確率・統計の基礎 攻略マニュアル

計量経済学 基礎からの解説シリーズ

目次

1　確率変数と確率分布　3

1.1　確率変数とは 3

1.2　離散型と連続型 4

1.3　分位点 7

2　期待値　8

2.1　期待値の定義 8

2.2　期待値の性質（最重要） 9

3　分散と標準偏差　11

3.1　分散と標準偏差の定義 11

3.2　分散の性質 12

3.3　標準化 12

4　2つの確率変数：同時分布・周辺分布・独立　14

4.1　同時分布と周辺分布 14

4.2　独立 15

5　共分散と相関　17

5.1　共分散と相関係数 17

5.2　共分散の計算ルール 18

6 条件付き分布と条件付き期待値 21

6.1 条件付き分布 21

6.2 条件付き期待値 22

6.3 条件付き分散 23

6.4 条件付き期待値の性質 24

6.5 誤差項の分解 24

7 繰り返し期待値の法則（LIE） 26

7.1 繰り返し期待値の法則 26

7.2 LIE の最重要の応用 27

7.3 全分散の公式 28

8 独立・平均独立・無相関：強さの違い 29

8.1 3つの条件 29

8.2 反例1：平均独立だが独立ではない 30

8.3 反例2：無相関だが平均独立ではない 30

9 重要な分布：正規分布とカイ二乗分布 33

9.1 正規分布 33

9.2 カイ二乗分布 34

10 母集団・標本・推定量 36

10.1 用語の整理 36

10.2 標本平均の性質 37

11 記号の読み方と練習問題 39

11.1 記号の読み方まとめ 39

11.2 練習問題 40

11.3 解答 40

第1章

確率変数と確率分布

ここまでの流れ　計量経済学では、データを「母集団から無作為に選ばれた人の記録」とみなし、そこから母集団の性質を推測します。その土台になる言葉が、確率・統計です。
この章の問い　「実際に観測するまで値が決まらない量」を、数学ではどう扱えばよいのでしょうか。
この章で分かること　確率変数、離散型と連続型、確率関数・密度関数・累積分布関数（CDF）、そして分位点です。

1.1　確率変数とは

要点：確率変数

確率変数とは、実際に観測するまで値が決まっておらず、「どの値をどのくらいの確率でとるか」が決まっている変数のことです。

身近な例を挙げます。

- サイコロの出目 X：$1, 2, \ldots, 6$ をそれぞれ確率 $1/6$ でとります。
- ランダムに選んだ労働者の賃金 Y：誰が選ばれるかで値が変わります。
- ランダムに選んだ労働者が大卒かどうか X（大卒なら 1、そうでなければ 0）：このように 0 か 1 をとる変数を**ダミー変数**と呼びます。

確率変数と実現値の区別　サイコロを振る前の「出目」は確率変数 X です。振った後に出た具体的な数（例えば 4）は X の**実現値**といいます。確率変数は「まだ決まっていない量」、実現値は「決まった後のただの数」です。期待値や分散は確率変数について定義されるもので、実現値（ただの数）には期待値もばらつきもありません。

補足：記号の使い分け

多くの統計学の教科書は、確率変数を大文字（X）、実現値を小文字（x）で書き分けます。一方、計量経済学では y_i, x_i のように小文字で確率変数を表すことが多く、同じ記号が文脈によって確率変数にも実現値にもなります。迷ったら「期待値をとっている（E の中にある）なら確率変数」と読んでください。この教材では、確率・統計の例では大文字を、回帰の話に近い部分では小文字を使います。

計量経済学では、データ $\{(y_i, x_i)\}_{i=1}^{N}$ の1人1人（$i = 1, \ldots, N$）を「母集団から無作為に選ばれた人の確率変数（の実現値）」と考えます。添字 i を外した (y, x) は「母集団から選ばれる典型的な1人」を表し、その分布が母集団の分布です。

発展：確率変数の正確な定義

起こりうる結果全体の集合を Ω（標本空間）、その部分集合を事象とし、各事象 A に確率 $\Pr(A)$ を割り当てます。確率は次の公理を満たします。

(i) $\Pr(A) \geq 0$

(ii) $\Pr(\Omega) = 1$

(iii) 互いに重ならない事象 $A_1, A_2, \ldots$ について、確率は足し合わせられる：

$$\Pr(A_1 \cup A_2 \cup \cdots) = \Pr(A_1) + \Pr(A_2) + \cdots$$

確率変数とは、各結果 $\omega \in \Omega$ に実数 $X(\omega)$ を対応させる**関数**です。例えばサイコロでは $\Omega = \{1\text{の目}, \ldots, 6\text{の目}\}$ で、$X(k\text{ の目}) = k$ です。「変数」という名前ですが、数学的には関数である点に注意してください。

1.2 離散型と連続型

要点：確率関数・密度関数・累積分布関数

- **離散型**（とびとびの値をとる）：各値の確率 $p(x) = \Pr(X = x)$ を**確率関数**といいます。$p(x) \geq 0$、$\sum_x p(x) = 1$。
- **連続型**（連続的な値をとる）：**密度関数** $f_X(x) \geq 0$ を使い、

$$\Pr(a \leq X \leq b) = \int_a^b f_X(x)\, dx, \qquad \int_{-\infty}^{\infty} f_X(x)\, dx = 1.$$

- **累積分布関数（CDF）**：$F_X(x) = \Pr(X \leq x)$。離散型・連続型のどちらでも定義できます。

数式の読み方

- 確率関数は「各値の確率そのもの」、密度関数は「確率の濃さ」で、確率は密度関数の**面積**として得られます。
- CDF は「x 以下になる確率」を x の関数として並べたもので、左から順に確率を積み上げていくイメージです。

連続型では、「ちょうど1つの値をとる確率」は 0 です（$\Pr(X = a) = \int_a^a f_X(x)\,dx = 0$）。確率になるのは密度関数の面積であって、密度の値そのものではありません。

補足：つまずきポイント

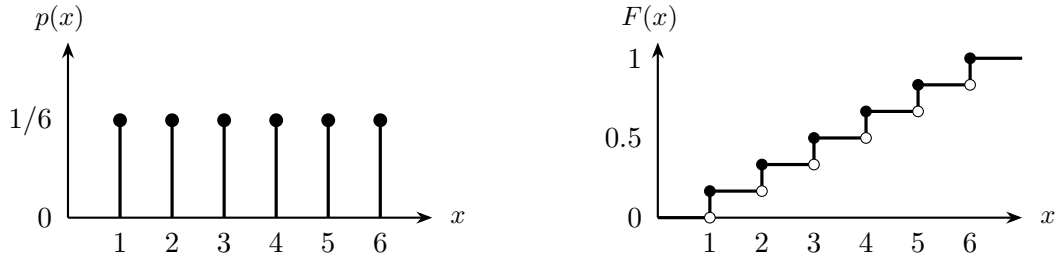
密度関数の値 $f_X(x)$ は確率ではないので、1 を超えることもあります。例えば区間 $[0, 0.5]$ 上の一様分布の密度は、その区間で $f_X(x) = 2$ です（面積は $2 \times 0.5 = 1$）。$f_X(x)\,dx$（密度 × 微小な幅）が「x の近くの値をとる確率」にあたると考えてください。

要点：CDF の性質

1. 単調非減少（x が増えると $F_X(x)$ は減らない）。
2. $\lim_{x \to -\infty} F_X(x) = 0$、$\lim_{x \to \infty} F_X(x) = 1$。
3. 右連続。

また $\Pr(a < X \leq b) = F_X(b) - F_X(a)$ で、連続型なら $F_X'(x) = f_X(x)$（CDF を微分すると密度）です。

図 1.1　サイコロ（離散型）の確率関数と CDF

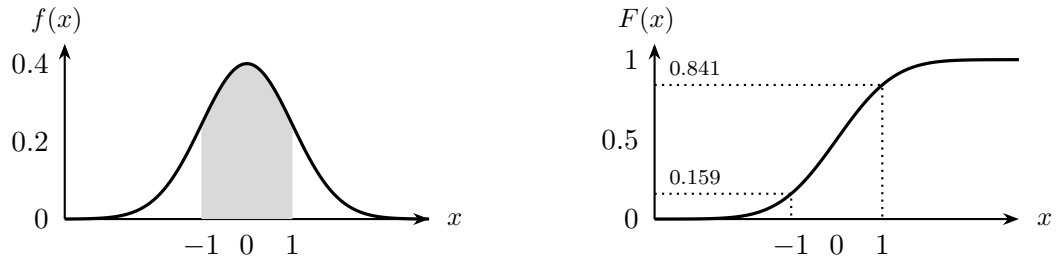


確率関数 $p(x) = 1/6$（$x = 1, \ldots, 6$）と、その CDF $F(x) = \lfloor x \rfloor / 6$（$1 \le x < 6$ の範囲、$x \ge 6$ で 1）を描いた。

図の読み方

1. 左の図は確率関数で、6つの値それぞれに 1/6 の確率があります。
2. 右の図は CDF で、値をとる点（1〜6）で 1/6 ずつ跳ね上がる階段状になります。
3. 黒丸は「跳んだ後の値を含む」こと（右連続性）、白丸は「跳ぶ前の値」を表します。例えば $F(3) = 3/6 = 0.5$ です。

図 1.2　標準正規分布（連続型）の密度と CDF



標準正規分布の密度 $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ と、その CDF $F(x) = \int_{-\infty}^{x} f(t)\,dt$（数値計算した値）を描いた。

図の読み方

1. 左の図の灰色部分は $-1 \le x \le 1$ の範囲の面積で、$\Pr(-1 \le Z \le 1) \approx 0.683$ です。
2. 右の図の CDF はなめらかに 0 から 1 へ増加します。
3. 灰色の面積は、CDF の値の差 $F(1) - F(-1) \approx 0.841 - 0.159 = 0.683$ に等しくなります。確率を面積で測ることと、CDF の差で測ることは同じことです。

1.3 分位点

要点：分位点

α を 0 と 1 の間の数とします。$F(q_\alpha) = \alpha$ を満たす値 q_α を、F のα **分位点**といいます。例えば中央値は $q_{0.5}$ です。

F が狭義単調増加（どこでも strictly に増える）なら、この式を満たす q_α はちょうど 1 つに決まります。「分布の左から α の割合のところにある点」と読めます。

