

確率統計入門

山崎 輝

法政大学 経営学部 ゼミ資料

内容

1. 基礎概念
2. 確率変数と期待値
3. 分散と標準偏差
4. 共分散と相関係数
5. 回帰分析

基礎概念

確率論の基礎概念

【標本空間・根元事象・事象】

- ❖ 確率論の出発点は**試行**

試行 (trial) . . . 前もって結果の決まっていない現象の観察や実験

- ❖ ただし、**起こりうる結果の全体は前もって分かっている**とし、この結果全体の集合を**標本空間** (sample space) という

- ❖ 1つ1つの結果を**根元事象** (elementary event) という

- ❖ 根元事象の集まり (標本空間の部分集合) を**事象** (event) という

サイコロの例 (1)

- ❖ **試行**：サイコロを1回振る
- ❖ **標本空間**： $\{1, 2, 3, 4, 5, 6\}$
- ❖ **根元事象**： $1, 2, 3, 4, 5, 6$
- ❖ **事象**：
 - ✓ 1の目が出る事象は $\{1\}$
 - ✓ 2の目が出ない事象は $\{1, 3, 4, 5, 6\}$
 - ✓ 4か6の目がでる事象は $\{4, 6\}$
 - ✓ 偶数の目が出る事象は $\{2, 4, 6\}$

例題

例題：コインを2回投げて表と裏をみる試行をする。このとき以下の問いに答えよ。

- ① この試行の標本空間と根元事象を書け
- ② 1回目が表、2回目が裏となる事象を書け
- ③ 少なくとも1回は表が出る事象を書け

確率

【確率とは？】

- ❖ **確率** (probability) とは、試行したときの**事象**の起こりやすさを表す数値
- ❖ 事象 A が起こる確率を $p(A)$ と書くことにする
- ❖ 確率 p は以下の3つの性質をみたす
 - 性質1.** $p(\text{標本空間}) = 1$
 - 性質2.** どんな事象 A に対しても $0 \leq p(A) \leq 1$
 - 性質3.** 互いに排反な2つの事象 A と B に対して $p(A \cup B) = p(A) + p(B)$

サイコロの例 (2)

- ❖ **試行：**サイコロを1回振る
- ❖ **性質1.** 標本空間が実現する確率は $p(\{1, 2, 3, 4, 5, 6\}) = 1$
- ❖ **性質2.**
 - ✓ 5の目が出る確率は $p(\{5\}) = 1/6$
 - ✓ 奇数の目が出る確率は $p(\{1, 3, 5\}) = 3/6 = 1/2$
- ❖ **性質3.** 4の目が出る事象と6の目が出る事象は互いに排反
 - ✓ 4もしくは6の目が出る確率は $p(\{4\} \cup \{6\}) = p(\{4, 6\}) = 2/6 = 1/3$
 - ✓ 4の目が出る確率と6の目が出る確率の和は $p(\{4\}) + p(\{6\}) = 1/6 + 1/6 = 1/3$

例題

例題：コインを2回投げて表と裏をみる試行をする。このとき以下の問いに答えよ。

- ① 1回目表、2回目裏となる確率は？
- ② 2回目に表が出る確率は？
- ③ 少なくとも1回は表が出る確率は？

経験的確率

【経験的確率とは？】

- ❖ 偏りのないコイン投げであれば、表の出る確率は1/2だと知っている
- ❖ コインに細工があるとき、表と裏の出る確率をどのように知るのか？
- ❖ 例えば、以下の手順で確率を推定する  **経験的確率** (empirical probability)
 - Step1.** コインを100回投げて、表と裏の出た回数を数え上げる
 - Step2.** 表の出る経験的確率は $p(\{\text{表}\}) = \text{表の出た回数} \div 100$
 - Step3.** 裏の出る経験的確率は $p(\{\text{裏}\}) = \text{裏の出た回数} \div 100$

アンケートの例 (1)

