

『AI・データサイエンスのための数学入門』学習補助資料

統計的推定を具体例で理解する

点推定・推定量・不偏性・標準誤差

1 何を知りたくて、何を観測するのか

ある大学の学生全員の「1日の学習時間の平均」を知りたいとします。全員を調べる代わりに、一部の学生を選び、その平均から全体の平均を推測します。ここでは、調査時点と学習時間の定義をそろえておきます。

用語	この調査での意味
母集団	対象となる大学の学生全員
標本	選ばれた学生について得られた学習時間のデータ
母数	全学生の平均学習時間 μ 、分散 σ^2 など
統計量	標本平均、標本分散など、標本から計算する量

大文字は「これから変わりうる量」、小文字は「得られた値」

	観測前：確率変数	観測後：実現した値
標本	$X_1, \dots, X_n$	$x_1, \dots, x_n$
標本平均	$\bar{X} = \frac{1}{n} \sum X_i$	$\bar{x} = \frac{1}{n} \sum x_i$
役割	母平均の 推定量	母平均の 推定値

推定量は「どんな標本が出るかによって値が変わる計算規則」です。推定値は、その規則に実際のデータを入れて得た数です。例えば 12, 15, 18, 21 が得られれば、 $\bar{x} = 16.5$ となります。

このノートでの約束

特に断らない限り、 $X_1, \dots, X_n$ は独立同分布で、 $E[X_i] = \mu$ 、 $V[X_i] = \sigma^2 < \infty$ とします。不偏分散では $n \geq 2$ とします。母数は固定して考え、標本の側に確率的な変動をもたせます。

読み方：2～4 節で具体例と図、5 節で証明、6～7 節で解釈と抽出方法を学びます。8 節に問題、9～10 節に解答があります。

2 すべての場合を書き出してみる

0 と 2 がそれぞれ確率 $1/2$ で出る母集団を考えます。例えば、0 と 2 の札を同数入れた箱から、1 枚引くたびに札を戻して、よく混ぜて引きます。このとき

$$\mu = \frac{0+2}{2} = 1, \quad \sigma^2 = \frac{(0-1)^2 + (2-1)^2}{2} = 1.$$

2 回独立に引くと、順序を区別した標本は次の 4 通りです。どれも確率 $1/4$ です。ここでは s_2^2 が分母 2 の標本分散、 s^2 が分母 1 の不偏分散です。

標本 (x_1, x_2)	確率	$\bar{x}$	s_2^2	s^2
(0, 0)	1/4	0	0	0
(0, 2)	1/4	1	1	2
(2, 0)	1/4	1	1	2
(2, 2)	1/4	2	0	0
確率で重み付けした平均		1	1/2	1

不偏とは「繰り返したときの中心が正しい」こと

$$E[\bar{X}] = \frac{0+1+1+2}{4} = 1 = \mu, \quad E[S^2] = \frac{0+2+2+0}{4} = 1 = \sigma^2.$$

一方、 $E[S_2^2] = 1/2 < \sigma^2$ です。ここで大文字の S_2^2, S^2 は、表の各列を確率変数として見たものです。

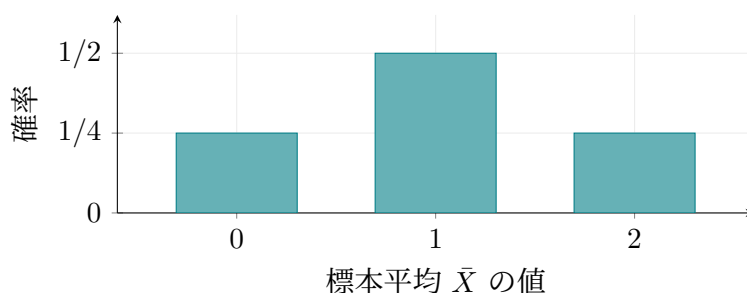


図 1：標本平均の分布。母平均は 1 でも、標本平均が 0 や 2 になることがあります。

立ち止まって考える

この例では、不偏分散の推定値は 0 か 2 で、真の母分散 1 には一度も一致しません。それでも不偏です。「不偏」と「1 回の推定が正確」は違うことを、表から説明してみましょう。