発展：離散型の分位点

サイコロのように CDF が階段状だと、$F(q) = \alpha$ を満たす q が存在しないことも、1 つに決まらないこともあります。例えば $\alpha = 0.4$ では、F は $2/6 \approx 0.33$ から $3/6 = 0.5$ に跳ぶので、解はありません。$\alpha = 0.5$ では、区間 $[3, 4)$ のすべての点で $F = 0.5$ です。そこで一般には

$$q_\alpha = \inf\{x : F(x) \geq \alpha\}$$

（$F(x) \geq \alpha$ となる最小の x）と定義します。

要点

- 確率変数は「観測するまで値が決まらない量」で、観測後の値（実現値）と区別します。
- 離散型は確率関数、連続型は密度関数で分布を表し、連続型の確率は密度の面積です。密度の値は 1 を超えることもあります。
- CDF $F(x) = \Pr(X \leq x)$ は離散型・連続型の両方で使え、$\Pr(a < X \leq b) = F(b) - F(a)$ です。
- α 分位点は $F(q_\alpha) = \alpha$ を満たす値で、検定の臨界値などに使います。

次の章へ　分布が分かれば、「平均的にどれくらいの値をとるか」を数値にできます。次の章で、分布の中心を表す期待値を学びます。

大学院とのつながり　CDF や分位点は、母集団の分布の特徴（中央値など）を推定する話題で使います。

第2章

期待値

ここまでの流れ　確率変数の分布は、確率関数・密度関数・CDF で表せました。
この章の問い　分布の「中心」を1つの数で表すにはどうすればよいのでしょうか。また、その数はどんな計算ルールに従うのでしょうか。
この章で分かること　期待値の定義と2通りの解釈、そして期待値の最重要の性質である線形性（計量経済学の式変形の大半を支える性質）です。

2.1　期待値の定義

要点：期待値（平均）

$$\mathrm{E}(X) = \sum_{x} x\, p(x) \ \text{（離散型）}, \qquad \mathrm{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \ \text{（連続型）}. \tag{2.1}$$

「とりうる値 $\times$ その確率（密度）」をすべて足し合わせたものです。

数式の読み方

- $p(x)$（または $f_X(x)\,dx$）は、値 x が出る確率の重みです。
- 期待値は、各値をその起こりやすさで重みづけして平均したものなので、「分布の重心」にあたります。

期待値には2つの解釈があります。

1. **重み付き平均**：各値を、その起こりやすさで重みづけして平均したもの。母集団全体の平均と考えてよい。
2. **長い目で見た平均**：同じ実験を何度もくり返したときの、実現値の平均が近づく値。これを正確にしたのが、漸近理論で学ぶ大数の法則です。

期待値は「1回の実験で期待される値」という意味ではありません。例えば、サイコロの期待値 3.5 は、1回振って出ることはありません。

計算例：サイコロの期待値

$$\mathrm{E}(X) = 1\cdot\tfrac{1}{6} + 2\cdot\tfrac{1}{6} + \cdots + 6\cdot\tfrac{1}{6} = \tfrac{21}{6} = 3.5.$$

2.2 期待値の性質（最重要）

要点：期待値の性質

a, b を定数、X, Y を確率変数とすると、次が成り立ちます。

1. $\mathrm{E}(a) = a$（定数の期待値はその定数）。
2. $\mathrm{E}(a + bX) = a + b\,\mathrm{E}(X)$。
3. $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$（独立でなくても常に成り立つ）。
4. 関数 g について $\mathrm{E}[g(X)] = \sum_x g(x)p(x)$、または $\int g(x) f_X(x)\,dx$。
5. 一般に $\mathrm{E}(XY) \neq \mathrm{E}(X)\mathrm{E}(Y)$、$\mathrm{E}[g(X)] \neq g(\mathrm{E}X)$（例：$\mathrm{E}(X^2) \neq (\mathrm{E}X)^2$）。

性質2の証明（離散型）　性質4で $g(x) = a + bx$ とおきます。

$$
\begin{aligned}
\mathrm{E}(a + bX) &= \sum_x (a + bx)\,p(x) && \text{（性質4で } g(x) = a + bx\text{）} \\
&= a \underbrace{\sum_x p(x)}_{=1} + b \underbrace{\sum_x x\,p(x)}_{=\mathrm{E}(X)} && \text{（和を2つに分ける）} \\
&= a + b\,\mathrm{E}(X)
\end{aligned}
$$

性質3の証明（離散型）　同時確率 $p(x, y) = \Pr(X = x, Y = y)$ を使います。

$$
\begin{aligned}
\mathrm{E}(X + Y) &= \sum_x \sum_y (x + y)\,p(x, y) && \text{（性質4（2変数版））} \\
&= \sum_x x \underbrace{\sum_y p(x, y)}_{=\Pr(X=x)} + \sum_y y \underbrace{\sum_x p(x, y)}_{=\Pr(Y=y)} && \text{（和を2つに分け、和の順序を入れ替える）} \\
&= \mathrm{E}(X) + \mathrm{E}(Y)
\end{aligned}
$$

独立性は一度も使っていないので、X と Y がどんな関係にあっても成り立ちます。

計算例：性質5の確認（サイコロ）

$\mathrm{E}(X^2) = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6} \approx 15.17$ ですが、$(\mathrm{E}X)^2 = 3.5^2 = 12.25$ で、等しくありません。この差がちょうど次の章の分散です。

性質2と3をまとめて、期待値の**線形性**といいます。計量経済学の式変形の大半は、この線形性を使っています。ベクトルや行列の期待値は「各要素の期待値を並べたもの」と定

義されるので、定数行列 A について $\mathrm{E}(AX) = A\,\mathrm{E}(X)$ も成り立ちます（各要素が「定数 $\times$ 確率変数」の和なので、線形性をくり返し使えばよい）。

補足：つまずきポイント

$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ が成り立つのは、X と Y が独立な場合など（より正確には $\mathrm{Cov}(X,Y) = 0$ の場合）に限られます。「期待値は何でも中に入れたり外に出したりできる」わけではなく、自由に使えるのは定数倍と和（線形性）だけです。

発展：期待値が存在しない場合

期待値は、和や積分が有限の値に定まるとき（$\mathrm{E}|X| < \infty$ のとき）にだけ定義されます。例えば、密度 $f(x) = \dfrac{1}{\pi(1+x^2)}$ のコーシー分布では $\int |x| f(x)\,dx = \infty$ となり、期待値が存在しません。大学院の教科書で「$\mathrm{E}\|x\| < \infty$ を仮定する」と書かれているのは、このような事態を排除して、期待値が定まることを保証するためです。

要点

- 期待値は「とりうる値 $\times$ 確率」の総和（積分）で、分布の重心を表します。1回の実験で出る値という意味ではありません。
- 線形性：$\mathrm{E}(a+bX) = a + b\mathrm{E}(X)$、$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$。後者は独立でなくても成り立ちます。
- $\mathrm{E}(XY) \neq \mathrm{E}(X)\mathrm{E}(Y)$、$\mathrm{E}[g(X)] \neq g(\mathrm{E}X)$ が一般的で、自由に外に出せるのは定数と和だけです。
- 期待値は存在しないこともあり、そのため $\mathrm{E}|X| < \infty$ のような条件が必要になります。

次の章へ　期待値は分布の「中心」しか教えてくれません。次の章で、分布の「ばらつき」を表す分散を学びます。

大学院とのつながり　期待値の線形性は、推定量の不偏性の証明をはじめ、あらゆる式変形の出発点になります。

第3章

分散と標準偏差

ここまでの流れ　期待値は分布の中心（重心）を表す数でした。
この章の問い　分布の「ばらつき」はどう測ればよいのでしょうか。
この章で分かること　分散と標準偏差の定義、計算公式、そして「平均を引いて標準偏差で割る」標準化です。

3.1　分散と標準偏差の定義

要点：分散・標準偏差

$\mu = \mathrm{E}(X)$ とするとき

$$\mathrm{Var}(X) = \mathrm{E}\big[(X-\mu)^2\big], \qquad \text{標準偏差} = \sqrt{\mathrm{Var}(X)}. \tag{3.1}$$

分散は「平均からのずれの2乗の平均」で、ばらつき（散らばり）の大きさを表します。

数式の読み方

- $X-\mu$ は、確率変数が平均からどれだけずれているかを表します。
- それを2乗して期待値をとるので、大きなずれが出やすい分布ほど分散が大きくなります。

なぜ2乗するのか　ずれ $X-\mu$ そのものの平均は、線形性より $\mathrm{E}(X)-\mu=0$ となり、常に0になってしまいます（プラスのずれとマイナスのずれが打ち消し合う）。2乗すれば符号が消え、大きなずれほど大きく数えられます。

なぜ標準偏差を使うのか　分散は2乗しているので、単位も2乗になります（月給が「万円」なら分散は「万円^2」）。平方根をとった標準偏差は元の単位に戻り、「平均からおよそどれくらい離れるか」の目安になります。

3.2 分散の性質

要点：分散の性質

1. 計算公式：$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$。
2. $\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$（定数を足してもばらつきは変わらない。$b$ 倍すると分散は b^2 倍）。
3. $\mathrm{Var}(X) \geq 0$。$\mathrm{Var}(X) = 0$ となるのは、X が（確率 1 で）定数のときだけ。

計算公式の導出　$(X-\mu)^2 = X^2 - 2\mu X + \mu^2$ の期待値をとります。μ は定数なので外に出せます。

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(X^2) - 2\mu\,\mathrm{E}(X) + \mu^2 && \text{（線形性。}\mu\text{ は定数）}\\
&= \mathrm{E}(X^2) - 2\mu^2 + \mu^2 && (\mathrm{E}(X) = \mu)\\
&= \mathrm{E}(X^2) - \mu^2
\end{aligned}$$

計算例：サイコロの分散

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2 = \tfrac{91}{6} - \tfrac{49}{4} = \tfrac{182-147}{12} = \tfrac{35}{12} \approx 2.92, \qquad \text{標準偏差} \approx 1.71.$$

3.3 標準化

要点：標準化

$\mu = \mathrm{E}(X)$、$\sigma = \sqrt{\mathrm{Var}(X)} > 0$ として、$Z = (X-\mu)/\sigma$ とおくと

$$\mathrm{E}(Z) = 0, \qquad \mathrm{Var}(Z) = 1.$$

「平均を引いて標準偏差で割る」と、平均 0・分散 1 になります。

導出は、期待値の線形性と分散の性質 2 を使います。

$$\begin{aligned}
\mathrm{E}(Z) &= \frac{\mathrm{E}(X) - \mu}{\sigma} = 0 && \text{（線形性）}\\
\mathrm{Var}(Z) &= \frac{1}{\sigma^2}\mathrm{Var}(X) = 1 && \text{（性質2で }b = 1/\sigma\text{、}\mathrm{Var}(X) = \sigma^2\text{）}
\end{aligned}$$

