情報科学 第13回
正規分布とt分布

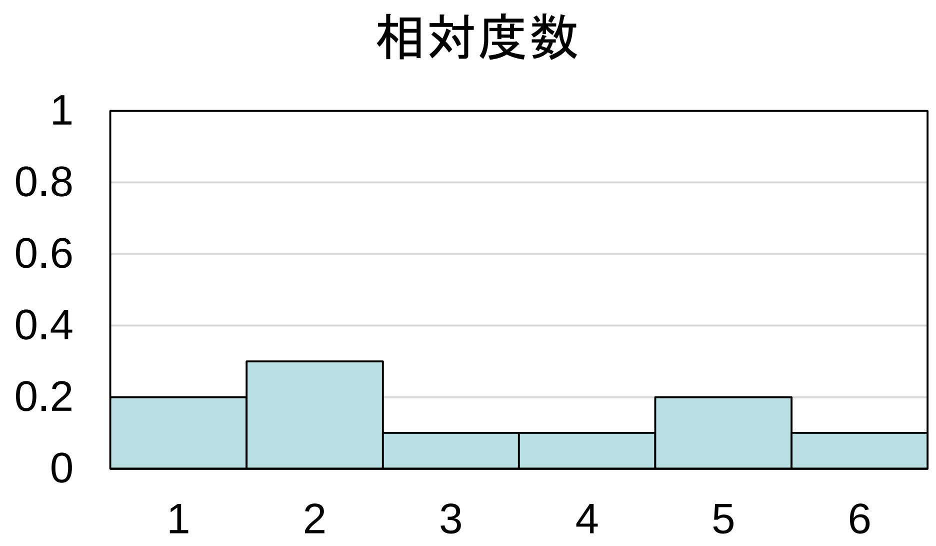
幸谷智紀
tkouya@ym.sist.ac.jp

本日の内容

- 正規分布･･･連続的確率分布
- 標準正規分布･･･正規分布の基本形
- Studentのt-分布

確率分布＝相対度数のヒストグラム

さいころの目	度数（出た回数）	相対度数（度数／振った回数）
1	3	0.3
2	1	0.1
3	2	0.2
4	3	0.3
5	0	0
6	1	0.1



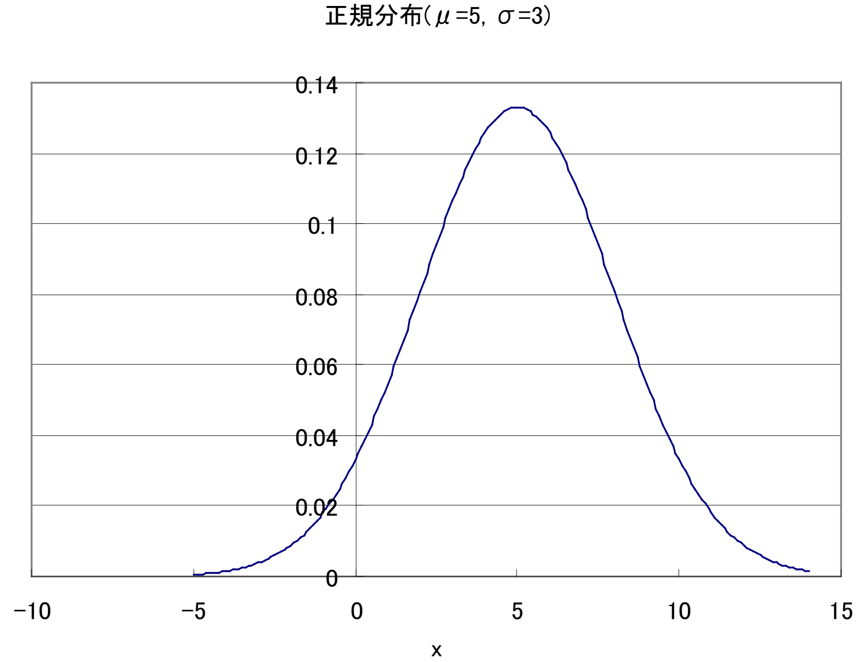
- 相対度数＝度数／全体の試行回数

→相対度数を「確率」と呼ぶ

→相対度数のヒストグラムを「確率分布」と呼ぶ

正規分布・・・連続的確率分布

- 平均μ, 標準偏差σで平均値に集中する分布
- 試験の得点など



$\mathrm{N}(\mu, \sigma^2)$ と表記する。

確率変数 ： X

確率密度関数 ： $$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

平均 ： $E(X) = \mu$

分散 ： $\sigma^2(X) = \sigma$

Excelで正規分布を体感する

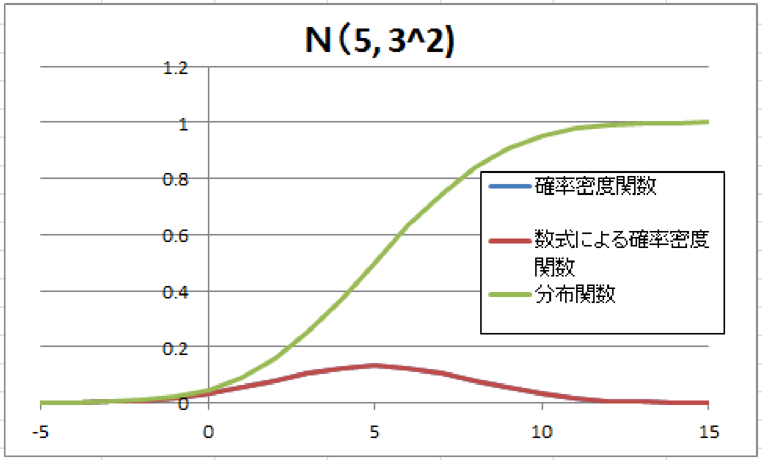
- **NORMDIST**(x, μ, σ, TRUE or FALSE)
  - μ **・・・成功数・・・ 試行回数**に含まれる成功の回数を指定
  - σ **・・・試行回数・・・**独立試行の回数を指定
  - **TRUE or FALSE・・・関数形式・・・**
  - TRUE を指定した場合は分布関数を計算
  - FALSE の場合は確率密度を計算

課題13A

1. μ=5, σ=3の時の正規分布分布の確率密度関数p(-5)～p(15)と分布関数の値をそれぞれ計算し，表とグラフをExcelファイル「課題13」の「課題13A」シートに作れ。

