

問. 正規分布 $N(0, 1^2)$, $N(0, 4^2)$, $N(3, 3^2)$ の3つのグラフ x 軸の区間を $[-5, 10]$ の範囲で1つのグラフに書きなさい。

正規分布のグラフを書く命令は

`curve(dnorm(x, {μの値}, {σの値}), {xの下限}, {xの上限})`

である。分散 σ^2 の代わりに標準偏差 σ を使っていることに注意して、

```
> curve(dnorm(x, 0, 1), -5, 10)
> par(new=T)
> curve(dnorm(x, 0, 4), -5, 10)
> par(new=T)
> curve(dnorm(x, 3, 3), -5, 10)
```

とすればよいが、このままだと y の範囲が揃わない (図1)。そこで、`ylim = c(0, 0.5)` をつけて y の範囲を揃えると

```
> curve(dnorm(x, 0, 1), -5, 10, ylim = c(0, 0.5))
> par(new=T)
> curve(dnorm(x, 0, 4), -5, 10, ylim = c(0, 0.5))
> par(new=T)
> curve(dnorm(x, 3, 3), -5, 10, ylim = c(0, 0.5))
```

y の範囲も揃ったグラフが描ける。また、ラベルを書きたくないときは、`xlab=""` や `ylab=""` をつければよい。

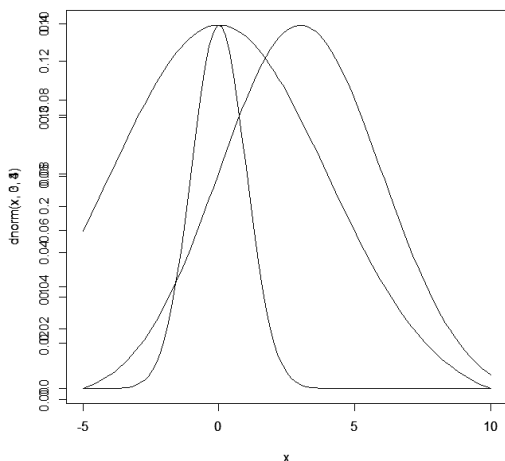


図 1: y 軸の揃っていないグラフ

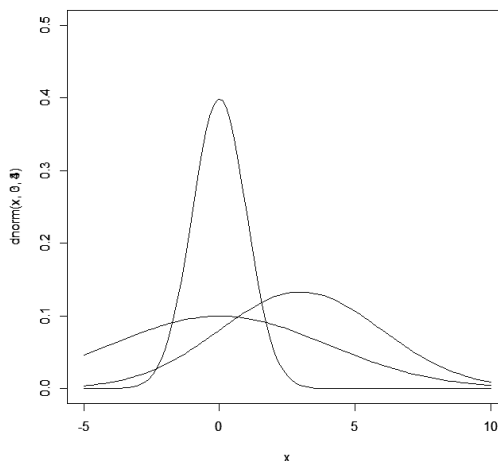


図 2: y 軸の揃っているグラフ

ちなみに次のようにするとグラフごとに色も付きわかりやすい。(各自で確認すること)

```
> curve(dnorm(x, 0, 1), -5, 10, ylim = c(0, 0.5), ylab="")
> par(new=T)
> curve(dnorm(x, 0, 4), -5, 10, ylim = c(0, 0.5), col="red", ylab="")
> par(new=T)
> curve(dnorm(x, 3, 2), -5, 10, ylim = c(0, 0.5), col="blue", ylab="正規分布のグラフ")
```

・ 標本平均と標本分散の定義

母集団から得られた n 個の標本 x_1, x_2, \dots, x_n を使って、次のように平均 μ , 分散 σ^2 を \bar{x} と s^2 を使って推定する。

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ s^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &\left(= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right) \quad \text{【教科書によってはこちらを使っている】} \end{aligned}$$

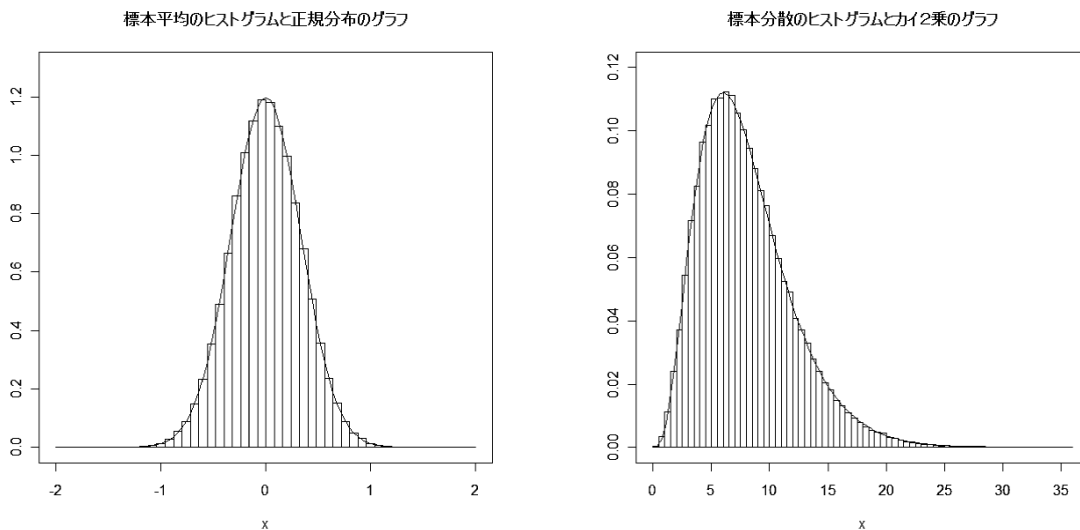
一般に、母集団から計算された平均・分散を母平均 (μ)・母分散 (σ^2) と呼び、標本から計算された平均・分散を標本平均 (\bar{x})・標本分散 (s^2) と呼び、記号も区別をしている。