数式の読み方

標準化した値は「平均から標準偏差の何個分離れているか」を表します。単位に依存しないので、異なる単位の量を比べられます。t 統計量は、この考え方でできています。

要点

- 分散は「平均からのずれの2乗の平均」、標準偏差はその平方根で、元の単位に戻ります。
- 計算公式 $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$ が最もよく使われます。
- $\mathrm{Var}(a + bX) = b^2\mathrm{Var}(X)$：定数を足してもばらつきは変わらず、$b$ 倍すると分散は b^2 倍になります。
- 標準化 $(X - \mu)/\sigma$ で、平均 0・分散 1 に直せます。

次の章へ　ここまでは1つの確率変数を扱いました。回帰分析は「x と y の関係」を調べるので、次の章から2つ以上の確率変数を同時に扱います。

大学院とのつながり　標本平均の分散が σ^2/N になること（第10章）は、漸近理論の出発点です。

第4章

2つの確率変数：同時分布・周辺分布・独立

ここまでの流れ　1つの確率変数については、期待値（中心）と分散（ばらつき）が分かりました。
この章の問い　2つの確率変数を同時に扱うとき、その「関係」を含む情報はどう表せばよいのでしょうか。「関係がない」とはどういう意味でしょうか。
この章で分かること　同時分布と周辺分布、独立の定義、独立なら期待値が分解できること、そして IID の意味です。この章から、教材を通して使う数値例を導入します。

4.1　同時分布と周辺分布

要点：同時分布・周辺分布

- **同時分布**：(X, Y) の組み合わせごとの確率 $\Pr(X = x, Y = y)$。2つの変数の関係を含むすべての情報を持ちます。
- **周辺分布**：片方だけの分布。もう一方について足し合わせて得られます。

$$\Pr(X = x) = \sum_{y} \Pr(X = x, Y = y), \qquad \Pr(Y = y) = \sum_{x} \Pr(X = x, Y = y).$$

連続型なら $f_X(x) = \int f_{XY}(x, y)\, dy$ です。

数式の読み方

「周辺」という名前は、表にしたとき、合計が表の周辺（端）に書かれることに由来します。**周辺分布だけからは同時分布は復元できません**。同じ周辺分布を持つ、異なる関係の同時分布がいくつもあるからです。

計算例：数値例（この教材でずっと使う仮想の母集団）

X を大卒ダミー（1：大卒、0：非大卒）、Y を月給（万円）とします。母集団での同時確率 $\Pr(X=x, Y=y)$ が次の表で与えられているとします。

	$Y=20$	$Y=30$	$Y=40$	合計 $\Pr(X=x)$
$X=0$	0.20	0.20	0.10	0.50
$X=1$	0.05	0.20	0.25	0.50
合計 $\Pr(Y=y)$	0.25	0.40	0.35	1.00

例えば左上の 0.20 は、「ランダムに1人選んだとき、非大卒かつ月給 20 万円である確率が 20%」という意味です。周辺分布から

$$\mathrm{E}(Y) = 20(0.25) + 30(0.40) + 40(0.35) = 31, \qquad \mathrm{E}(X) = 0 \times 0.5 + 1 \times 0.5 = 0.5$$

です。ダミー変数の期待値は「1 をとる確率」に等しくなります（$\mathrm{E}(X) = \Pr(X=1)$）。

4.2 独立

要点：独立

すべての x, y で

$$\Pr(X=x, Y=y) = \Pr(X=x)\Pr(Y=y) \tag{4.1}$$

（連続型なら $f_{XY}(x,y) = f_X(x) f_Y(y)$）が成り立つとき、$X$ と Y は**独立**であるといいます。

数式の読み方

独立なら、次章の条件付き確率の定義より

$$\Pr(Y=y \mid X=x) = \frac{\Pr(X=x, Y=y)}{\Pr(X=x)} = \Pr(Y=y)$$

となります。「X の値を知っても、Y の分布について何の情報も得られない」ということです。

計算例：数値例は独立か

$\Pr(X=0, Y=20) = 0.20$ ですが、$\Pr(X=0)\Pr(Y=20) = 0.5 \times 0.25 = 0.125$ なので、独立ではありません。独立であるにはすべての組み合わせで等式が成り立つ必要があるので、1か所でも成り立たなければ独立ではありません。

要点：独立なら期待値が分解できる

X と Y が独立なら、任意の関数 g, h について

$$\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\,\mathrm{E}[h(Y)]. \tag{4.2}$$

特に $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ です。

離散型での証明は次のとおりです。

$$\begin{aligned}
\mathrm{E}[g(X)h(Y)] &= \sum_x \sum_y g(x)h(y)\,p_X(x)p_Y(y) && \text{（独立なので同時確率が積になる）} \\
&= \sum_x g(x)p_X(x) \cdot \sum_y h(y)p_Y(y) && \text{（x の和と y の和に分けられる）} \\
&= \mathrm{E}[g(X)]\,\mathrm{E}[h(Y)]
\end{aligned}$$

補足：IID の意味

データ $\{(y_i, x_i)\}_{i=1}^N$ が **IID**（independent and identically distributed、独立同一分布）であるとは、「異なる個人 i と j のデータ (y_i, x_i) と (y_j, x_j) は互いに独立で、全員が同じ母集団分布から抽出されている」ことです。これは無作為抽出のことです。同じ個人の中の y_i と x_i が独立という意味ではありません（もし独立なら、x は y と無関係で、回帰分析をする意味がありません）。

要点

- 同時分布は2変数の関係を含むすべての情報を持ち、周辺分布は一方だけの分布です。周辺分布から同時分布は復元できません。
- 独立とは、すべての (x, y) で同時確率が周辺確率の積になることです。1か所でも崩れれば独立ではありません。
- 独立なら $\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$。特に $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$。
- IID は「異なる個人が互いに独立で、同じ母集団分布に従う」こと、つまり無作為抽出です。

次の章へ　独立でない2変数の「関係の強さ」を、1つの数で表したくなります。次の章で、共分散と相関係数を学びます。

大学院とのつながり　IID の仮定は、大数の法則・中心極限定理を使うときの標準的な前提です。

第5章

共分散と相関

ここまでの流れ　2つの確率変数は、同時分布で表せることと、独立なら「関係がない」ことが分かりました。
この章の問い　独立ではない2変数について、「一緒に動く傾向」の強さを1つの数で表すにはどうすればよいのでしょうか。
この章で分かること　共分散と相関係数の定義、それらの計算ルール（和の分散の公式など）、そして $|\mathrm{Corr}| \le 1$ の証明です。

5.1　共分散と相関係数

要点：共分散・相関係数

$\mu_X = \mathrm{E}(X)$、$\mu_Y = \mathrm{E}(Y)$ として

$$\mathrm{Cov}(X,Y) = \mathrm{E}\big[(X-\mu_X)(Y-\mu_Y)\big], \qquad \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}. \tag{5.1}$$

$\mathrm{Cov}(X,Y)=0$ のとき、X と Y は**無相関**であるといいます。

数式の読み方

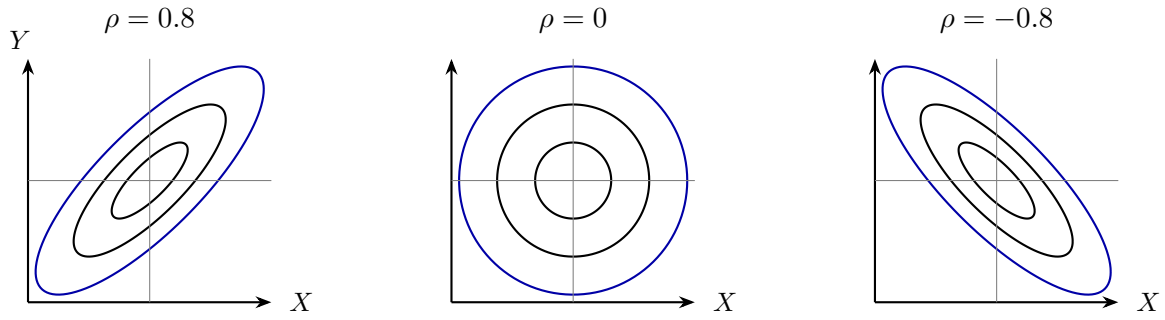
$(X-\mu_X)(Y-\mu_Y)$ の符号が、共分散の符号の意味を決めます。

- X も Y も平均より大きい、または両方とも平均より小さいとき、積は正になります。
- 片方が平均より大きく、もう片方が小さいとき、積は負になります。

その平均である共分散が正なら「X が大きいとき Y も大きい傾向」、負なら「X が大きいとき Y は小さい傾向」を表します。

なぜ相関係数を使うのか　共分散は単位に依存します。月給を「万円」から「円」に変えると Y は 10000 倍になり、共分散も 10000 倍になります。相関係数は標準偏差で割って単位をなくした（基準化した）もので、$-1 \le \mathrm{Corr}(X,Y) \le 1$ を満たし、単位を変えても値が変わりません。

図 5.1　相関係数が 0.8、0、-0.8 の2変数の等高線



平均 0・分散 1 の2変数の正規分布で、相関係数が ρ のとき、密度の等高線は $x = r\cos t$、$y = r\left(\rho\cos t + \sqrt{1-\rho^2}\sin t\right)$（$r = 1, 2, 3$）で描ける。灰色の線は平均（0）の位置。

図の読み方

1. 各パネルは、平均 0・分散 1 の2変数の分布の等高線（同じ密度の点を結んだ線）で、灰色の線は平均の位置です。
2. 相関が正（左）なら、分布は右上と左下の領域（$(X-\mu_X)(Y-\mu_Y) > 0$）に細長く伸びます。負（右）なら、左上と右下に伸びます。
3. 相関 0（中央）では、等高線が円になり、4つの領域に均等に広がります。正の積と負の積が打ち消し合うため、共分散は 0 です。

5.2　共分散の計算ルール

要点：共分散・分散の計算ルール

1. 計算公式：$\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$。
2. 特に $\mathrm{E}(X) = 0$ または $\mathrm{E}(Y) = 0$ なら $\mathrm{Cov}(X, Y) = \mathrm{E}(XY)$。
3. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$、$\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$。
4. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$。
5. $\mathrm{Cov}(a + bX,\, c + dY) = bd\,\mathrm{Cov}(X, Y)$。
6. 独立 $\Rightarrow$ 無相関（逆は一般に成り立たない。第8章）。