2. 確率密度の式(1)に基づいて，確率密度関数値を計算し，1の値と一致することを確認せよ。

1	平均	5		
2	標準偏差	3		
3				
4	x	確率密度関数	数式による確率密度関数	分布関数
5	-5	0.000514093	0.000514093	0.000429
6	-4	0.001477283	0.001477283	0.00135
7	-3	0.003798662	0.003798662	0.00383
8	-2	0.00874063	0.00874063	0.009815
9	-1	0.017996989	0.017996989	0.02275
10	0	0.033159046	0.033159046	0.04779
11	1	0.054670025	0.054670025	0.091211
12	2	0.080656908	0.080656908	0.158655
13	3	0.106482669	0.106482669	0.252493
14	4	0.125794409	0.125794409	0.369441
15	5	0.13298076	0.13298076	0.5
16	6	0.125794409	0.125794409	0.630559
17	7	0.106482669	0.106482669	0.747507
18	8	0.080656908	0.080656908	0.841345
19	9	0.054670025	0.054670025	0.908789
20	10	0.033159046	0.033159046	0.95221
21	11	0.017996989	0.017996989	0.97725
22	12	0.00874063	0.00874063	0.990185
23	13	0.003798662	0.003798662	0.99617
24	14	0.001477283	0.001477283	0.99865
25	15	0.000514093	0.000514093	0.999571