- ❖ **試行**：「デリバティブ入門」の授業アンケートをとる
- ❖ **標本空間**：{満足, 普通, 不満}
- ❖ **アンケート結果**：回答200人中、満足80人、普通100人、不満20人
- ❖ **経験的確率**：
 - ✓ 「満足」の経験的確率は $p(\{\text{満足}\}) = 80/200 = 2/5$
 - ✓ 「普通」の経験的確率は $p(\{\text{普通}\}) = 100/200 = 1/2$
 - ✓ 「普通」もしくは「不満」の経験的確率は $p(\{\text{普通}, \text{不満}\}) = 120/200 = 3/5$

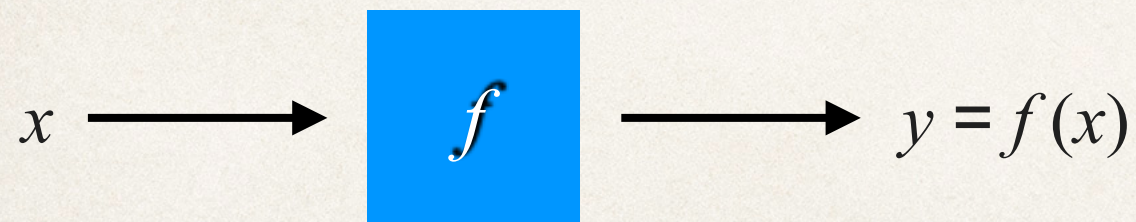
確率変数と期待値

確率変数

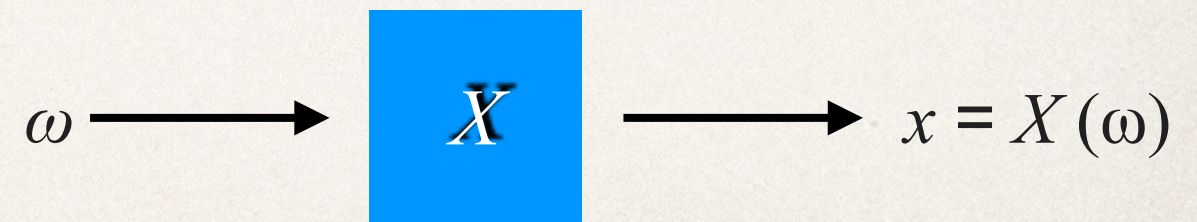
【確率変数とは？】

- ❖ 試行したとき、得られた結果に対して得点を付けることがある
- ❖ 根元事象に対して数値を対応させるルールが**確率変数** (random variable)
- ❖ このルール (確率変数) は**関数**である！

関数の概念



確率変数



サイコロの例 (3)

- ❖ **試行**：サイコロを1回振る
- ❖ **ルール**：サイコロの目が奇数なら1、偶数なら0
- ❖ **確率変数の対応**：
 - ✓ $X(1) = X(3) = X(5) = 1$ 、 $X(2) = X(4) = X(6) = 0$
 - ✓ もしくは $x_1 = x_3 = x_5 = 1$ 、 $x_2 = x_4 = x_6 = 0$

得点のルールである確率変数はいろいろと設定できる

アンケートの例 (2)

- ❖ **試行**：「ファイナンス入門」の授業アンケートをとる
- ❖ **標本空間**：{満足, 普通, 不満}
- ❖ **アンケート結果**：回答200人中、満足80人、普通100人、不満20人
- ❖ **確率変数の対応**：
 - ✓ $X(\text{満足}) = 2$ 、 $X(\text{普通}) = 1$ 、 $X(\text{不満}) = 0$

アンケートの集計結果を分析するには数値化する必要がある！

例題

例題：次の問いに答えよ。

- ① サイコロを1回振って出た目を2で割った数を確率変数 X のアウトプットとするとき、この確率変数 X の対応をすべて書き出さない
- ② コインを2回投げて表の出た回数を確率変数 Y のアウトプットとするとき、この確率変数 Y の対応をすべて書き出さない