証明

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{E}\big[XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y\big] && \text{（公式1：括弧を展開）}\\
&= \mathrm{E}(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y && \text{（線形性。}\mu\text{ は定数）}\\
&= \mathrm{E}(XY) - \mu_X\mu_Y
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{E}\big[\{(X-\mu_X)+(Y-\mu_Y)\}^2\big] && \text{（公式4：}\mathrm{E}(X+Y)=\mu_X+\mu_Y\text{）}\\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y) && \text{（展開すると2乗の項と交差項）}
\end{aligned}$$

公式5は、$(a+bX) - \mathrm{E}(a+bX) = b(X-\mu_X)$、同様に $d(Y-\mu_Y)$ なので、積の期待値が $bd\mathrm{Cov}(X,Y)$ になることから分かります。公式6は、独立なら $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$（第4章）なので、公式1より $\mathrm{Cov} = 0$ です。

計算例：数値例での計算

$X=0$ の行は $xy=0$ なので $X=1$ の行だけが残り、

$$\mathrm{E}(XY) = 1\cdot(20\times 0.05 + 30\times 0.20 + 40\times 0.25) = 1+6+10 = 17.$$

よって $\mathrm{Cov}(X,Y) = 17 - 0.5\times 31 = 1.5 > 0$ で、大卒の方が月給が高い傾向があります。また $\mathrm{Var}(X) = \mathrm{E}(X^2) - 0.5^2 = 0.5 - 0.25 = 0.25$（$X^2 = X$ なので $\mathrm{E}(X^2) = 0.5$）、$\mathrm{E}(Y^2) = 400(0.25)+900(0.40)+1600(0.35) = 1020$ より $\mathrm{Var}(Y) = 1020 - 31^2 = 59$ です。したがって

$$\mathrm{Corr}(X,Y) = \frac{1.5}{\sqrt{0.25\times 59}} \approx 0.39.$$

発展：$|\mathrm{Corr}| \le 1$ の証明（コーシー・シュワルツの不等式）

任意の実数 t について $\mathrm{Var}(Y - tX) \ge 0$ です。公式4・5より

$$\mathrm{Var}(Y-tX) = \mathrm{Var}(Y) - 2t\mathrm{Cov}(X,Y) + t^2\mathrm{Var}(X) \ge 0.$$

これは t の2次式で、常に 0 以上なので、判別式は 0 以下です：$4\mathrm{Cov}(X,Y)^2 - 4\mathrm{Var}(X)\mathrm{Var}(Y) \le 0$。すなわち $\mathrm{Cov}(X,Y)^2 \le \mathrm{Var}(X)\mathrm{Var}(Y)$ で、両辺の平方根をとって $|\mathrm{Corr}(X,Y)| \le 1$ を得ます。等号は $Y - tX$ が定数になる（Y が X の1次関数になる）ときに限り成り立ちます。

補足：回帰分析でよく使う読み方

説明変数 $x_1, \ldots, x_K$ の1つ目が定数 $x_1 = 1$ のとき、「すべての j について $\mathrm{E}(x_j u) = 0$」は、次の2つを意味します。

- $j = 1$ のとき $x_1 = 1$ なので、$\mathrm{E}(1 \cdot u) = \mathrm{E}(u) = 0$。
- すると公式2より、$\mathrm{Cov}(x_j, u) = \mathrm{E}(x_j u) - \mathrm{E}(x_j) \cdot 0 = \mathrm{E}(x_j u) = 0$。

つまり、定数項を含む回帰では、「誤差の平均が 0」と「説明変数と誤差が無相関」が、同時に成り立つことになります。

要点

- 共分散は「一緒に動く傾向」の向きと強さ、相関係数は単位に依存しないように基準化したもので、$-1 \leq \mathrm{Corr} \leq 1$ です。
- $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$。$\mathrm{E}(X) = 0$ または $\mathrm{E}(Y) = 0$ なら $\mathrm{Cov} = \mathrm{E}(XY)$。
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$、$\mathrm{Cov}(a + bX, c + dY) = bd\mathrm{Cov}(X, Y)$。
- 独立ならば無相関ですが、逆は一般に成り立ちません。

次の章へ　ここからが回帰分析の核心です。「X の値ごとに、Y の平均はどう変わるか」を表す条件付き期待値を学びます。

大学院とのつながり　「誤差と説明変数が無相関」という条件（直交条件）は、OLS の一致性の最も基本的な仮定です。

第6章

条件付き分布と条件付き期待値

ここまでの流れ　2つの確率変数の関係は、同時分布や共分散で表せました。
この章の問い　「X の値が分かったとき、Y の平均はいくらか」を、どう定義し、どう扱えばよいのでしょうか。
この章で分かること　条件付き分布、条件付き期待値、条件付き分散、そして条件付き期待値の計算ルールです。回帰分析の出発点である「誤差項の分解」まで到達します。

6.1　条件付き分布

要点：条件付き確率・条件付き分布

$\Pr(X=x)>0$ のとき

$$\Pr(Y=y \mid X=x)=\frac{\Pr(X=x, Y=y)}{\Pr(X=x)}. \tag{6.1}$$

x を固定して y の関数とみたものを、「$X=x$ のときの Y の**条件付き分布**」といいます。連続型なら $f_{Y|X=x}(y)=f_{XY}(x,y)/f_X(x)$ です。

数式の読み方

- 「$X=x$ であると分かったときの Y の分布」です。母集団全体ではなく、$X=x$ の人たちだけに絞り込んだ部分集団の中での Y の分布です。
- 分母 $\Pr(X=x)$ で割るのは、絞り込んだ部分集団の中で確率の合計を 1 にするためです。実際、$\sum_y \Pr(Y=y \mid X=x)=\Pr(X=x)/\Pr(X=x)=1$ です。

計算例：数値例の条件付き分布

$X=1$（大卒）の行 $(0.05, 0.20, 0.25)$ を $\Pr(X=1)=0.5$ で割ると、大卒者の月給の分布は、$(20, 30, 40)$ に対して $(0.1, 0.4, 0.5)$（合計 1）です。非大卒者は $(0.4, 0.4, 0.2)$ となります。

6.2 条件付き期待値

要点：条件付き期待値

$$\mathrm{E}(Y \mid X = x) = \sum_{y} y \Pr(Y = y \mid X = x) \qquad \Big(\text{連続型なら } \int y\, f_{Y|X=x}(y)\, dy\Big). \quad (6.2)$$

条件付き分布を使って計算した期待値で、「$X = x$ のグループの中での Y の平均」です。

計算例：数値例の条件付き期待値

$$\mathrm{E}(Y \mid X = 0) = 20(0.4) + 30(0.4) + 40(0.2) = 28,$$

$$\mathrm{E}(Y \mid X = 1) = 20(0.1) + 30(0.4) + 40(0.5) = 34.$$

非大卒の平均月給は 28 万円、大卒は 34 万円です。

要点：$\mathrm{E}(Y \mid X)$ は確率変数

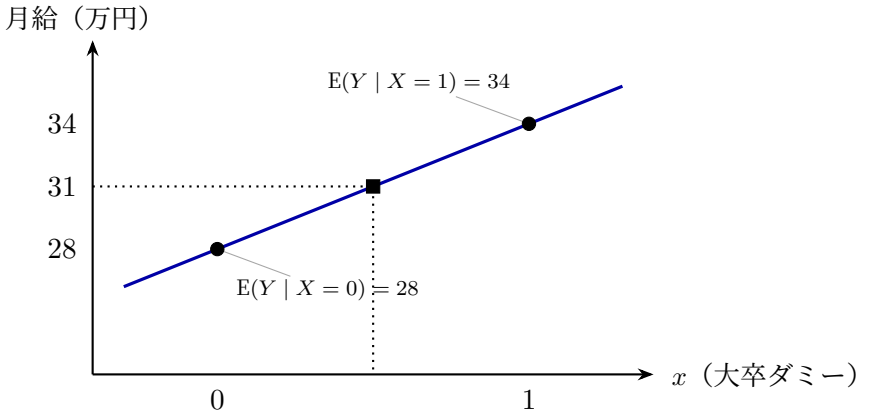
関数 $\mu(x) = \mathrm{E}(Y \mid X = x)$ を定義して、その x に確率変数 X を代入したもの $\mu(X)$ を $\mathrm{E}(Y \mid X)$ と書きます。$\mathrm{E}(Y \mid X)$ は「X の関数」であり、それ自体が確率変数です。$\mathrm{E}(Y \mid X = x)$（x に具体的な値を入れたもの、ただの数）と $\mathrm{E}(Y \mid X)$（確率変数）の区別が大切です。

数値例では、$\mathrm{E}(Y \mid X)$ は $X = 0$ なら 28、$X = 1$ なら 34 という値をとる確率変数で、

$$\mathrm{E}(Y \mid X) = 28 + 6X$$

と書けます。この関数 $\mu(x)$ は **回帰関数**とも呼ばれます。

図 6.1　数値例の条件付き期待値 $\mathrm{E}(Y \mid X = x)$



条件付き期待値 $\mathrm{E}(Y \mid X = 0) = 28$、$\mathrm{E}(Y \mid X = 1) = 34$ を通る直線 $28 + 6x$ を描いた。$x = 0.5$ での値 31 は、全体の平均 $\mathrm{E}(Y) = 31$ に等しい。

図の読み方

1. 黒い丸が、非大卒（$x = 0$）と大卒（$x = 1$）のグループごとの平均月給です。
2. 青い直線は、2点を通る直線 $28 + 6x$ です。X がダミー変数のときは、2点を通る直線が必ず引けるので、条件付き期待値は必ず線形で書けます。
3. 黒い四角は、$x = 0.5$（2グループの比率が半々）での直線の値で、全体の平均 $\mathrm{E}(Y) = 31$ に一致します。これは次章の繰り返し期待値の法則の例です。

6.3　条件付き分散

条件付き分布を使えば、分散も条件付きで定義できます。

$$\mathrm{Var}(Y \mid X = x) = \mathrm{E}\Big[\big(Y - \mathrm{E}(Y \mid X = x)\big)^2 \,\Big|\, X = x\Big] = \mathrm{E}(Y^2 \mid X = x) - \big[\mathrm{E}(Y \mid X = x)\big]^2. \tag{6.3}$$

