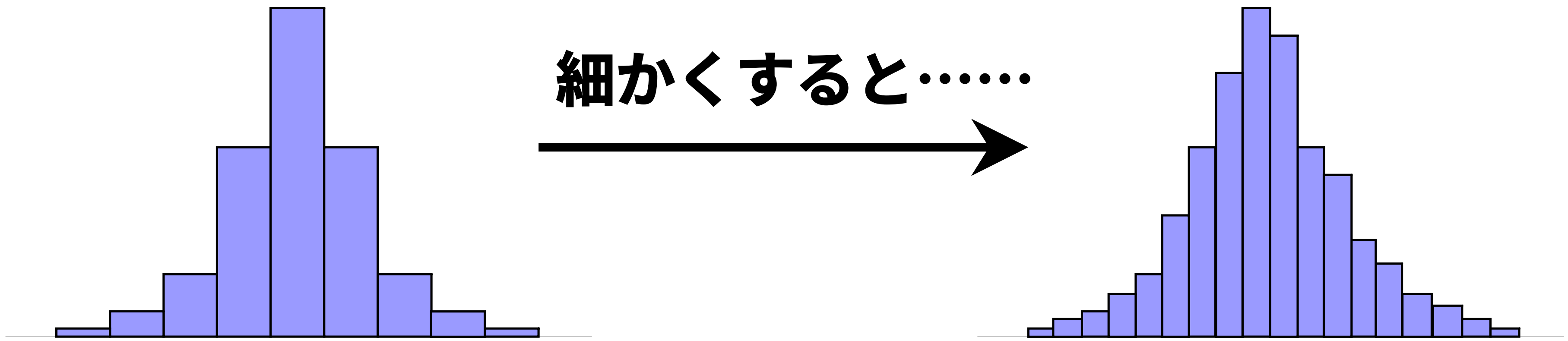
比較現代日本論研究演習/現代日本論演習 (田中重人)

第 1 回「確率密度と理論分布」(2004.10.7)

