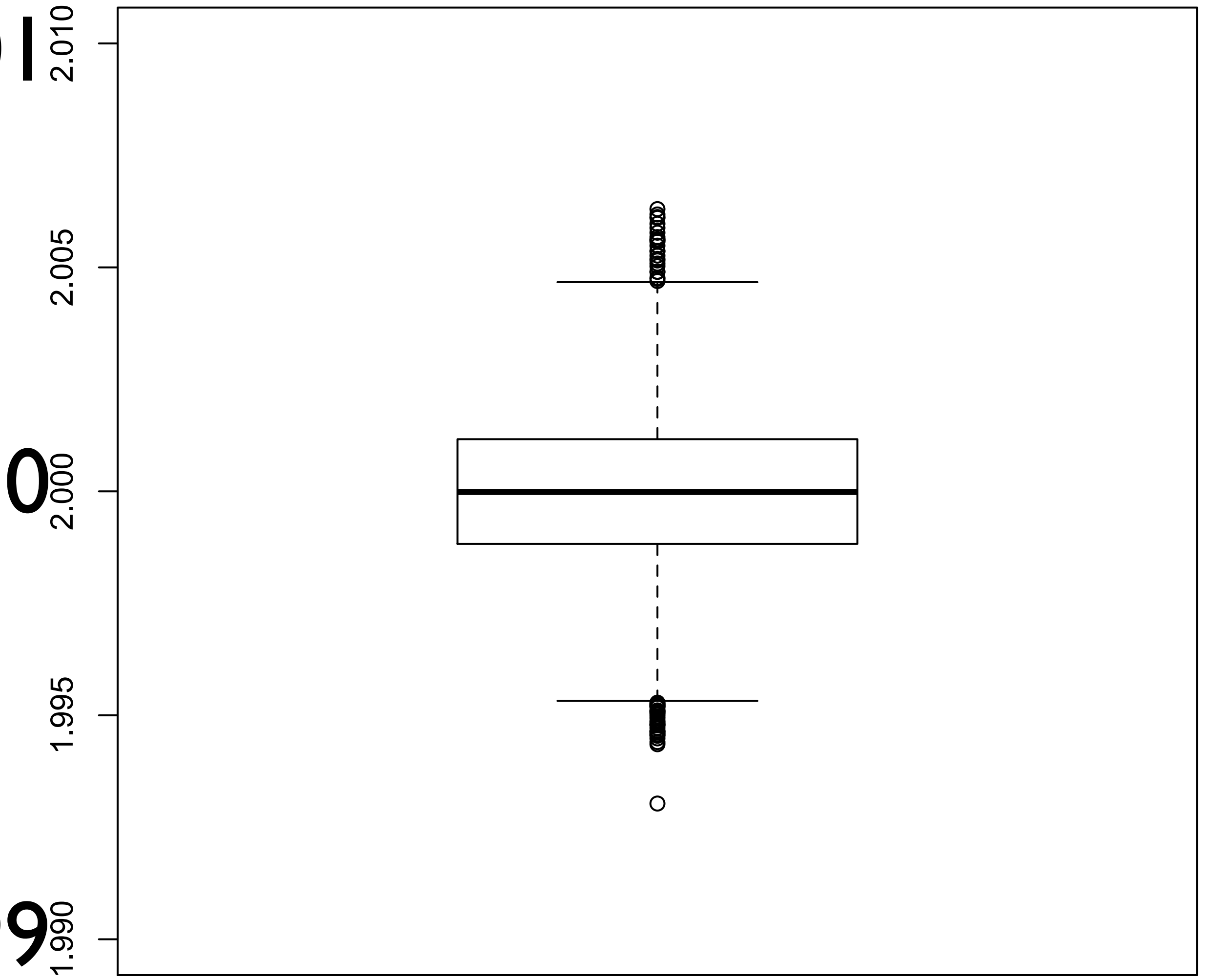
βは2.0付近に分布してもよさそうだが・・・

n=10000のサンプル10000個

$y \quad \dfrac{\text{平均}2x+3.5\text{、分散1の正規分布}}{\text{平均1、分散1の正規分布}}$

$y \quad$ 平均2x+3.5、分散1の正規分布





βの推定値の平均　0.674

βの推定値の標準偏差　77.76

βの推定値の平均　1.999994

βの推定値の標準偏差　0.001731

βは2.0付近に分布してもよさそうだが・・・

n=10000のサンプル10000個

正規分布 →割り算→ **正規分布の比**

よく使われる

よく知られている

伝統

老舗

魔物

”maverick”

正規分布する変数の比

単純な分布の「割り算」の結果

他とは質的にちがった分布
「ばらつきかた」が違う
データの量が増えることが'善'でない

グラフでさほど変に見えないこともある

サンプルサイズがかなり大きくないと
目立たないことがある

他の分布のつもりで分析→相当な過大評価

サンプルの平均値のような当然に見える推定量→良くない

回避 正規分布する変数の比

注意 正規分布っぽい量の比

計算 計算して分布をチェックしよう

nは変えて、大きいところまで

余談

正規分布する変数の比

1つの正規分布／（２つの正規分布の和）なども含む

他の分布

ガンマ分布する変数

対数正規分布する変数

切断正規分布する変数

説明変数が複数の場合

リンクがidentity以外の場合

母平均とともに母分散が変化するとき