期待値

【期待値とは？】

- ❖ 試行したときの平均的に期待される得点が**期待値** (expectation)
- ❖ 確率変数 X の期待される結果が期待値
- ❖ 期待値 $\mathbb{E}[X]$ の定義：

$$\mathbb{E}[X] := p_1x_1 + p_2x_2 + \cdots + p_nx_n$$

$x_1, x_2, \dots, x_n$ は確率変数 X の結果、 $p_1, p_2, \dots, p_n$ はそれぞれの結果が出る確率

n は根元事象の数であることに注意！

サイコロの例 (4)

❖ **試行とルール**：サイコロを1回振り、出た目が奇数なら1、偶数なら0

❖ **確率変数**：

✓ $x_1 = x_3 = x_5 = 1, \quad x_2 = x_4 = x_6 = 0$

❖ **確率**：

✓ $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$

❖ **期待値**：

$$\mathbb{E}[X] = \frac{1}{6} \times 1 + \frac{1}{6} \times 0 + \frac{1}{6} \times 1 + \frac{1}{6} \times 0 + \frac{1}{6} \times 1 + \frac{1}{6} \times 0 = \frac{1}{2}$$

例題

例題：次の問いに答えよ。

- ① サイコロを1回振って出た目を2で割った数を確率変数 X のアウトプットとするとき、この確率変数 X の期待値を計算せよ
- ② コインを2回投げて表の出た回数を確率変数 Y のアウトプットとするとき、この確率変数 Y の期待値を計算せよ

平均

【平均とは？】

- ❖ 経験的確率で計算される期待値を**平均** (average) という
- ❖ 平均の計算：

$$\mathbb{E}[X] = \frac{1}{N}x_1 + \frac{1}{N}x_2 + \cdots + \frac{1}{N}x_N = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

確率変数 X の N 個の観測値 $x_1, x_2, \dots, x_N$ がそれぞれ $1/N$ の確率で出たとする

N は観測値の数であることに注意！

サイコロの例 (5)

- ❖ サイコロに細工があるかもしれないので、100回投げて平均を計算
- ❖ **確率変数**：サイコロの出た目が奇数なら1、偶数なら0
- ❖ **平均**：
$$\mathbb{E}[X] = \frac{1}{100} \times 1 + \frac{1}{100} \times 0 + \dots + \frac{1}{100} \times 0 = \frac{1 + 0 + \dots + 0}{100}$$

根元事象の数 $n = 6$ 、観測値の数 $N = 100$ であることに注意！

	1回目	2回目	3回目	...	100回目
サイコロの目	3	6	3	...	4
確率	1/100	1/100	1/100	...	1/100
確率変数	1	0	1	...	0

アンケートの例 (3)

❖ 観測結果：

- ✓ 「満足」の回答者は80人なので、 $X(\text{満足}) = 2$ が80回観測された
- ✓ 「普通」の回答者は100人なので、 $X(\text{普通}) = 1$ が100回観測された
- ✓ 「不満」の回答者は20人なので、 $X(\text{不満}) = 0$ が20回観測された

❖ 平均：

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{200} \times 2 \times 80 + \frac{1}{200} \times 1 \times 100 + \frac{1}{200} \times 0 \times 20 \\ &= \frac{2}{5} \times 2 + \frac{1}{2} \times 1 + \frac{1}{10} \times 0 \\ &= \frac{2 \times 80 + 1 \times 100 + 0 \times 20}{200} = 1.3\end{aligned}$$

例題

例題： A株とB株の過去の株価および収益率が下表のとおりであった。
A株とB株の収益率の平均（**平均リターン**）を求めよ。

		3月末	4月末	5月末	6月末
A株	株価（円）	5,000	4,950	5,049	5,301
	収益率	—	-1.0%	2.0%	5.0%
B株	株価（円）	300	285	291	317
	収益率	—	-5.0%	2.0%	9.0%

分散と標準偏差

分散

【分散とは？】

- ❖ 試行したときの得点の散らばり度合いを表す値が**分散** (variance)
- ❖ 確率変数 X の結果の散らばり度合いが分散
- ❖ 分散 $\mathbf{Var}[X]$ の定義：

$$\mathbf{Var}[X] := \mathbb{E} [(X - \mu)^2] = p_1(x_1 - \mu)^2 + p_2(x_2 - \mu)^2 + \cdots + p_n(x_n - \mu)^2$$