・ 標本平均と標本分散の分布

母集団の分布が平均 μ , 分散 σ^2 の正規分布 $N(\mu, \sigma^2)$ に従うとき、 n 個の標本 x_1, x_2, \dots, x_n が得られたとき、標本平均と標本分散は次のような分布に従う。

- ・ 標本平均 \bar{x} は平均は同じ μ , 分散は n で割った $\frac{\sigma^2}{n}$ の正規分布 $N(\mu, \frac{\sigma^2}{n})$ に従う
- ・ 標本分散 s^2 は $\frac{n-1}{\sigma^2} s^2$ で変換すると自由度 $(n-1)$ の χ^2 分布に従う。

例. 標準正規分布から標本 x_1, \dots, x_9 を得た場合のヒストグラムと理論分布



標本平均の分散は n が大きくなると小さくなるので、推定がよくなることがわかる（分散が小さいと μ の近くにしか出現しない）。そのため、実験などではなるべく多くのデータを得るように努力をしている。標本平均だけでなく、一般にパラメータの推定はデータの個数が多いほうが良い。

・ 不偏推定量と最尤推定量

推定には必ず誤差が生じるため、推定の良し悪しを判断する基準が必要になる。一般的に用いられる基準としては下記の2つがよく用いられている。

・ 不偏推定量（期待値による基準）

推定量 $h(X)$ の期待される値として、期待値が次のように定義されている。

$$E[h(X)] = \begin{cases} \sum h(x_i) \times \Pr\{X = x_i\} & (\text{離散型の場合}) \\ \int_{-\infty}^{\infty} h(x) \times f(x) dx & (\text{連続型の場合}) \end{cases}$$

推定量の期待値が実際の値と一致している場合、不偏推定量（意識：偏りの無い推定量）と呼ぶ。実際に標本平均 \bar{x} の期待値を計算すると、

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] \\ &= \frac{1}{n} \sum_{i=1}^n \mu = \mu \end{aligned}$$

となるので、標本平均は母平均の不偏推定量である。さらに標本分散も母分散の不偏推定量となっている。

$$E[S^2] = E\left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2\right] = \sigma^2$$

・ 最尤推定量（尤度関数による基準）

母集団のパラメータが θ で確率密度関数が $f(x|\theta)$ とする。このとき、 n 個の標本 x_1, x_2, \dots, x_n が得られた場合の尤もらしさを表す値として、尤度関数が次のように定義されている。

$$L(\theta|x_1, \dots, x_n) = f(x_1|\theta) \times f(x_2|\theta) \times \dots \times f(x_n|\theta)$$

離散型の場合は確率密度関数 $f(x|\theta)$ のかわりに確率 $p(x|\theta)$ で計算するが、いずれにしても標本 x_i が現れた出現確率の積とみなすことができる（厳密に言えば連続型の場合、確率密度の積）。よって、この尤度関数の値が大きいほど分布の当てはまりが良いことを示している。そこで n 個の標本が得られた場合の尤度関数の値を最大にするような、パラメータ θ の推定量を最尤推定量と呼ぶ。正規分布を仮定した場合、標本平均 \bar{x} と n で割った標本分散が最尤推定量となっている。

ただし、最尤推定量と不偏推定量が常に同時に成立しているわけではないことに注意が必要である。実際正規分布を仮定した場合、母平均 μ の推定量 \bar{x} は不偏推定量かつ最尤推定量であるが、母分散 σ^2 の不偏推定量は $n-1$ で割った標本分散 s^2 であり、最尤推定量は n で割った標本分散である。実際、分散の最尤推定量の期待値は実際の値より少し小さい値となる。

また、不偏推定量になる推定量はいくつでも作ることができるが、推定量の分散が小さいほど推定の精度が良いことになるので、一般的には分散が最小になる推定量を選ぶ。

・ R を使った標本平均と標本分散の計算

R には平均や分散を計算する関数として下記が用意されている。

`mean(x)` : 標本平均 (母平均 μ の推定量)

`var(x)` : 標本分散 (母分散 σ^2 の推定量・不偏推定量)