ミニ実験：この抽出を 20 組繰り返し、20 個の標本平均をさらに平均してみましょう。結果を他の人と比べてください。有限回の実験の平均が、期待値 1 にぴったり一致するとは限りません。

3 標準偏差と標準誤差は、何が違うか

散らばっている対象が違う

母標準偏差 σ ：個々の観測値 X_i の散らばり。

標準誤差 $\sigma/\sqrt{n}$ ：標本を取り直したときの標本平均 $\bar{X}$ の散らばり。

独立なデータを平均すると、上側へのずれと下側へのずれが打ち消し合うことがあり、平均のばらつきは小さくなります。正確には

$$E[\bar{X}] = \mu, \quad V[\bar{X}] = \frac{\sigma^2}{n}, \quad SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

標本数を 4 倍にすると、標準誤差は半分

母平均 50、母標準偏差 10 の場合を考えます。

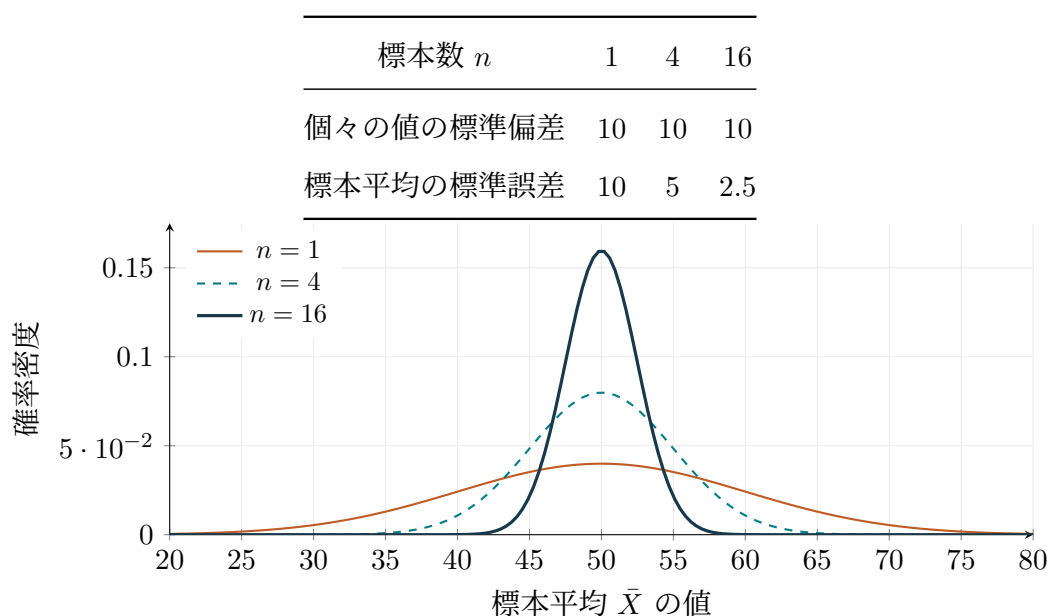


図 2：母集団が正規分布 $N(50, 10^2)$ の場合の、標本平均の厳密な分布。縦軸は確率そのものではなく密度で、各曲線の下全体の面積は 1 です。

母集団が正規分布でなくても、上の期待値・分散の公式は成り立ちます。一方、分布の形を釣鐘型で描けるとは限りません。大きい n で正規分布に近づくという説明には、中心極限定理を使います。**手元のデータから**： σ が未知なら $s/\sqrt{n}$ を標準誤差の推定値として使います。これ自体は、実際の誤差 $\bar{x} - \mu$ ではありません。

4 4 個のデータで分散を計算する

本文のデータ 12, 15, 18, 21 を使います。まず

$$\bar{x} = \frac{12 + 15 + 18 + 21}{4} = 16.5.$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
12	-4.5	20.25
15	-1.5	2.25
18	1.5	2.25
21	4.5	20.25
合計	0	45