$x_1, x_2, \dots, x_n$ は確率変数 X の結果、 $p_1, p_2, \dots, p_n$ はそれぞれの結果が出る確率、
 $\mu := \mathbb{E}[X]$ は確率変数 X の期待値

サイコロの例 (6)

❖ **試行とルール** : サイコロを1回振り、出た目が奇数なら1、偶数なら0

❖ **期待値** : $\mu = \mathbb{E}[X] = \frac{1}{2}$

❖ **分散** :
$$\begin{aligned}\mathbf{Var}[X] &= \frac{1}{6} \times \left(1 - \frac{1}{2}\right)^2 + \frac{1}{6} \times \left(0 - \frac{1}{2}\right)^2 + \frac{1}{6} \times \left(1 - \frac{1}{2}\right)^2 \\ &\quad + \frac{1}{6} \times \left(0 - \frac{1}{2}\right)^2 + \frac{1}{6} \times \left(1 - \frac{1}{2}\right)^2 + \frac{1}{6} \times \left(0 - \frac{1}{2}\right)^2 \\ &= \frac{1}{4}\end{aligned}$$

標準偏差

【標準偏差とは？】

- ❖ 試行したときの得点の標準化された散らばり度合いを表す値が標準偏差 (standard deviation)
- ❖ 確率変数 X の結果の標準化された散らばり度合いが標準偏差
- ❖ 標準偏差 σ_X の定義：

$$\sigma_X := \sqrt{\text{Var}[X]}$$

(標準化とは？) 分散を計算するときに値を2乗したので、標準偏差では分散の平方根をとることで単位を元に戻した

サイコロの例 (7)

❖ **試行とルール** : サイコロを1回振り、出た目が奇数なら1、偶数なら0

❖ **分散** : $\text{Var}[X] = \frac{1}{4}$

❖ **標準偏差** : $\sigma_X = \sqrt{\frac{1}{4}} = \frac{1}{2}$

❖ **期待値 $\pm$ 標準偏差** :

$$\mathbb{E}[X] \pm \sigma_X = \frac{1}{2} \pm \frac{1}{2} = \begin{cases} 1 \\ 0 \end{cases}$$

例題

例題：コインを1回投げたとき、次の確率変数に対する期待値、分散、標準偏差をそれぞれ求めよ。

- ① 表が出たら $X(H) = 1$ 、裏が出たら $X(T) = -1$ となる確率変数 X
- ② 表が出たら $Y(H) = 10$ 、裏が出たら $Y(T) = -10$ となる確率変数 Y

経験的な分散・標準偏差

【経験的な分散と標準偏差】

❖ 経験的な分散と標準偏差は...

$$\begin{aligned}\text{Var}[X] &= \frac{1}{N}(x_1 - \mu)^2 + \frac{1}{N}(x_2 - \mu)^2 + \cdots + \frac{1}{N}(x_N - \mu)^2 \\ &= \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2}{N}\end{aligned}$$

$$\sigma_X = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2}{N}}$$

ただし、この統計値は適当とはいえない？ ➡ **標本分散、標本標準偏差**

例題

例題： A株とB株の過去の株価および収益率が下表のとおりであった。
A株とB株の収益率の分散と標準偏差（**ボラティリティ**）を求めよ。

		3月末	4月末	5月末	6月末
A株	株価（円）	5,000	4,950	5,049	5,301
	収益率	—	-1.0%	2.0%	5.0%
B株	株価（円）	300	285	291	317
	収益率	—	-5.0%	2.0%	9.0%

期待値、分散、標準偏差の性質

X を確率変数、 a と b を定数とすると、次の性質が成り立つ

- ❖ **性質1 :** $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- ❖ **性質2 :** $\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- ❖ **性質3 :** $\mathbf{Var}[aX + b] = a^2\mathbf{Var}[X]$
- ❖ **性質4 :** $\sigma_X^2 = \mathbf{Var}[X]$
- ❖ **性質5 :** $\sigma_{aX+b} = |a| \sigma_X$