確率変数の標準化

X を $\mathbf{N}(\mu, \sigma^2)$ に従う確率変数とする時，

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

と変換した Z は $\mathbf{N}(0, 1)$ に従う。

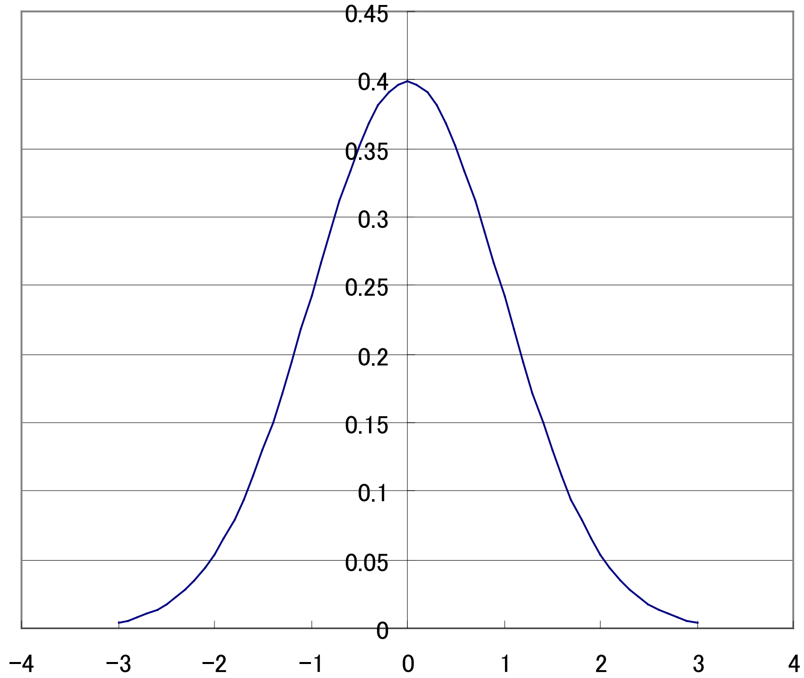
同様にして，X を $\mathbf{N}(\mu, \sigma^2/n)$ に従う確率変数とする時，

$$Z = \frac{X - \mu}{\sqrt{\sigma^2/n}} = \frac{X - \mu}{\sigma/\sqrt{n}}$$

と変換した Z も $\mathbf{N}(0, 1)$ に従う。

標準正規分布・・・連続的確率分布

- 平均μ=0，標準偏差σ=1の正規分布



$\mathrm{N}(0,1)$ と表記する。

確率変数 ： X

確率密度関数 ： $$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{1}$$

平均 ： $E(X) = 0$

分散 ： $\sigma^2(X) = 1$

Excelで標準正規分布を体感する

- **NORMSDIST**(x)・・・標準正規分布の分布関数
- **NORMSINV**(z)・・・標準正規分布の逆関数

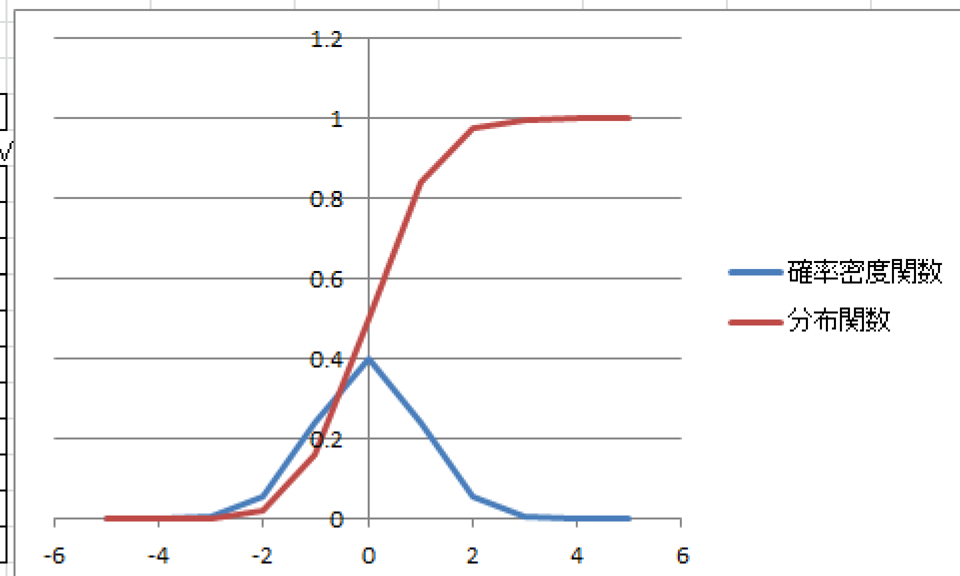
分布関数　：　$\Phi(x_\beta) = \int_{-\infty}^{x_\beta} p(x)dx = \beta$

$$x_\beta = \Phi^{-1}(\beta)$$

課題13B

1. μ=0, σ=1の時の標準正規分布の確率密度関数p(-5)～p(5)と分布関数の値をそれぞれNORMDIST関数を用いて計算せよ。
2. NORMSDIST関数を用いて分布関数の値を求め，1と同じになることを確認せよ。
3. NORMSINV関数を用いて，2の値からxの値を求められることを確認せよ。