「$X = x$ のグループの中での Y のばらつき」です。グループによってばらつきが異なることを**不均一分散**といい、回帰分析の標準誤差の議論で重要になります。

計算例：数値例の条件付き分散

$\mathrm{Var}(Y \mid X = 0) = 56$、$\mathrm{Var}(Y \mid X = 1) = 44$ です（練習問題 問2）。2つが異なるので、この例は不均一分散の状況です。

6.4 条件付き期待値の性質

要点：条件付き期待値の性質

1. 線形性：$\mathrm{E}(a + bY + cZ \mid X) = a + b\,\mathrm{E}(Y \mid X) + c\,\mathrm{E}(Z \mid X)$。
2. X の関数は「定数のように」外に出せる：$\mathrm{E}\big[g(X)Y \mid X\big] = g(X)\,\mathrm{E}(Y \mid X)$。
3. 特に $\mathrm{E}\big[g(X) \mid X\big] = g(X)$。
4. X と Y が独立なら $\mathrm{E}(Y \mid X) = \mathrm{E}(Y)$。

性質2の証明（離散型）　$X = x$ と分かっている世界では、$g(X) = g(x)$ はただの数です。

$$
\begin{aligned}
\mathrm{E}[g(X)Y \mid X = x] &= \sum_{y} g(x)\, y \Pr(Y = y \mid X = x) && (g(X) = g(x)) \\
&= g(x) \sum_{y} y \Pr(Y = y \mid X = x) && (g(x) \text{ を外へ}) \\
&= g(x)\,\mathrm{E}(Y \mid X = x)
\end{aligned}
$$

これがすべての x で成り立つので、x に X を代入して性質2を得ます。性質3は $Y = 1$ とした場合です。性質4は、独立なら条件付き分布が周辺分布に等しいことから従います。

6.5 誤差項の分解

任意の y について $u := y - \mathrm{E}(y \mid x)$ と定義すると

$$
y = \mathrm{E}(y \mid x) + u, \qquad \mathrm{E}(u \mid x) = 0 \tag{6.4}
$$

が自動的に成り立ちます。実際、性質1と3から（$\mathrm{E}(y \mid x)$ は x の関数なので性質3が使えます）、

$$
\begin{aligned}
\mathrm{E}(u \mid x) &= \mathrm{E}(y \mid x) - \mathrm{E}\big[\mathrm{E}(y \mid x) \mid x\big] && (\text{性質1（線形性）}) \\
&= \mathrm{E}(y \mid x) - \mathrm{E}(y \mid x) = 0 && (\text{性質3：}\mathrm{E}(y \mid x) \text{ は } x \text{ の関数})
\end{aligned}
$$

数式の読み方

どんな y も、「x で説明できる部分 $\mathrm{E}(y \mid x)$」と「x では平均的に説明できない部分 u」に分解できます。これが回帰分析の出発点です。

補足：つまずきポイント

$\mathrm{E}(u \mid x) = 0$ が「自動的に成り立つ」のは、u を $y - \mathrm{E}(y \mid x)$ と定義したからです。回帰モデル $y = x\beta + u$ で $\mathrm{E}(u \mid x) = 0$ を「仮定する」とき、本当に仮定しているのは、「$\mathrm{E}(y \mid x)$ が $x\beta$ という線形の形をしている」ことです。

発展：条件付き期待値は最良の予測

x の関数 $g(x)$ で y を予測するとき、平均2乗誤差 $\mathrm{E}[(y - g(x))^2]$ を最小にするのは $g(x) = \mathrm{E}(y \mid x)$ です。実際、$y - g(x) = u + \{\mathrm{E}(y \mid x) - g(x)\}$ と分け、$h(x) = \mathrm{E}(y \mid x) - g(x)$ とおくと、次章の結果（$\mathrm{E}[u\,h(x)] = 0$）より、交差項が消えて

$$\mathrm{E}[(y - g(x))^2] = \mathrm{E}(u^2) + \mathrm{E}[h(x)^2] \geq \mathrm{E}(u^2)$$

となります。等号は $g(x) = \mathrm{E}(y \mid x)$ のとき成り立ちます。予測の候補を線形関数 xb に限ったときの最良予測は、**線形射影**と呼ばれ、$\beta = \arg\min_b \mathrm{E}[(y - xb)^2]$ で定まります。

要点

- 条件付き分布は「$X = x$ の人たちだけに絞った Y の分布」で、同時確率を周辺確率で割って得ます。
- $\mathrm{E}(Y \mid X = x)$ は x ごとの平均（ただの数）、$\mathrm{E}(Y \mid X)$ は X の関数としての確率変数です。
- X の関数は条件付き期待値の外に出せます（$\mathrm{E}[g(X)Y \mid X] = g(X)\mathrm{E}(Y \mid X)$）。
- $u = y - \mathrm{E}(y \mid x)$ とおけば、$y = \mathrm{E}(y \mid x) + u$、$\mathrm{E}(u \mid x) = 0$ が自動的に成り立ちます。

次の章へ　グループごとの平均を、グループの比率で重みづけして平均すると、全体の平均に戻ります。これを定理にしたのが、次の章の繰り返し期待値の法則です。

大学院とのつながり　線形射影と条件付き期待値の違いは、回帰分析の仮定（何を推定しているのか）を考えるうえで中心的です。

第7章

繰り返し期待値の法則（LIE）

ここまでの流れ　条件付き期待値 $\mathrm{E}(Y \mid X)$ は「X のグループごとの平均」で、X の関数（確率変数）でした。
この章の問い　グループごとの平均と、全体の平均は、どう結びつくのでしょうか。「誤差が説明変数と無関係」という条件は、どんな帰結を持つのでしょうか。
この章で分かること　繰り返し期待値の法則、その最重要の応用（$\mathrm{E}(u \mid x) = 0$ から $\mathrm{E}(u) = 0$、$\mathrm{E}(xu) = 0$ などが従うこと）、そして全分散の公式です。

7.1　繰り返し期待値の法則

要点：繰り返し期待値の法則（Law of Iterated Expectations）

$$\mathrm{E}\big[\mathrm{E}(Y \mid X)\big] = \mathrm{E}(Y). \tag{7.1}$$

数式の読み方

内側の $\mathrm{E}(Y \mid X)$ は X の関数（確率変数）でした。その期待値（X の分布についての平均）をとると、Y の全体の平均に戻る、という定理です。「グループごとの平均」を「グループの比率」で加重平均すると、全体の平均に戻る、ということです。

計算例：数値例での確認

$\mathrm{E}(Y \mid X)$ は確率 0.5 で 28、確率 0.5 で 34 をとる確率変数なので

$$\mathrm{E}\big[\mathrm{E}(Y \mid X)\big] = 28\Pr(X=0) + 34\Pr(X=1) = 28(0.5) + 34(0.5) = 31 = \mathrm{E}(Y).$$

離散型での証明

$$\begin{aligned}
\mathrm{E}\big[\mathrm{E}(Y \mid X)\big] &= \sum_x \mathrm{E}(Y \mid X = x)\Pr(X = x) && \text{（X の関数の期待値（第2章の性質4））} \\
&= \sum_x \sum_y y \frac{\Pr(X = x, Y = y)}{\Pr(X = x)} \Pr(X = x) && \text{（条件付き期待値の定義）} \\
&= \sum_y y \sum_x \Pr(X = x, Y = y) && \text{（分母が消え、和の順序を入れ替える）} \\
&= \sum_y y \Pr(Y = y) = \mathrm{E}(Y) && \text{（周辺分布の定義）}
\end{aligned}$$

連続型でも、和を積分に、確率を密度に置き換えれば同じ計算になります。

7.2 LIE の最重要の応用

要点：$\mathrm{E}(u \mid x) = 0$ から従うこと

$\mathrm{E}(u \mid x) = 0$ ならば

(a) $\mathrm{E}(u) = 0$、

(b) $\mathrm{E}(xu) = 0$、

(c) $\mathrm{Cov}(x, u) = 0$。

さらに一般に、x のどんな関数 $h(x)$ についても $\mathrm{E}[h(x)u] = 0$ です。

$$
\begin{aligned}
&\text{(a)}\quad \mathrm{E}(u) = \mathrm{E}\big[\mathrm{E}(u \mid x)\big] = \mathrm{E}[0] = 0 && \text{（LIE）}\\
&\text{(b)}\quad \mathrm{E}(xu) = \mathrm{E}\big[\mathrm{E}(xu \mid x)\big] && \text{（LIE）}\\
&\qquad\qquad = \mathrm{E}\big[x\,\mathrm{E}(u \mid x)\big] = \mathrm{E}[x \cdot 0] = 0 && \text{（x は条件付き期待値の外に出せる）}\\
&\text{(c)}\quad \mathrm{Cov}(x, u) = \mathrm{E}(xu) - \mathrm{E}(x)\mathrm{E}(u) = 0 && \text{（(a)(b) と共分散の計算公式）}
\end{aligned}
$$

一般の $h(x)$ でも (b) と同じ計算が成り立ちます。

補足：OLS の不偏性の証明で使う形

推定量 $\hat{\beta}$ について、説明変数 X を条件にした期待値が $\mathrm{E}(\hat{\beta} \mid X) = \beta$ と示せたとします。このとき LIE により

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}\big[\mathrm{E}(\hat{\beta} \mid X)\big] = \mathrm{E}[\beta] = \beta$$

の1行で、条件を外した不偏性 $\mathrm{E}(\hat{\beta}) = \beta$ が得られます。

補足：LIE の一般形

情報の細かさが違う2つの条件付けについて、$\mathrm{E}[y \mid x] = \mathrm{E}\big[\mathrm{E}[y \mid w] \mid x\big]$（$x$ が w で決まるとき）が成り立ちます。「細かい情報 w で条件付けてから、粗い情報 x で条件付け直すと、最初から x で条件付けたのと同じ」という意味です。x として「何もない（定数）」をとると、上の LIE になります。

7.3 全分散の公式

発展：全分散の公式

LIE の分散版として、次が成り立ちます。

$$\mathrm{Var}(Y) = \underbrace{\mathrm{E}\big[\mathrm{Var}(Y \mid X)\big]}_{\text{グループ内のばらつきの平均}} + \underbrace{\mathrm{Var}\big[\mathrm{E}(Y \mid X)\big]}_{\text{グループ平均どうしのばらつき}} .$$