例題

例題： X を確率変数とするとき、次の問いに答えよ。

- ① $\mathbb{E}[X] = 3$ のとき、 $\mathbb{E}[-2X + 3]$ を求めよ
- ② $\mathbb{E}[X] = 5$ かつ $\mathbf{Var}[X] = 3$ のとき、 $\mathbb{E}[X^2]$ を求めよ
- ③ $\mathbf{Var}[X] = 4$ のとき、 $\mathbf{Var}[-3X + 1]$ を求めよ
- ④ $\sigma_X = 2$ のとき、 σ_{-3X+1} を求めよ

共分散と相関係数

2つのコインを投げたときの確率

表1：独立

		コインA		
		H	T	計
コインB	H	1/4	1/4	1/2
	T	1/4	1/4	1/2
	計	1/2	1/2	

表2：正の相関

		コインA		
		H	T	計
コインB	H	3/8	1/8	1/2
	T	1/8	3/8	1/2
	計	1/2	1/2	

表3：負の相関

		コインA		
		H	T	計
コインB	H	1/8	3/8	1/2
	T	3/8	1/8	1/2
	計	1/2	1/2	

共分散

【共分散とは？】

- ❖ 2つの確率変数 X と Y の連動性が**共分散** (covariance)
- ❖ 共分散 $\mathbf{Cov}[X, Y]$ の定義（もしくは σ_{XY} ）：

$$\mathbf{Cov}[X, Y] := \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

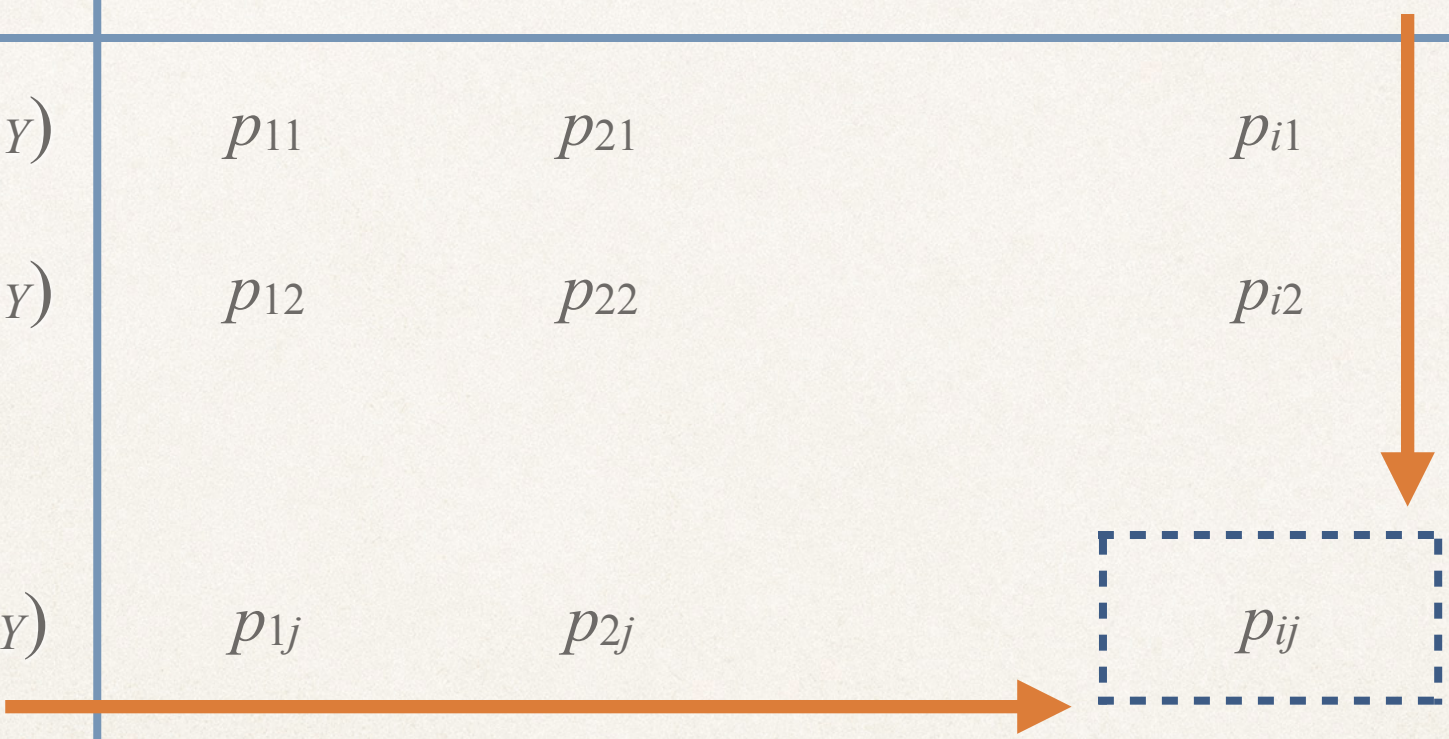
$$\begin{aligned} &= p_{11}(x_1 - \mu_X)(y_1 - \mu_Y) + p_{21}(x_2 - \mu_X)(y_1 - \mu_Y) + \cdots + p_{m1}(x_m - \mu_X)(y_1 - \mu_Y) \\ &+ p_{12}(x_1 - \mu_X)(y_2 - \mu_Y) + p_{22}(x_2 - \mu_X)(y_2 - \mu_Y) + \cdots + p_{m2}(x_m - \mu_X)(y_2 - \mu_Y) \\ &+ \cdots \\ &+ p_{1n}(x_1 - \mu_X)(y_n - \mu_Y) + p_{2n}(x_2 - \mu_X)(y_n - \mu_Y) + \cdots + p_{mn}(x_m - \mu_X)(y_n - \mu_Y) \end{aligned}$$