1. ヒストグラムと確率密度
2. 確率の理論分布
3. 正規分布とそのファミリー
4. 確率密度関数の使いかた

【ヒストグラム】

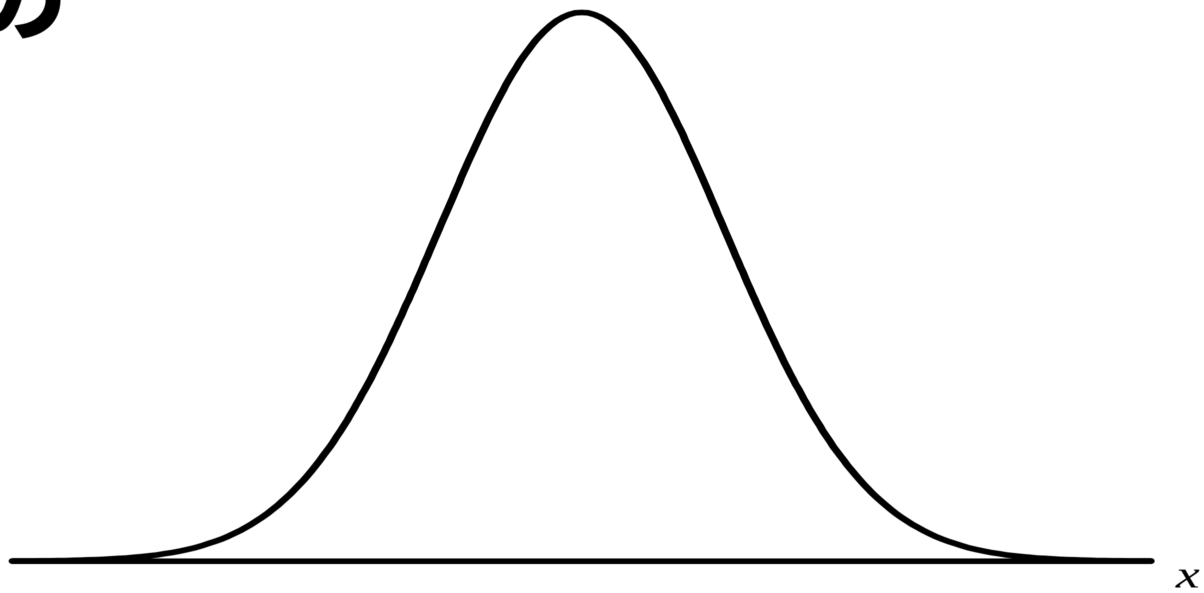
連続量を階級分けして度数分布を示したもの



【確率密度のグラフ】

Probability density

連続量に対応して、連続的に変化する確率を表したもの



【確率の理論分布】

特定の仮定から
理論的に導出された確率の分布

例：硬貨を投げるとき

表が出る　→

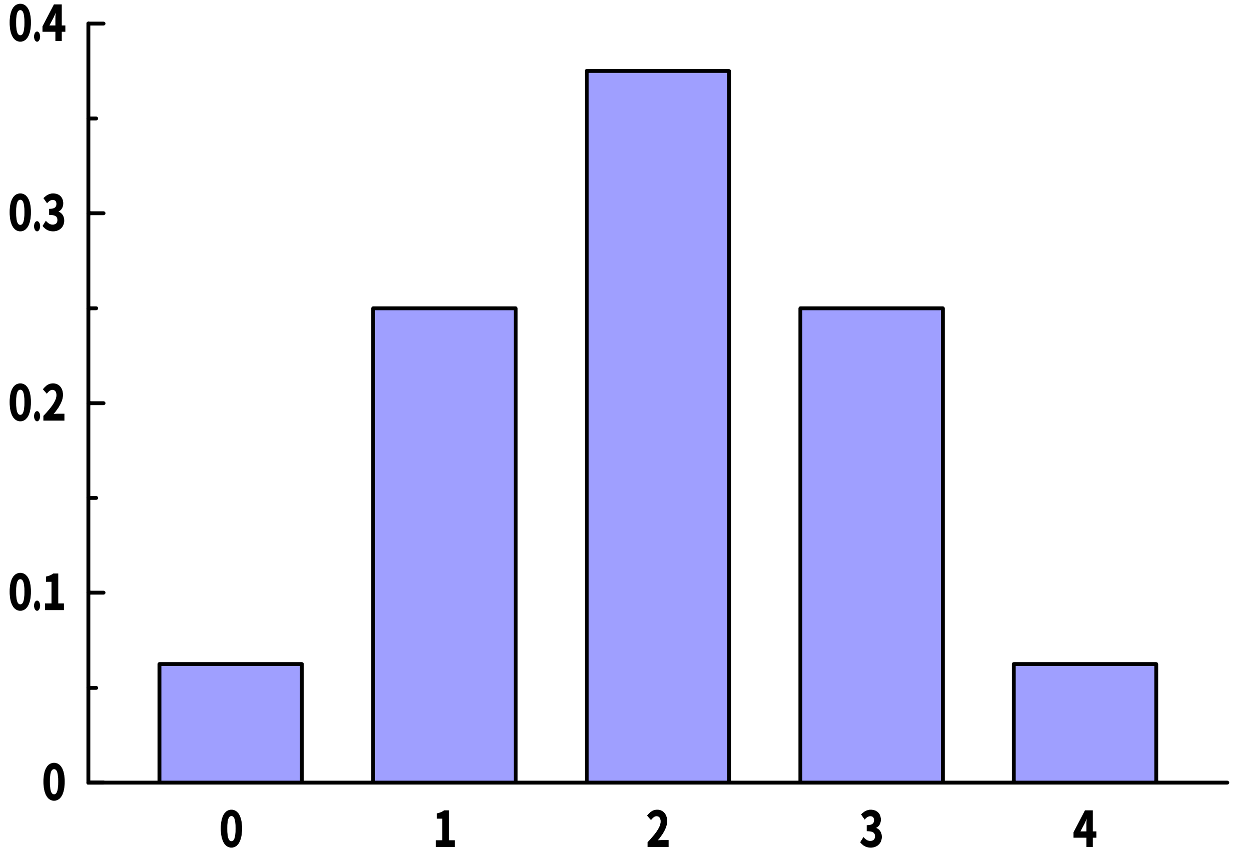
裏が出る　→

【2 項分布】

硬貨を n 回投げる。
表が出る回数を x とする。

$n = 4$ のとき、 x はどのような値をどのような確率でとるか?