したがって、標本分散と不偏分散は、それぞれ

$$s_4^2 = \frac{45}{4} = 11.25, \quad s^2 = \frac{45}{3} = 15.$$

この標本による母平均の推定値は 16.5、母分散の不偏推定量 S^2 の実現値は 15 です。母分散そのものが 15 と分かったわけではありません。

なぜ標本平均を中心にとると、小さくなりやすいのか

任意の数 a に対して、次の恒等式が成り立ちます。

$$\sum_{i=1}^n (x_i - a)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - a)^2$$

右辺の第 2 項は 0 以上なので、偏差平方和は $a = \bar{x}$ のとき最小になります。標本平均は、その標本に最もよく合う中心なのです。

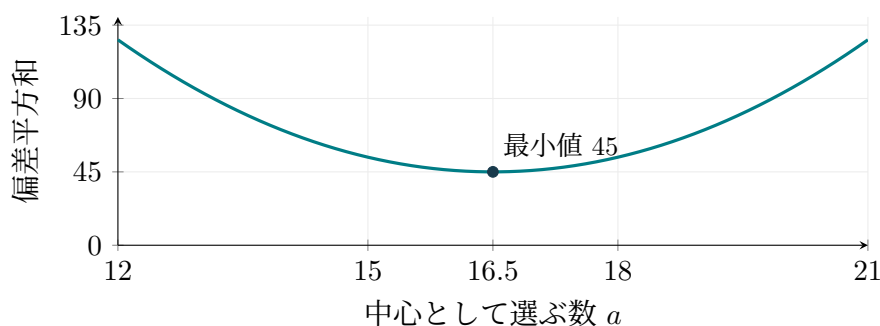


図 3：このデータでは $\sum (x_i - a)^2 = 45 + 4(a - 16.5)^2$ 。最小化による差を平均的に補正するのが、次節の $n - 1$ です。

5 分母が $n - 1$ になる理由

自由度から見る：最後の偏差は、すでに決まっている

4 個の偏差を $d_i = x_i - \bar{x}$ と書くと、 $d_1 + d_2 + d_3 + d_4 = 0$ です。最初の 3 個を決めると、最後は

$$d_4 = -(d_1 + d_2 + d_3)$$

で決まります。一般に偏差には 1 本の線形な制約があるので、自由度は $n - 1$ です。ここでいう「自由」は、確率変数の独立性ではなく、線形代数的に自由に指定できる成分の数を表します。

期待値から見る：補正係数を正確に求める

自由度の説明に加えて、なぜこの補正で不偏になるのかを計算しましょう。前節の恒等式で中心を真の母平均 μ にすると

$$Q := \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

$E[(X_i - \mu)^2] = \sigma^2$ と $E[(\bar{X} - \mu)^2] = \sigma^2/n$ から

$$\begin{aligned} E[Q] &= n\sigma^2 - n \cdot \frac{\sigma^2}{n} \\ &= (n - 1)\sigma^2. \end{aligned}$$

よって、次の 2 つが得られます。

標本分散の偏りと、不偏分散

$$\begin{aligned} S_n^2 &= \frac{Q}{n}, & E[S_n^2] &= \frac{n-1}{n}\sigma^2, \\ S^2 &= \frac{Q}{n-1}, & E[S^2] &= \sigma^2. \end{aligned}$$

不偏性には正規分布の仮定は不要です。独立同分布で、有限の分散をもつことが、ここで用いた条件です。

もし母平均が既知なら？

中心を推定せず、既知の μ を使えば

$$T = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2, \quad E[T] = \sigma^2.$$