p_{ij} は $X = x_i$ かつ $Y = y_j$ となる確率、 m 、 n は根元事象の数であることに注意！

$\mu_X := \mathbb{E}[X]$ は確率変数 X の期待値、 $\mu_Y := \mathbb{E}[Y]$ は確率変数 Y の期待値

共分散の計算イメージ

		確率変数 X					
		$(x_1 - \mu_X)$	$(x_2 - \mu_X)$	...	$(x_i - \mu_X)$	...	$(x_m - \mu_X)$
確率変数 Y	$(y_1 - \mu_Y)$	p_{11}	p_{21}		p_{i1}		p_{m1}
	$(y_2 - \mu_Y)$	p_{12}	p_{22}		p_{i2}		p_{m2}
	$\vdots$						
	$(y_j - \mu_Y)$	p_{1j}	p_{2j}		p_{ij}		p_{mj}
	$\vdots$						
	$(y_n - \mu_Y)$	p_{1n}	p_{2n}		p_{in}		p_{mn}



相関係数

【相関係数とは？】

- ❖ 2つの確率変数 X と Y の 標準化された連動性が **相関係数** (correlation)
- ❖ 相関係数 ρ_{XY} の定義：

$$\rho_{XY} := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

(相関係数の性質)

- 相関係数は必ず -1 から 1 の間の値をとる
- 相関係数が正のとき **正相関**、負のとき **負相関**という
- 相関係数が 0 のとき **無相関**といい、独立ならば無相関だが、逆は成り立たない

例題

例題：コインAとコインBを同時に投げた。コインAの確率変数を $X(H) = 2$ 、 $X(T) = -2$ 、コインBの確率変数を $Y(H) = 2$ 、 $Y(T) = -2$ とするとき、次の問いに答えよ。

- ① 2つのコインの表裏確率が表1のとき、共分散と相関係数を求めよ
- ② 2つのコインの表裏確率が表2のとき、共分散と相関係数を求めよ
- ③ 2つのコインの表裏確率が表3のとき、共分散と相関係数を求めよ

経験的な共分散・相関係数

【経験的な共分散と相関係数】

- ❖ 観測値のペア $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ で計算される共分散は...