	A	B	C	D	E
1	平均	0			
2	標準偏差	1			
3					
4		NORMSDIST	NORMDIST	NORMSDIST	
5	x	確率密度関数	分布関数	確率密度関	NORMSINV
6	-5	1.48672E-06	2.86652E-07	2.87E-07	-5
7	-4	0.00013383	3.16712E-05	3.17E-05	-4
8	-3	0.004431848	0.001349898	0.00135	-3
9	-2	0.053990967	0.02275013	0.02275	-2
10	-1	0.241970725	0.158655254	0.158655	-1
11	0	0.39894228	0.5	0.5	0
12	1	0.241970725	0.841344746	0.841345	1
13	2	0.053990967	0.977249868	0.97725	2
14	3	0.004431848	0.998650102	0.99865	3
15	4	0.00013383	0.999968329	0.999968	4
16	5	1.48672E-06	0.999999713	1	5
17					



Studentのt-分布

X を $\mathbf{N}(\mu, \sigma^2/n)$ に従う確率変数とする時，

$$Z = \frac{X-\mu}{\sqrt{\sigma^2/n}} = \frac{X-\mu}{\sigma/\sqrt{n}}$$

と変換した Z も $\mathbf{N}(0,1)$ に従う。ならば，分散 σ^2 を標本分散 s^2（標本数 n) で置き換えた時の

$$t = \frac{X-\mu}{\sqrt{s^2/n}} = \frac{X-\mu}{s/\sqrt{n}}$$

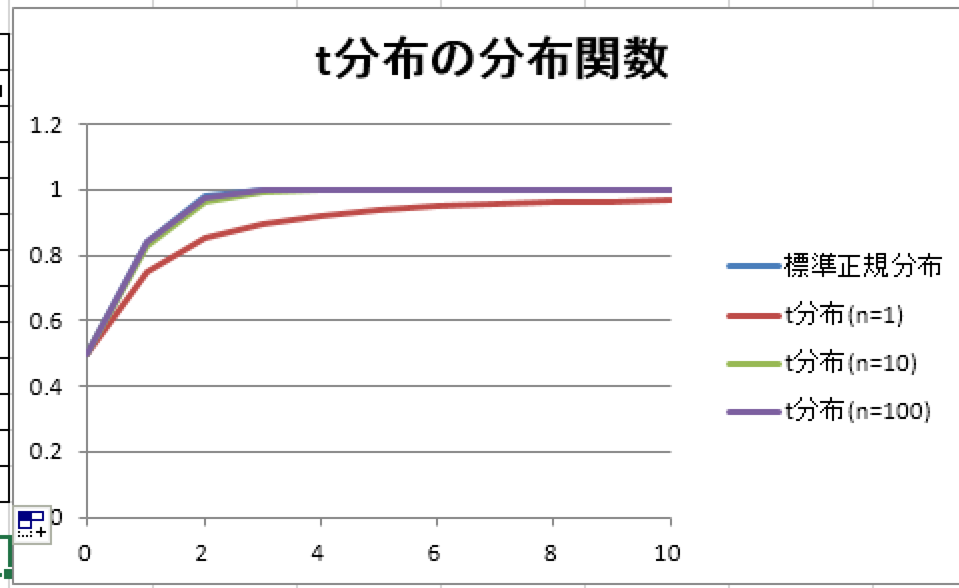
と変換した t は？$\longrightarrow$ **Student** の **t-**分布（自由度 $n-1$）に従う。