n = 4, 確率= 0.5 の 2 項分布



【期待値】

Expected value

値 (x) **に確率** (p) **を掛けたものの総和：**

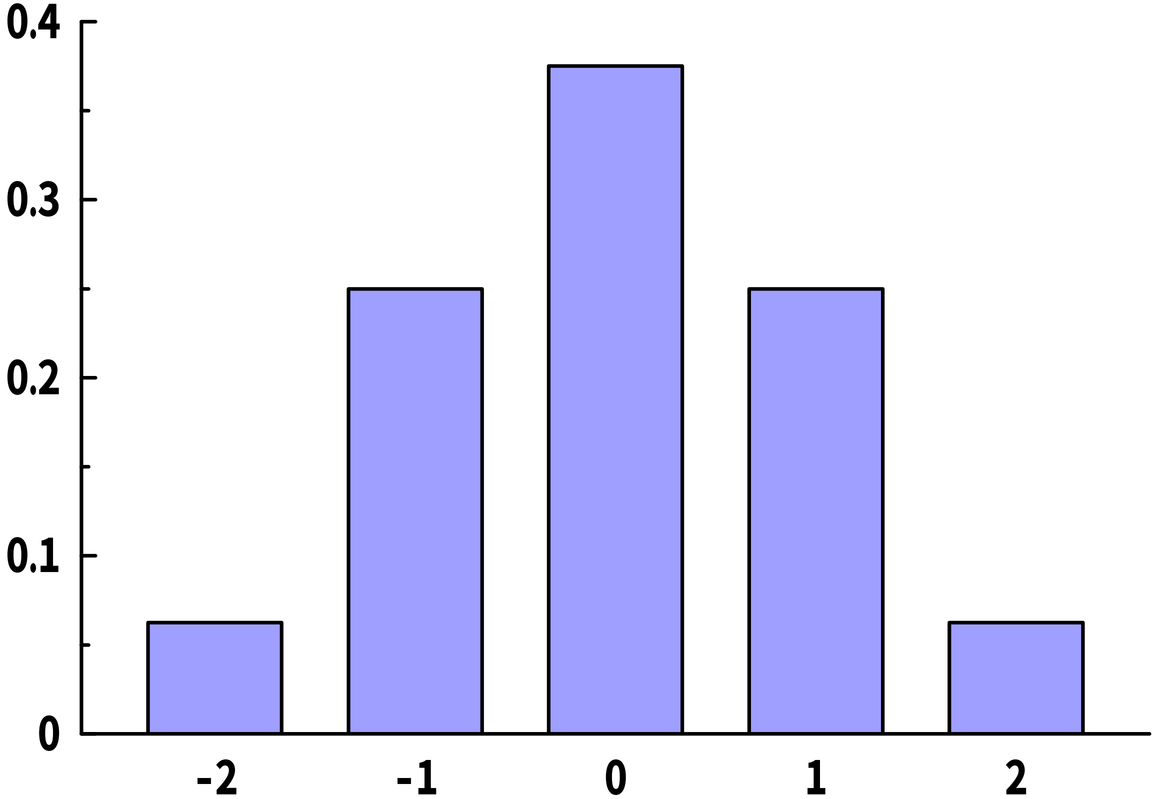
$$E = \sum (x \times p)$$

※「平均値」と呼ばれることもある

$n = 4$ の 2 項分布の期待値は?