`sd(x)` : 標本標準偏差 (母標準偏差 σ の推定量・`sqrt(var(x))`)

例 1. 平均と分散の計算

5 個の標本 15, 13, 21, 8, 18 が与えられた場合

1) `x` に標本を代入する

```
> x <- c(15,13,21,8,18)
```

2) 標本の確認

```
> x
```

```
[1] 15 13 21 8 18
```

3) 平均と分散の計算

```
> mean(x)
```

```
[1] 15
```

```
> var(x)
```

```
[1] 24.5
```

つまり、標本平均 $\bar{x} = 15$ 、標本分散 $s^2 = 24.5$ であることがわかる。

例題. 次の 10 個の標本の平均と分散を求めよ。

10.1, 6.1, 12.3, 9.6, 9.0, 7.1, 6.8, 12.1, 14.3, 12.6

例と同じように `x` にデータを入力して標本平均 \bar{x} 、標本分散 s^2 、標本標準偏差 s を求めるとそれぞれ

$$\bar{x} = 10.0$$

$$s^2 = 7.775556 = 7.78$$

$$s = 2.788468 = 2.79$$

となる。

例 2. 乱数を使った平均・分散の計算 (不偏推定量の確認)

R にはいくつかの基本的な分布に従った乱数をつくるのが簡単にできる。そこで、何回か平均・分散を計算して、真の値の近くなるかどうか確認してみよう。

1) 母平均 $\mu = 0$ 、標準偏差 $\sigma = 1$ の正規分布 $N(0, 1^2)$ に従う 10 個の標本を代入

```
> x <- rnorm(10,0,1)
```

2) 標本の確認 (乱数なので実際には異なる値になる)

```
> x
```

```
[1] -0.70414983 -1.79536612 0.08472537 -0.34633101 0.05832257
```

```
[6] 2.01385432 0.69930744 0.44551811 0.07515012 0.63905958
```

3) 平均と分散の計算 (乱数なので実際には異なる値になる)

```
> mean(x)
```

```
[1] 0.1170091
```

```
> var(x)
```

```
[1] 0.9855502
```

もう一度、1) を行うと違う標本になるので、平均や分散も異なる値になる。しかしながらほとんどの場合、`mean(x)` は0の近くの値を取り、`var(x)` は1の近くの値を取ることがわかる。

実際の例. 標本平均の場合 (小数第4位を四捨五入)

```
0.429, -0.720, -0.184, -0.255, 0.497, 0.148, -0.085, -0.130, 0.282, 0.215
```

今日の問題

正規分布 $N(4, 2^2)$ に従う乱数を10個発生させて、その平均と分散を計算する

0) 学籍番号を使って乱数を初期化 `set.seed(学籍番号)` 例 `set.seed(1025001)`

1) `x <- rnorm(10,4,2)` で乱数を発生させ、標本平均と標本分散を計算する。

2) 1) を10回繰り返し、解答用紙に結果を書く (小数第3位を四捨五入)。

1回目 $\bar{x}_1 = 4.51$, $s_1^2 = 3.50$

2回目 $\bar{x}_2 = 4.12$, $s_2^2 = 4.79$

⋮

10回目 $\bar{x}_{10} = 3.66$, $s_{10}^2 = 2.19$

2) 10回分の標本平均の $\bar{x}_1, \dots, \bar{x}_{10}$ の標本平均 $\bar{\bar{x}}$, 標本分散 $s_{\bar{x}}^2$ と

10回分の標本分散 s_1^2, \dots, s_{10}^2 の標本平均 \bar{s}^2 , 標本分散 $s_{s^2}^2$ を計算して結果を書く

標本平均: 標本平均 $\bar{\bar{x}} = 3.74$, 標本分散 $s_{\bar{x}}^2 = 0.446$

標本分散: 標本平均 $\bar{s}^2 = 3.61$, 標本分散 $s_{s^2}^2 = 2.02$

ヒント: \bar{x} の標本平均は4.00前後、 s^2 の標本平均は4.00前後の値になります。

ただし、 s^2 の平均はバラツキが大きいです。

プログラム例 (学籍番号1325001の場合)

```
> set.seed(1325001)
```

```
> x <- rnorm(10,4,2)
```

```
> c(mean(x), var(x))
```

```
[1] 3.557622 2.246756
```

```
> x <- rnorm(10,4,2)
```

```
> c(mean(x), var(x))
```

```
[1] 2.785827 3.415087
```

```
⋮
```

・ 標本平均と標本分散

正規分布 $N(4, 2^2)$ に従う乱数を 10 個発生させて、その平均と分散を計算せよ。

	標本平均	標本分散
1回目		
2回目		
3回目		
4回目		
5回目		
6回目		
7回目		
8回目		
9回目		
10回目		
10回の平均		
10回の分散		

2019年度神奈川工科大学 数理統計学 演習問題	学科	学年	組	学 籍 番 号	氏 名	

提出先：K3-3309号室前 20番のボックス 提出期限：10月28日（月）17時頃まで