$$\begin{aligned}\text{Cov}[X, Y] &= \frac{1}{N}(x_1 - \mu_X)(y_1 - \mu_Y) + \dots + \frac{1}{N}(x_N - \mu_X)(y_N - \mu_Y) \\ &= \frac{(x_1 - \mu_X)(y_1 - \mu_Y) + \dots + (x_N - \mu_X)(y_N - \mu_Y)}{N}\end{aligned}$$

- ❖ 経験的な相関係数は経験的な共分散と経験的な標準偏差で計算される

経験的な共分散の計算イメージ

コインの例：コインAとコインBの表裏確率が事前にはわからないので、2つのコインを同時に20回投げて経験的な共分散を求める

		1回目	2回目	...	20回目
コインA	表裏	T	T	...	H
	確率変数	-2	-2	...	2
コインB	表裏	H	T	...	H
	確率変数	2	-2	...	2

例題

例題： A株とB株の過去の株価および収益率が下表のとおりであった。
A株とB株の収益率の共分散と相関係数を求めよ。

		3月末	4月末	5月末	6月末
A株	株価（円）	1,000	1,020	1,051	1,093
	収益率	—	2.0%	3.0%	4.0%
B株	株価（円）	500	495	520	530
	収益率	—	-1.0%	5.0%	2.0%

期待値、分散、共分散の性質

X と Y を確率変数、 a と b を定数とすると、次の性質が成り立つ

❖ **性質6 :** $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$

❖ **性質7 :** $\mathbf{Var}[aX + bY] = a^2\mathbf{Var}[X] + b^2\mathbf{Var}[Y] + 2ab\mathbf{Cov}[X, Y]$

❖ **性質8 :** $\mathbf{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

確率変数 X と Y が**独立**であるとき、次の性質が成り立つ

❖ **性質9 :** $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

❖ **性質10 :** $\mathbf{Cov}[X, Y] = 0, \quad \rho_{XY} = 0$

例題

例題： X と Y を確率変数とするとき、次の問いに答えよ。

- ① $\mathbb{E}[X] = 3$ 、 $\mathbb{E}[Y] = 4$ のとき、 $\mathbb{E}[-2X + Y + 3]$ を求めよ
- ② $\mathbf{Var}[X] = 3$ 、 $\mathbf{Var}[Y] = 4$ 、 $\mathbf{Cov}[X, Y] = -2$ のとき、 $\mathbf{Var}[-2X + Y]$ を求めよ
- ③ X と Y が独立であり、 $\mathbf{Var}[X] = 3$ 、 $\mathbf{Var}[Y] = 4$ のとき、 $\mathbf{Var}[-2X + Y]$ を求めよ