【標準化】

$$Z = (x - E)\frac{2}{\sqrt{n}}$$ に変換すると



【標準正規分布】

Standard normal distribution

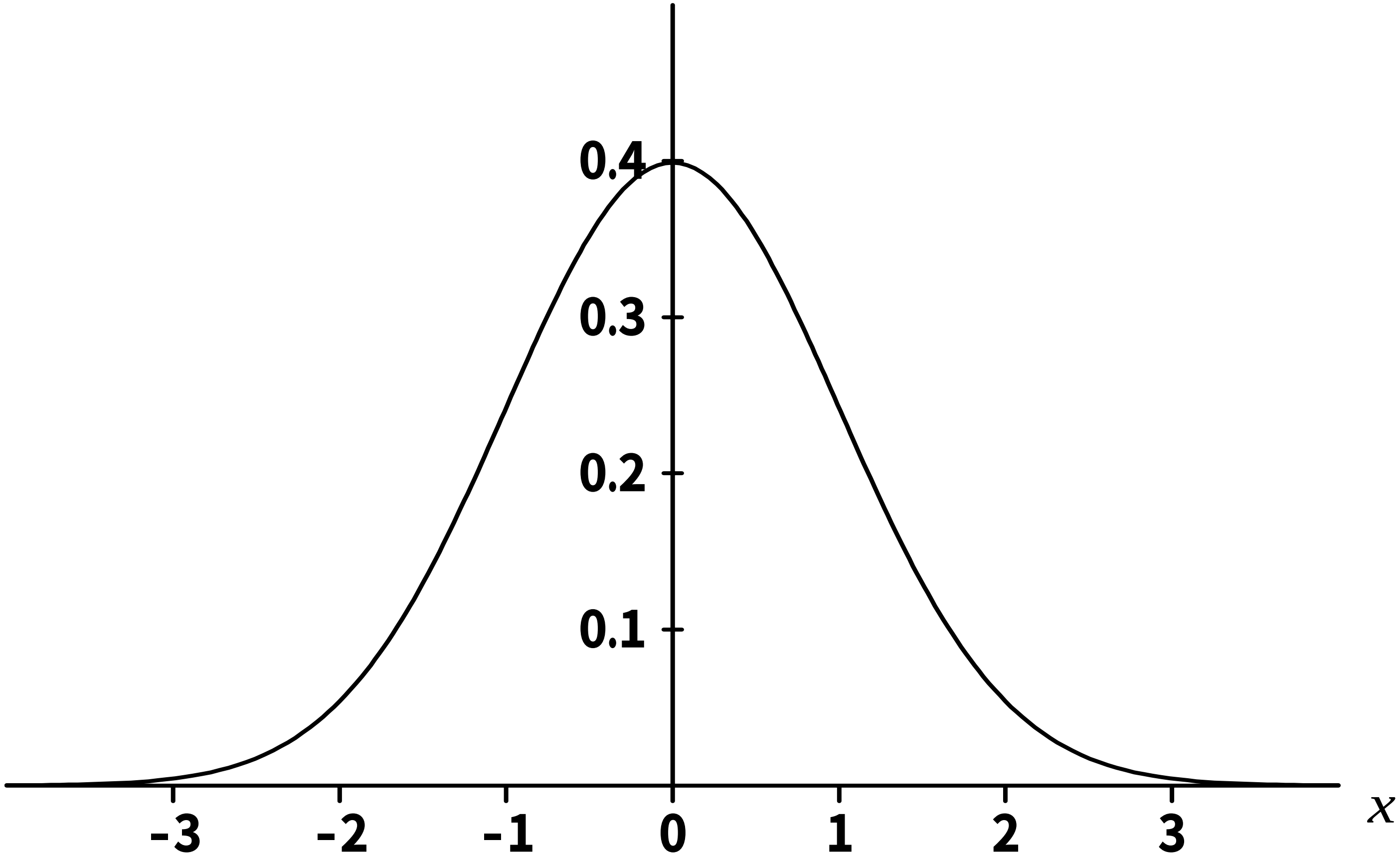
n が大きければ、Z は

標準正規分布の確率密度関数

$$\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{x^2}{2}\right]$$