数値例では、$\mathrm{E}[\mathrm{Var}(Y \mid X)] = 0.5(56) + 0.5(44) = 50$、$\mathrm{Var}(\mathrm{E}(Y \mid X)) = \mathrm{Var}(28 + 6X) = 36 \times 0.25 = 9$ で、合計 $59 = \mathrm{Var}(Y)$ です。右辺第2項の割合 $9/59 \approx 0.15$ は「Y のばらつきのうち X で説明できる割合」で、回帰分析の決定係数 R^2 の母集団版にあたります。

要点

- 繰り返し期待値の法則：$\mathrm{E}[\mathrm{E}(Y \mid X)] = \mathrm{E}(Y)$。グループ平均を比率で加重平均すると全体の平均に戻ります。
- $\mathrm{E}(u \mid x) = 0$ なら、$\mathrm{E}(u) = 0$、$\mathrm{E}(xu) = 0$、$\mathrm{Cov}(x, u) = 0$、さらに任意の $h(x)$ で $\mathrm{E}[h(x)u] = 0$ です。
- 条件付き期待値で示した性質（例：不偏性）から、条件を外した性質は LIE で一気に得られます。

次の章へ　「u と x に関係がない」という条件は、実は強さの違う3つの段階があります。次の章で、独立・平均独立・無相関を整理します。

大学院とのつながり　LIE は、不偏性の証明、誤差の仮定の整理、条件付き分散の分解など、大学院の議論で最も頻繁に使う道具の1つです。

第8章

独立・平均独立・無相関：強さの違い

ここまでの流れ　独立（第4章）、無相関（第5章）、$\mathrm{E}(u \mid x) = 0$（第6・7章）と、「関係がない」ことを表す条件がいくつか登場しました。
この章の問い　これらは同じ意味でしょうか。違うなら、どの条件がどの条件を含むのでしょうか。
この章で分かること　独立・平均独立・無相関の3段階の強さの違いと、逆向きの含意が成り立たないことを示す反例です。

8.1　3つの条件

u と x の「関係がない」ことを表す条件には、強さの異なる3段階があります。

条件	式	意味
独立（最も強い）	u の分布が x に依存しない	x を知っても、u の分布（平均・ばらつき・形）について何も分からない
平均独立	$\mathrm{E}(u \mid x) = \mathrm{E}(u)\ (= 0)$	x を知っても、u の平均は分からない（ばらつきは x で変わってよい）
無相関・直交（最も弱い）	$\mathrm{E}(xu) = 0$（$\mathrm{E}(u) = 0$ のもとで $\mathrm{Cov}(x, u) = 0$）	u と x の間に直線的な関係がない（曲がった関係はあってよい）

要点：含意の向き

$$\text{独立} \implies \text{平均独立} \implies \text{無相関}$$

逆向きの矢印は、一般には成り立ちません。

矢印の証明

- **独立 ⇒ 平均独立**：独立なら u の条件付き分布は周辺分布と同じなので、その期待値も同じです：$\mathrm{E}(u \mid x) = \mathrm{E}(u)$（第6章の性質4）。

- **平均独立 ⇒ 無相関**：第7章の (b)(c) です（$\mathrm{E}(u \mid x) = 0$ なら $\mathrm{E}(xu) = 0$、$\mathrm{Cov}(x, u) = 0$）。

8.2 反例1：平均独立だが独立ではない

計算例：数値例

$u = Y - \mathrm{E}(Y \mid X)$ とすると、第6章の分解より $\mathrm{E}(u \mid X) = 0$ なので、平均独立です。しかし

$$\mathrm{Var}(u \mid X = 0) = \mathrm{Var}(Y \mid X = 0) = 56, \qquad \mathrm{Var}(u \mid X = 1) = 44$$

で、u のばらつきが X によって変わるので、独立ではありません。不均一分散がある状況は、いつもこの例にあたります。

8.3 反例2：無相関だが平均独立ではない

x が $-1, 0, 1$ をそれぞれ確率 $1/3$ でとり、$u = x^2 - \frac{2}{3}$ とします。

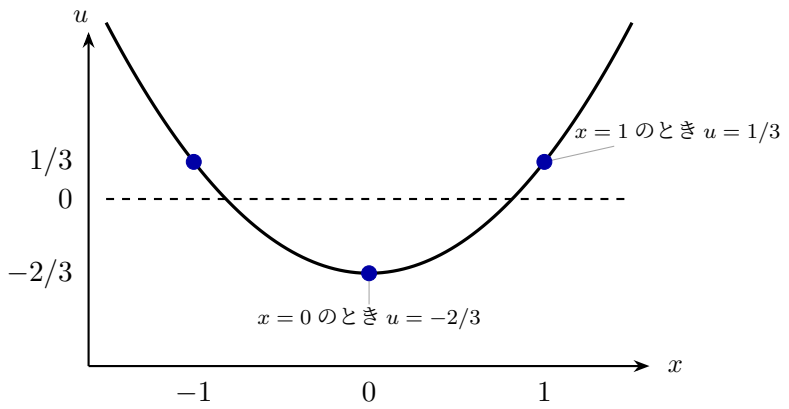
$$\mathrm{E}(u) = \tfrac{1}{3}(1 + 0 + 1) - \tfrac{2}{3} = 0 \qquad (\mathrm{E}(x^2) = 2/3)$$
$$\mathrm{E}(xu) = \mathrm{E}(x^3) - \tfrac{2}{3}\mathrm{E}(x) = 0 - 0 = 0 \qquad (x^3 \text{ も } -1, 0, 1 \text{ をとるので平均 } 0)$$

よって、u と x は無相関です。しかし

$$\mathrm{E}(u \mid x) = x^2 - \tfrac{2}{3} = \begin{cases} -\frac{2}{3} & (x = 0) \\ \frac{1}{3} & (x = \pm 1) \end{cases}$$

は x によって変わるので、平均独立ではありません。

図 8.1　反例2：$u = x^2 - \frac{2}{3}$ は x と無相関だが、関係がある



$u = x^2 - 2/3$ の曲線と、$x \in \{-1, 0, 1\}$ でとりうる3点（青）、および $\mathrm{Cov}(x, u) = 0$ から決まる最良の直線（傾き $\mathrm{Cov}(x, u)/\mathrm{Var}(x) = 0$、水平な破線）を描いた。

図の読み方

1. 黒い曲線が $u = x^2 - 2/3$ で、青い点が $x = -1, 0, 1$ のときにとりうる値です。
2. u は x によって明らかに変わっています（$x = 0$ では $-2/3$、$x = \pm 1$ では $1/3$）。
3. ところが、点に最もよく当てはまる直線は水平な破線（傾き 0）で、共分散は 0 になってしまいます。共分散が測るのは「直線的な関係」だけで、このような曲がった関係を見落とすからです。

計算例：線形射影と条件付き期待値の違い

反例2で $y = x^2$ とすると、$\mathrm{E}(y \mid x) = x^2$ は非線形です。ところが、定数と x だけを使った線形の最良予測（線形射影）は、傾きが $\mathrm{Cov}(x, y)/\mathrm{Var}(x) = 0$（$\mathrm{Cov}(x, x^2) = \mathrm{E}(x^3) - \mathrm{E}(x)\mathrm{E}(x^2) = 0$）、切片が $\mathrm{E}(y) = 2/3$ で、定数 $2/3$ になります。つまり、線形射影は常に定義できますが、$\mathrm{E}(y \mid x)$ が線形だとは限りません。

補足：回帰分析での使い分け

線形射影を推定したいなら、無相関（$\mathrm{E}(xu) = 0$）は定義上自動的に成り立ちます。一方、条件付き期待値そのものを推定したいなら、より強い平均独立（$\mathrm{E}(u \mid x) = 0$）を正当化する必要があります。不偏性には後者が必要で、一致性には前者で足ります。

要点

- 強さの順は、独立 > 平均独立 > 無相関。独立なら平均独立、平均独立なら無相関です。
- 逆は成り立ちません。不均一分散は「平均独立だが独立ではない」例、曲がった関係は「無相関だが平均独立ではない」例です。
- 共分散は直線的な関係しか測りません。
- 一致性には無相関（直交）、不偏性には平均独立が必要という違いが、回帰分析の仮定の核心です。

次の章へ　ここまでで、確率の言葉（分布・期待値・条件付き期待値）がそろいました。次の章では、統計的推論で最もよく使う2つの分布、正規分布とカイ二乗分布を学びます。

大学院とのつながり　内生性とは「無相関（直交）が成り立たないこと」です。この章の3段階を区別できると、大学院での仮定の議論が読みやすくなります。

第9章

重要な分布：正規分布とカイ二乗分布

ここまでの流れ　確率変数の分布を、期待値・分散・条件付き期待値などで特徴づけられるようになりました。
この章の問い　検定や信頼区間で使う「1.96」や「3.84」という数字は、どの分布のどんな点なのでしょうか。
この章で分かること　正規分布の性質（標準化・対称性・68–95–99.7 則）、臨界値 1.96 の意味、そしてカイ二乗分布の定義・性質・臨界値です。

9.1　正規分布

要点：正規分布

平均 μ、分散 σ^2 の**正規分布** $\mathrm{Normal}(\mu, \sigma^2)$ の密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right). \tag{9.1}$$

$\mu = 0$、$\sigma^2 = 1$ のものを**標準正規分布**といい、その CDF を Φ と書きます。

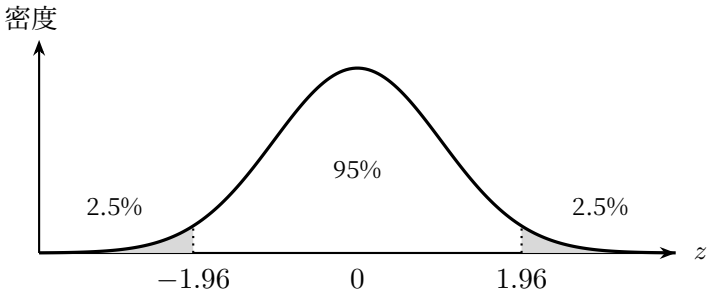
数式の読み方

- $\exp(\cdot)$ の中身 $-(x-\mu)^2/(2\sigma^2)$ は、x が平均 μ から離れるほど急速に小さくなります。そのため、密度は μ を頂点とする左右対称の釣鐘型になります。
- σ^2 が大きいほど、裾が広がります。
- 前の係数 $1/\sqrt{2\pi\sigma^2}$ は、全体の面積を 1 にするための調整です。