回歸分析

回帰分析

【回帰分析とは？】

- ❖ 2つの確率変数 X と Y の観測値のペアの系列

$$(X, Y) = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_N, y_N)\}$$

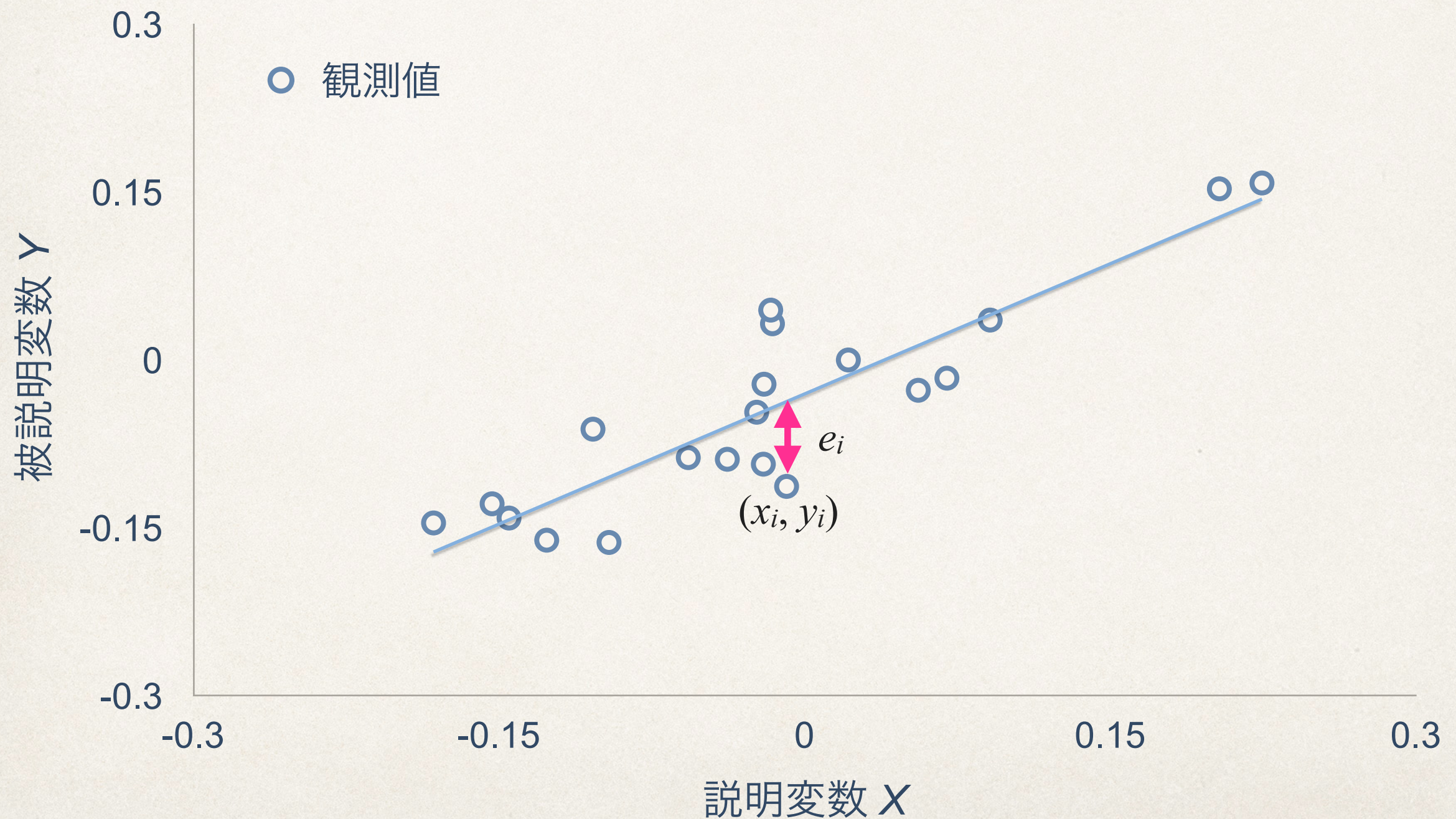
を x - y 平面上にプロットしたグラフを**散布図**という

- ❖ **回帰分析**とは、散布図に当てはまりのよい直線（**回帰直線**）を引く手法

$$Y = \beta X + \alpha$$

X を**説明変数**、 Y を**被説明変数**という

散布図と回帰直線



回帰係数と誤差

【回帰係数と誤差】

- ❖ 回帰直線の傾き β :

$$\beta = \frac{\text{Cov}[X, Y]}{\text{Var}[X]}$$

- ❖ 回帰直線の切片 α :

$$\alpha = \mathbb{E}[Y] - \beta \mathbb{E}[X]$$

- ❖ 誤差 e :

$$y_i = \beta x_i + \alpha + e_i$$

誤差 e は他の確率変数とは独立な確率変数で、平均 $\mathbb{E}[e] = 0$ 、標準偏差 $\sigma_e > 0$

例題

例題： A株とB株の過去の株価および収益率が下表のとおりであった。
A株の収益率を説明変数、B株の収益率を被説明変数とする回帰直線の方程式を求めよ。

		3月末	4月末	5月末	6月末
A株	株価（円）	1,000	1,020	1,051	1,093
	収益率	—	2.0%	3.0%	4.0%
B株	株価（円）	500	495	520	530
	収益率	—	-1.0%	5.0%	2.0%