で近似できる

標準正規分布の確率密度のグラフ：



※ 標準正規分布に定数による加減乗除を加えたものを総称して「正規分布」(normal distribution) という

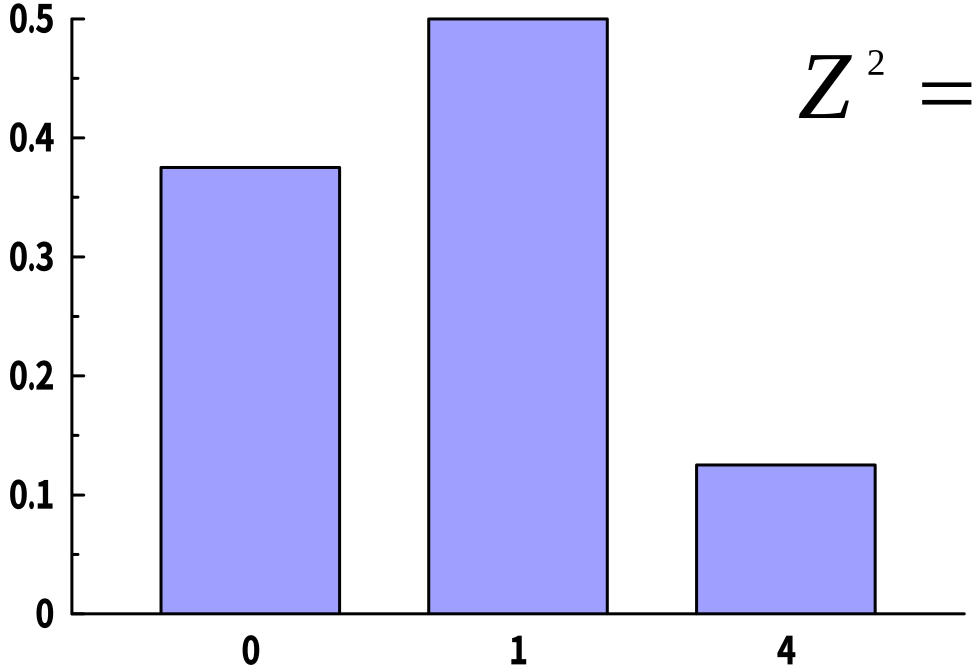
※ 0.5 以外の確率による 2 項分布でも、適当な標準化を行って n を増加させると正規分布に近づく