要点：正規分布の性質

1. **線形変換しても正規**：$X \sim \mathrm{Normal}(\mu, \sigma^2)$ なら $a+bX \sim \mathrm{Normal}(a+b\mu,\, b^2\sigma^2)$。
2. **標準化**：$Z = (X-\mu)/\sigma \sim \mathrm{Normal}(0,1)$。どんな正規分布の確率も Φ だけで計算できます：$\Pr(X \le c) = \Phi\big((c-\mu)/\sigma\big)$。
3. **対称性**：$\Phi(-z) = 1 - \Phi(z)$。
4. **68–95–99.7 則**：$\Pr(|Z| \le 1) \approx 0.683$、$\Pr(|Z| \le 2) \approx 0.954$、$\Pr(|Z| \le 3) \approx 0.997$。

図 9.1　標準正規分布の密度と、$|z| > 1.96$ の裾



標準正規分布の密度 $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ を描いた。$\phi(1.96) \approx 0.0584$。灰色の両裾の面積は、それぞれ約 0.025。

図の読み方

1. 白い部分 $[-1.96, 1.96]$ の面積が 0.95、両側の灰色部分がそれぞれ 0.025 です。
2. したがって、左端から 1.96 までの面積（CDF の値）は $\Phi(1.96) = 0.95 + 0.025 = 0.975$ です。
3. 言い換えると、1.96 は「その左側に 97.5% がある点」、つまり 0.975 分位点 $\Phi^{-1}(0.975)$ です（Φ^{-1} は Φ の逆関数で、確率を与えると分位点を返す）。

両側の有意水準 α	0.10	0.05	0.01
$\Phi^{-1}(1-\alpha/2)$	1.645	1.960	2.576

数式の読み方

信頼区間 $[\hat{\beta} \pm 1.96 \cdot \mathrm{se}(\hat{\beta})]$ や、「$|t| > 1.96$ なら有意水準 5% で棄却」という検定の「1.96」は、ここから来ています。α は有意水準で、例えば $\alpha = 0.05$ なら $\Phi^{-1}(0.975) = 1.96$ です。

9.2　カイ二乗分布

要点：カイ二乗分布

$Z_1, \ldots, Z_Q$ が互いに独立に $\mathrm{Normal}(0,1)$ に従うとき、

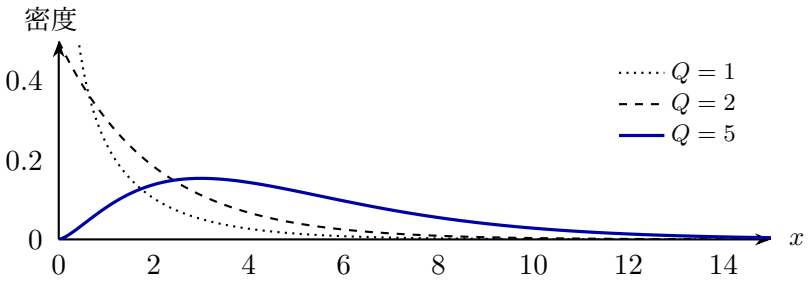
$$Z_1^2 + Z_2^2 + \cdots + Z_Q^2 \sim \chi^2_Q \quad (\text{自由度 } Q \text{ のカイ二乗分布}). \tag{9.2}$$

2乗の和なので、常に 0 以上の値をとり、右に裾が長い形をしています。平均は

$$\mathrm{E}(Z_q^2) = \mathrm{Var}(Z_q) = 1 \implies \mathrm{E}(\chi^2_Q) = Q$$

（分散は $2Q$）です。自由度 Q が大きいほど、分布は右に移り、広がります。

図 9.2　自由度 $Q = 1, 2, 5$ のカイ二乗分布の密度



χ^2_Q の密度：$Q = 1$ は $\frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2}$、$Q = 2$ は $\frac{1}{2} e^{-x/2}$、$Q = 5$ は $\frac{1}{3\sqrt{2\pi}} x^{3/2} e^{-x/2}$（係数は約 0.133）。$Q = 1$ は $x \to 0$ で密度が発散するため、縦軸の上限で切って描いた。

図の読み方

1. 点線が $Q = 1$、破線が $Q = 2$、青の実線が $Q = 5$ です。
2. $Q = 1$ では、0 の近くに大きく集中します（Z^2 は小さい値をとりやすい）。
3. Q が大きくなると、山が右に移ります（平均が Q）。

数式の読み方

複数の制約をまとめて検定する Wald 検定では、この分布の右裾 5% の境界（0.95 分位点）を、棄却の基準に使います。統計表の値は、$Q = 1$ で 3.84（$= 1.96^2$）、$Q = 2$ で 5.99、$Q = 5$ で 11.07 です。$Q = 1$ のとき、$|t| > 1.96$ と $t^2 > 3.84$ は同じことです。

要点

- 正規分布は、線形変換しても正規で、標準化 $Z = (X - \mu)/\sigma$ により標準正規分布の表だけで確率が計算できます。
- 1.96 は標準正規分布の 0.975 分位点で、両側 5% の検定・95% 信頼区間の基準です。
- カイ二乗分布 χ^2_Q は、独立な標準正規変数 Q 個の2乗和で、平均は Q。χ^2_1 の 0.95 分位点 3.84 は 1.96^2 です。

次の章へ　最後に、「母集団」「標本」「推定量」の関係を整理し、標本平均が持つ性質を導きます。

大学院とのつながり　漸近理論では、推定量が $N \to \infty$ で正規分布に近づくこと（中心極限定理）と、検定統計量が χ^2 分布に近づくことを示します。この章の分布が、そこでの「行き先」です。

第10章

母集団・標本・推定量

ここまでの流れ　確率変数・分布・期待値・分散・条件付き期待値と、「母集団」を記述する言葉がそろいました。
この章の問い　母集団の量（平均や回帰係数）は未知です。手元のデータ（標本）から、それをどう推測し、その良し悪しをどう議論すればよいのでしょうか。
この章で分かること　パラメータ・標本・推定量・推定値の区別、標本平均の期待値と分散、そして標本分散がなぜ $N-1$ で割られるのかです。

10.1　用語の整理

要点：用語

- **母集団の量（パラメータ）**：$\mu = \mathrm{E}(Y)$ や β のように、母集団の分布で決まる未知の定数。
- **標本**：母集団から無作為に抽出したデータ $Y_1, \ldots, Y_N$（IID）。
- **推定量**：標本から計算するルール（式）。例：標本平均 $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$。
- **推定値**：実際のデータを推定量に代入して得られた数。
- **不偏性**：$\mathrm{E}(\hat{\theta}) = \theta$。

推定量は確率変数　標本 $Y_1, \ldots, Y_N$ が確率変数なので、そこから計算される推定量 $\bar{Y}$ も確率変数です。別の標本を抽出すれば、別の値になります。推定量の分布（**標本分布**）を考えることで、「推定がどれくらい正確か」を議論できます。

補足：つまずきポイント

「パラメータ（定数）」「推定量（確率変数）」「推定値（ただの数）」の3つを区別してください。例えば、μ は未知だが動かない定数、$\bar{Y}$ は標本によって変わる確率変数、「今回のデータで計算した $\bar{Y} = 30.8$」は1つの実現値です。期待値や分散を考えられるのは、確率変数である推定量だけです。

10.2 標本平均の性質

要点：標本平均の期待値と分散

Y_i が IID で、$\mathrm{E}(Y_i) = \mu$、$\mathrm{Var}(Y_i) = \sigma^2$ のとき

$$\mathrm{E}(\bar{Y}) = \mu \ \text{(不偏)}, \qquad \mathrm{Var}(\bar{Y}) = \frac{\sigma^2}{N}. \tag{10.1}$$

$$\begin{aligned}
\mathrm{E}(\bar{Y}) &= \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}(Y_i) && \text{（線形性（独立性は不要））} \\
&= \frac{1}{N}\cdot N\mu = \mu \\
\mathrm{Var}(\bar{Y}) &= \frac{1}{N^2}\mathrm{Var}\Big(\sum_{i=1}^{N} Y_i\Big) && \text{（}\mathrm{Var}(bX) = b^2\mathrm{Var}(X)\text{、}b = 1/N\text{）} \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\mathrm{Var}(Y_i) && \text{（独立なら共分散がすべて 0）} \\
&= \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}
\end{aligned}$$

分散の計算の2つ目の等号は、第5章の公式 4 を N 個に拡張した

$$\mathrm{Var}\Big(\sum Y_i\Big) = \sum \mathrm{Var}(Y_i) + 2\sum_{i<j}\mathrm{Cov}(Y_i, Y_j)$$

で、独立なら共分散がすべて 0 になることを使っています。

数式の読み方

- $\mathrm{Var}(\bar{Y}) = \sigma^2/N$ は、N を大きくすると 0 に近づきます。標本が大きいほど、$\bar{Y}$ は μ の近くに集中します。これを厳密にしたのが大数の法則です。
- $\bar{Y}$ の分布の「形」が正規分布に近づくことを述べたのが、中心極限定理です。
- $\bar{Y}$ の標準偏差 $\sigma/\sqrt{N}$ は「推定の誤差の典型的な大きさ」で、σ を推定値で置き換えたものが**標準誤差**です。
- $\sqrt{N}(\bar{Y} - \mu)$ の分散は $N \cdot \sigma^2/N = \sigma^2$ となり、N が増えても 0 にも無限大にもなりません。漸近理論で $\sqrt{N}$ が登場するのは、このためです。

発展：標本分散はなぜ $N-1$ で割るのか

$\sum(Y_i-\bar{Y})^2$ を、真の平均 μ を使って展開します。

$$\begin{aligned}\sum(Y_i-\bar{Y})^2 &= \sum\big[(Y_i-\mu)-(\bar{Y}-\mu)\big]^2 && (\mu \text{ を足し引き})\\ &= \sum(Y_i-\mu)^2 - 2(\bar{Y}-\mu)\sum(Y_i-\mu) \\ &\quad + N(\bar{Y}-\mu)^2 && (\text{展開})\\ &= \sum(Y_i-\mu)^2 - N(\bar{Y}-\mu)^2 && \big(\textstyle\sum(Y_i-\mu)=N(\bar{Y}-\mu)\big)\end{aligned}$$

期待値をとると、$\mathrm{E}[(Y_i-\mu)^2]=\sigma^2$、$\mathrm{E}[(\bar{Y}-\mu)^2]=\mathrm{Var}(\bar{Y})=\sigma^2/N$ なので

$$\mathrm{E}\Big[\sum_{i=1}^{N}(Y_i-\bar{Y})^2\Big] = N\sigma^2 - N\cdot\frac{\sigma^2}{N} = (N-1)\sigma^2.$$