TDIST(*x, 自由度, 1*)･･･t-分布の1-分布関数
TINV(*確率, 自由度*)･･･t-分布の逆関数

課題13C

- X=0, 1,…, 10における標準正規分布の値F(x)とt-分布の分布関数の値を比較せよ。自由度は1, 10, 100とすること。

		t分布の自由度		
x	標準正規分布	1	10	100
0	0.5	0.5	0.5	0.5
1	0.841344746	0.75	0.829553	0.840138
2	0.977249868	0.852416	0.963306	0.975894
3	0.998650102	0.897584	0.993328	0.998296
4	0.999968329	0.922021	0.998741	0.999939
5	0.999999713	0.937167	0.999731	0.999999
6	0.999999999	0.947432	0.999934	1
7	1	0.954833	0.999981	1
8	1	0.960417	0.999994	1
9	1	0.964777	0.999998	1
10	1	0.968274	0.999999	1



中心極限定理

- 中心極限定理・・・区間推定・検定の基本定理

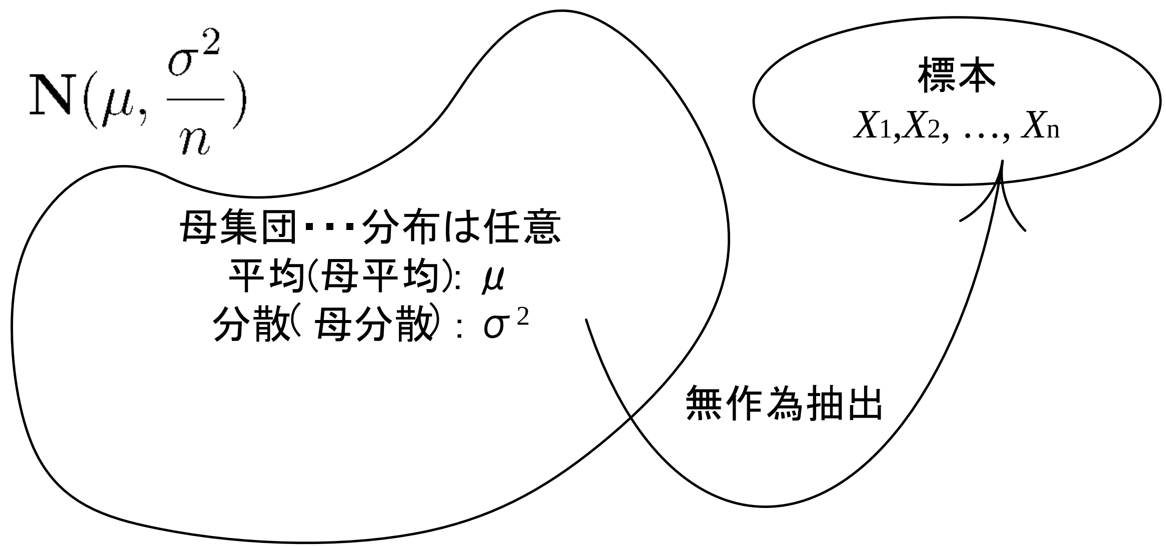
　母集団から n 個の標本 $X_1, X_2, \ldots, X_n$ を無作為集出する。この時，標本平均 $\overline{X}$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

は，母集団の分散が何であれ，近似的に，正規分布

$$\mathbf{N}(\mu, \frac{\sigma^2}{n})$$

に従う。



中心極限定理の意味づけ

母集団から n 個のデータ x_1, x_2, ..., x_n を無作為抽出する

$\downarrow$

無限個の取り出し方は，n 個の確率変数にまとめられる

$$\begin{aligned} X_1 &= x_1, x_1', x_1'', \ldots \\ X_2 &= x_2, x_2', x_2'', \ldots \\ &\vdots \\ X_n &= x_n, x_n', x_n'', \ldots \end{aligned}$$

$\downarrow$

この n 個の確率変数の平均

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

は，母分散の平均 μ を持ち，分散が σ^2/n である正規分布に「近い」分散を示す。

$||$

$n \longrightarrow \infty$ とすれば，母平均の平均付近に $\overline{X}$ は集中する。

課題１３D

- サイコロを10回(n=10), 20回(n=20), ・・・，50回(n=50)振って（RANDBETWEEN関数を使うこと），目の和の平均を求める。この時，平均の分散がだんだん1/nに比例して小さくなり，平均は3.6666・・・に近づくことを確かめよ。

（注意！） 10回標本を10回以上，20回標本を10回以上・・・取得してその度ごとに平均を求めて分散を出すこと。