【正規分布の応用上の意義】

偶然による現象の生起確率や、
その組み合わせで決まる物事は、
正規分布（またはそのファミリー）
で近似できることが多い

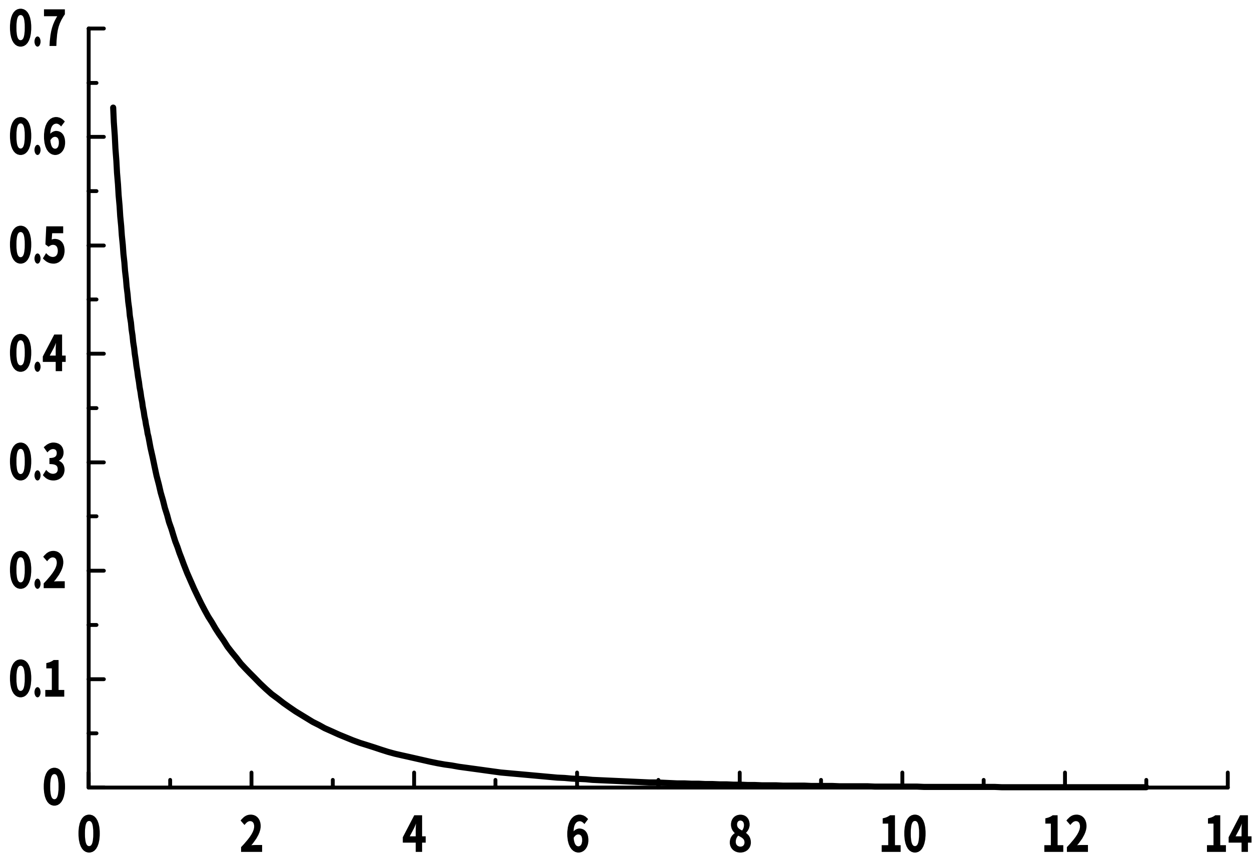
【自由度１のχ^2分布】

先に標準化した変数の２乗を考える：



$$Z^2 = \left((x - E)\frac{2}{\sqrt{n}} \right)^2$$

n が増加すると、Z^2 の確率分布は
自由度 1 の χ^2 分布に近づく



【χ²分布の一般形】

硬貨を n 回投げる作業を c 回繰り返す。

それぞれについて表が出る回数 x_i を数え、それを標準化して２乗して総和を求める：

$$\chi_c^2 = \sum_{i=1}^{c}\left((x_i - E)\frac{2}{\sqrt{n}}\right)^2$$

n が大きければ、自由度 c の χ² 分布に近似

【χ^2分布の応用上の意義】

期待値や平均値からのずれを予測するときに使う

【t 分布】

(1) **硬貨を** n **回投げる作業を** 1 **回おこない、表が出た回数を** x **とする**

(2) **硬貨を** n **回投げる作業を** d **回繰り返し、それぞれについて表が出る回数** y_j (j=1...d)**を数える**

このとき

$$t_d = (x - E)\sqrt{\frac{d}{\sum_{j=1}^{d}\left(y_j - E\right)^2}}$$

の確率分布は、n が大きければ、
自由度 d の t 分布で近似できる。

d が大きければ、標準正規分布で近似できる。

【t 分布の応用上の意義】

平均値とそこからのずれの両方を同時に予測するときに使う

【F 分布】

(1) **硬貨を** n **回投げる作業を** c **回おこない、それぞれについて表が出る回数** x_i (i=1 ... c)**を数える**

(2) **硬貨を** n **回投げる作業を** d **回繰り返し、それぞれについて表が出る回数** y_j (j=1 ... d)**を数える**

このとき

$$F_{(c,d)} = \frac{\sum_{i=1}^{c}(x_i - E)^2}{\sum_{j=1}^{d}(y_j - E)^2} \times \frac{d}{c}$$