よって $s^2=\frac{1}{N-1}\sum(Y_i-\bar{Y})^2$ は σ^2 の不偏推定量です。N で割ると少し小さめに偏ります。真の μ の代わりに、データに合わせて決めた $\bar{Y}$ からのずれを測っているためです。回帰分析で誤差分散の推定に使う $\frac{1}{N-2}\sum\hat{u}_i^2$ の「$N-2$」も、同じ理由です。

数式の読み方：標本対応物

母集団の期待値 $\mathrm{E}(\cdot)$ を、標本平均 $\frac{1}{N}\sum_{i=1}^{N}(\cdot)$ で置き換える操作を**標本対応物**（sample analogue）といいます。$\mathrm{E}(Y)\to\bar{Y}$ がその最も簡単な例です。計量経済学の多くの推定量は、母集団で成り立つ関係式の標本対応物として作られます。

要点

- パラメータ（未知の定数）、推定量（確率変数）、推定値（ただの数）を区別します。
- 標本平均は不偏で、分散は σ^2/N です。標本が大きいほど母平均の近くに集中します。
- 標本平均の標準偏差 $\sigma/\sqrt{N}$ が推定の誤差の目安で、σ を推定したものが標準誤差です。
- 標本分散は N ではなく $N-1$ で割ると不偏になります。

次の章へ　最後に、記号の読み方を一覧にまとめ、練習問題で理解を確かめます。

大学院とのつながり　$N\to\infty$ での標本平均の振る舞い（大数の法則・中心極限定理）を厳密に扱うのが、漸近理論です。推定量の性質を議論する言葉は、すべてこの章の延長にあります。

第11章

記号の読み方と練習問題

ここまでの流れ　確率変数から標本平均まで、確率・統計の言葉を順に学びました。
この章の問い　学んだ記号を正しく読めるでしょうか。計算や証明を自分でできるでしょうか。
この章で分かること　記号の読み方の一覧と、練習問題（解答つき）です。

11.1　記号の読み方まとめ

記号	意味
$\mathrm{E}(\cdot)$, $\mathrm{E}[\cdot]$	期待値。丸括弧と角括弧に意味の違いはない
$\mathrm{E}(y \mid x)$	x を条件とする y の条件付き期待値。縦棒「$\mid$」は「〜が与えられたとき」
Var, Cov, Corr	分散、共分散、相関係数
$\Pr(\cdot)$	確率
$\sum_{i=1}^{N}$	$i=1$ から N までの和
$\int_{-\infty}^{\infty} \cdots dx$	積分（連続型の「和」）
$X \sim F$	X は分布 F に従う
$\mathrm{Normal}(\mu, \sigma^2)$, χ^2_Q	正規分布、自由度 Q のカイ二乗分布
Φ, Φ^{-1}	標準正規分布の CDF とその逆関数（分位点）
IID	独立同一分布
$\hat{\beta}$（ハット）、$\bar{Y}$（バー）	推定量（データから計算したもの）、標本平均
$:=$	「左辺を右辺で定義する」
$\Rightarrow$, $\Leftrightarrow$	「ならば」、「同値（必要十分）」
$\mathbb{R}$, $\mathbb{R}^K$	実数全体、K 個の実数を並べたベクトル全体
$\arg\min_b g(b)$	$g(b)$ を最小にする b の値（最小値そのものではない）
$\mathbf{1}(A)$	指示関数。A が正しければ 1、そうでなければ 0

11.2 練習問題

数値例（第4章の表）を使う問題は、その表を参照してください。

問1 $\mathrm{Var}(a+bX) = b^2\mathrm{Var}(X)$ を、分散の定義と期待値の線形性から示せ。

問2 数値例の表を使って、$\mathrm{Var}(Y \mid X=0)$ と $\mathrm{Var}(Y \mid X=1)$ を計算せよ（$\mathrm{Var}(Y \mid X=x) = \mathrm{E}(Y^2 \mid X=x) - [\mathrm{E}(Y \mid X=x)]^2$）。2つが異なることは何を意味するか。

問3 数値例で $u = Y - \mathrm{E}(Y \mid X)$ と定義する。$\mathrm{E}(u)=0$ と $\mathrm{E}(Xu)=0$ を、表から直接計算して確かめよ。

問4 $\mathrm{E}(u \mid x)=0$ のとき、$\mathrm{E}(x^2u)=0$ も成り立つことを示せ。

問5 Y_i が IID で $\mathrm{Var}(Y_i)=59$ のとき、$N=100$ の標本平均 $\bar{Y}$ の標準偏差を求めよ。

問6 「$\mathrm{E}(xu)=0$ なら $\mathrm{E}(u \mid x)=0$ である」は正しいか。理由を述べよ。

問7 $\mathrm{Var}(X-Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y)$ を、第5章の公式 4・5 を使って示せ。数値例で $\mathrm{Var}(Y-20X)$ を計算せよ。

問8 $X \sim \mathrm{Normal}(31, 59)$ のとき、$\Pr\big(X \le 31 + 1.96\sqrt{59}\big)$ を求めよ。

11.3 解答

問1

$$
\begin{aligned}
(a+bX) - \mathrm{E}(a+bX) &= b(X-\mu) && (\mathrm{E}(a+bX) = a+b\mu) \\
\mathrm{Var}(a+bX) &= \mathrm{E}\big[b^2(X-\mu)^2\big] && (\text{分散の定義}) \\
&= b^2\mathrm{E}\big[(X-\mu)^2\big] = b^2\mathrm{Var}(X) && (b^2 \text{ を外へ})
\end{aligned}
$$

問2

$$
\begin{aligned}
\mathrm{E}(Y^2 \mid X=0) &= 400(0.4)+900(0.4)+1600(0.2) = 840 \\
\mathrm{Var}(Y \mid X=0) &= 840 - 28^2 = 56 \\
\mathrm{E}(Y^2 \mid X=1) &= 400(0.1)+900(0.4)+1600(0.5) = 1200 \\
\mathrm{Var}(Y \mid X=1) &= 1200 - 34^2 = 44
\end{aligned}
$$

2つは異なります。x の値によって誤差のばらつき $\mathrm{Var}(u \mid x) = \mathrm{Var}(Y \mid x)$ が異なることを意味し、これを**不均一分散**（heteroscedasticity）といいます。不均一分散があるときは、不均一分散に頑健な標準誤差を使う必要があります。

問3　u は、$X=0$ のとき $Y-28$ で $(-8,2,12)$、$X=1$ のとき $Y-34$ で $(-14,-4,6)$ をとります。

$$\begin{aligned}\mathrm{E}(u) &= -8(0.20)+2(0.20)+12(0.10)\\ &\quad -14(0.05)-4(0.20)+6(0.25)\\ &= -1.6+0.4+1.2-0.7-0.8+1.5=0\\ \mathrm{E}(Xu) &= -14(0.05)-4(0.20)+6(0.25)=-0.7-0.8+1.5=0\end{aligned}$$

$\mathrm{E}(Xu)$ では、$X=0$ の行は $Xu=0$ なので、$X=1$ の行だけが残ります。

問4　x^2 は x の関数なので、条件付き期待値の外に出せます。

$$\begin{aligned}\mathrm{E}(x^2u) &= \mathrm{E}\big[\mathrm{E}(x^2u\mid x)\big] && \text{(LIE)}\\ &= \mathrm{E}\big[x^2\,\mathrm{E}(u\mid x)\big]=\mathrm{E}[x^2\cdot 0]=0 && \text{(x^2 を外に出し、$\mathrm{E}(u\mid x)=0$)}\end{aligned}$$

一般に、$\mathrm{E}(u\mid x)=0$ なら、x のどんな関数 $g(x)$ についても $\mathrm{E}[g(x)u]=0$ です。平均独立が無相関（直交）より強い条件であるのは、このためです。

問5　$\mathrm{Var}(\bar{Y})=59/100=0.59$ なので、標準偏差は $\sqrt{0.59}\approx 0.77$（万円）です。

問6　正しくありません。第8章の反例2（$x\in\{-1,0,1\}$、$u=x^2-\frac{2}{3}$）では $\mathrm{E}(xu)=0$ ですが、$\mathrm{E}(u\mid x)=x^2-\frac{2}{3}\neq 0$ です。問4から、$\mathrm{E}(u\mid x)=0$ なら $\mathrm{E}(x^2u)=0$ のはずですが、この例では

$$\mathrm{E}(x^2u)=\mathrm{E}(x^4)-\tfrac{2}{3}\mathrm{E}(x^2)=\tfrac{2}{3}-\tfrac{4}{9}=\tfrac{2}{9}\neq 0$$

となっており、$\mathrm{E}(u\mid x)=0$ が成り立たないことが確かめられます。

問7　公式 4 で Y を $-Y$ に置き換えます。

$$\begin{aligned}\mathrm{Var}(X-Y) &= \mathrm{Var}(X)+\mathrm{Var}(-Y)+2\mathrm{Cov}(X,-Y) && \text{（公式4で $Y\to -Y$）}\\ &= \mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y)\end{aligned}$$

最後の等号は、$\mathrm{Var}(-Y)=\mathrm{Var}(Y)$ と、公式5より $\mathrm{Cov}(X,-Y)=-\mathrm{Cov}(X,Y)$ であることによります。 数値例では

$$\mathrm{Var}(Y-20X)=\mathrm{Var}(Y)+400\mathrm{Var}(X)-2\cdot 20\,\mathrm{Cov}(X,Y)=59+100-60=99.$$

問8　標準化すると $\Pr(Z \leq 1.96) = \Phi(1.96) \approx 0.975$ です。

要点

- 記号の読み方、特に $\mathrm{E}(y \mid x)$（条件付き期待値）、ハットとバー（推定量と標本平均）、$:=$（定義）を確認しましょう。
- 計算では、期待値の線形性、分散・共分散の計算ルール、LIE が最もよく使われます。
- 「独立 $\Rightarrow$ 平均独立 $\Rightarrow$ 無相関」の向きと、逆が成り立たない反例を説明できることが理解の目安です。

次の教材へ　次は線形代数の基礎（ベクトルと行列、転置、行列の積、逆行列、ランク、正定値、ベクトルでの微分、分散共分散行列）を学びます。最終的な目標は、$\hat{\beta} = (X^\top X)^{-1} X^\top Y$ という式が読めるようになることです。

大学院とのつながり　この教材の期待値・分散・共分散・条件付き期待値は、すべて次の教材以降で行列とベクトルの形に拡張されます。