の確率分布は、*n* が大きければ、

自由度(*c*, *d*) の *F* 分布で近似できる。

※ $\sqrt{F_{(1,d)}} = t_d$ である。

※また、d が大きければ、$F_{(c,d)}$ は χ^2_d に近似する。

【F 分布の応用上の意義】

平均値からのずれの大きさを
比較するときに使う

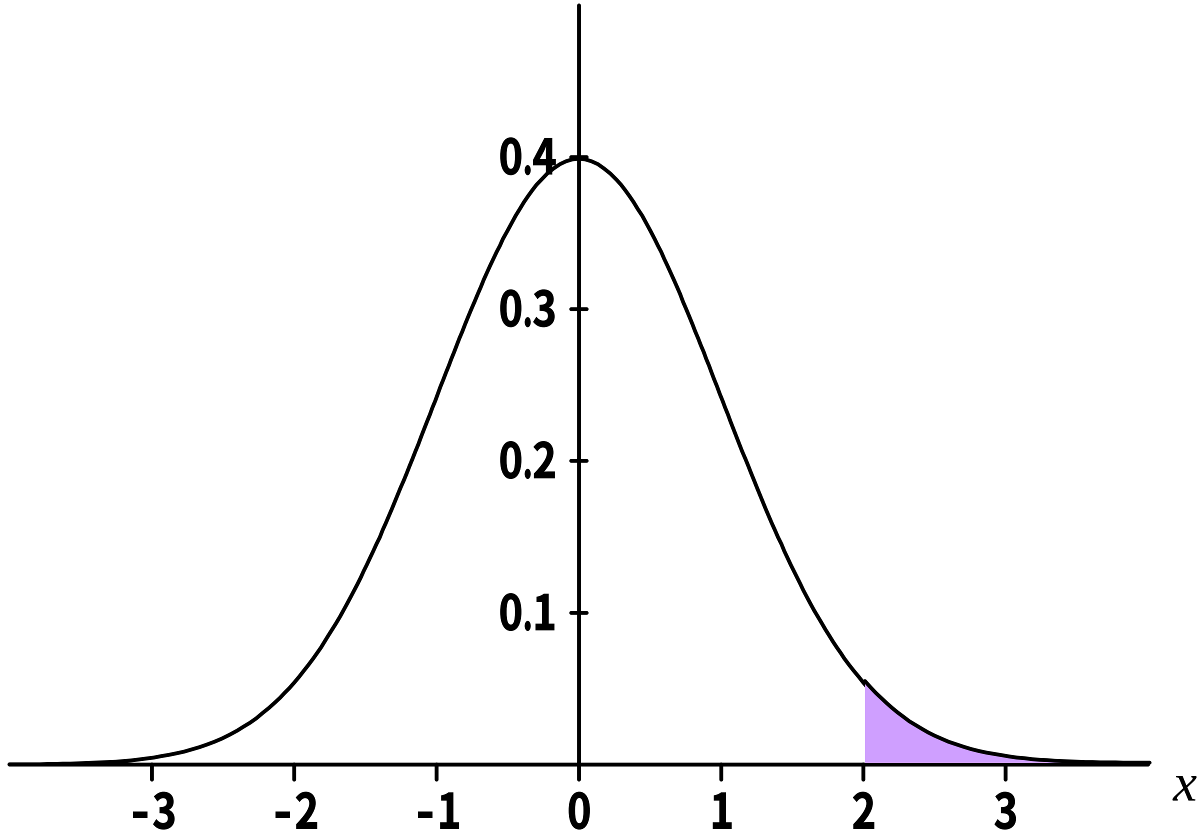
【確率密度関数の使いかた】

積分して、特定の区間内の値をとる確率

を求める。

例：標準正規分布において、2 以上の値をとる確率

$$\int_{2}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[- \frac{x^2}{2} \right]$$


0.4
0.3
0.2
0.1
-3
-2
-1
0
1
2
3
x

実際には、

$$0.05 = \int_{v}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[- \frac{x^2}{2} \right]$$

のような方程式を (v について) 解いて

臨界値 (critical value) を求めることが多い

【宿題】

別紙の乱数表から、1 桁の数字を 5 個と 50 個抜き出し、それぞれ平均値を求めよ。

乱数表の使いかた：

(1) **適当なところに目をつける**

(2) **適当な方向 (上・下・右・左・斜めなど)**
へ移動しながらひとつずつ数字を拾う

【文献】

宮川公男（1999）**『基本統計学［第 3 版］』有斐閣。**