この場合は分母 n のままで不偏です。 $n - 1$ は「分散ならいつでも使う数」ではなく、同じデータから平均も推定したことに対応する補正です。

6 不偏なら、よい推定量なのか

不偏でも、ばらつきが大きければ使いにくい

母平均を推定するために、次の 2 つの推定量を考えます。

$$T_1 = X_1, \quad T_2 = \bar{X}.$$

どちらも期待値は μ なので不偏ですが、分散は

$$V[T_1] = \sigma^2, \quad V[T_2] = \frac{\sigma^2}{n}.$$

$n > 1$ 、 $\sigma^2 > 0$ なら、全データの平均を使う T_2 の方がばらつきは小さくなります。不偏性は中心の正しさを表す性質で、推定の精度をそれだけで決めるものではありません。

発展：偏りとばらつきをまとめて測る

母数 θ の推定量 T に対して、平均二乗誤差を

$$\text{MSE}(T) = E[(T - \theta)^2]$$

と定めます。 T の 2 次モーメントが有限なら

$$\text{MSE}(T) = V[T] + (E[T] - \theta)^2$$

です。 $T - \theta = (T - E[T]) + (E[T] - \theta)$ と分けて二乗し、期待値を取ると、交差項の期待値が 0 になります。

小さな例で「不偏が常に最良ではない」と分かる

2 節の例では $\sigma^2 = 1$ です。不偏分散 S^2 は 0 と 2 をそれぞれ確率 1/2 で取るので

$$\text{MSE}(S^2) = \frac{1}{2}(0 - 1)^2 + \frac{1}{2}(2 - 1)^2 = 1.$$

一方、分母 2 の標本分散 S_2^2 は 0 と 1 を半々で取るため

$$\text{MSE}(S_2^2) = \frac{1}{2}(0 - 1)^2 + \frac{1}{2}(1 - 1)^2 = \frac{1}{2}.$$

この例では、偏りのある推定量の方が平均二乗誤差は小さくなります。これだけで、すべての母集団でもそうだとは言えません。

不偏性は、平方根を取るとそのまま残らない

$S = \sqrt{S^2}$ としても、一般には $E[S] = \sigma$ とはなりません。2 節の例なら

$$E[S] = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \sqrt{2} = \frac{1}{\sqrt{2}} < 1 = \sigma.$$

$s/\sqrt{n}$ は標準誤差の推定値として使いますが、不偏な推定値を与える規則だとは一般にいえません。

7 データ数だけでなく、集め方を確かめる

標本が大きくても、選択の偏りは消えない

ある大学では、学生の半数が A 群、半数が B 群で、学習時間の群平均がそれぞれ 1 時間、3 時間だとします。大学全体の平均は

$$\mu = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 3 = 2 \text{ 時間.}$$

しかし、回答しやすさの違いにより、得られる回答が A 群から 90%、B 群から 10% となる仕組みなら、回答者の分布の平均は

$$0.9 \cdot 1 + 0.1 \cdot 3 = 1.2 \text{ 時間.}$$

各群の中では偏りなく回答が得られるものとしします。この回答者の分布から独立に回答を大量に集めても、標本平均は全体の平均 2 ではなく、回答者の平均 1.2 付近に落ち着きます。

小さな標準誤差は、偏りが無いことを意味しない

データ数を増やすと偶然のばらつきは小さくできますが、選択バイアスや測定のずれが自動的に直るわけではありません。AI の学習・評価データでも、実際に対象とする集団をカバーしているかが重要です。

無作為抽出と独立同分布を区別する

- **復元抽出**：毎回戻して同じ条件で独立に選べば、独立同分布のモデルになります。
- **非復元抽出**：一度選んだ個体を戻さない場合、後の選択は前の選択の影響を受けます。無作為でも、標本は一般に独立ではありません。
- **実際の調査**：層別抽出など、異なる抽出確率を用いる適切な設計もあります。抽出設計に合った推定や重み付けが必要です。

無作為抽出は、1 回ごとの標本が必ず母集団と同じ構成になることを保証するものではありません。確率に基づいて誤差を評価するための土台です。

発展：有限母集団から戻さずに抽出するとき

$N > 1$ 個の値 $a_1, \dots, a_N$ に対し、 $\mu = N^{-1} \sum a_j$ 、 $\sigma_N^2 = N^{-1} \sum (a_j - \mu)^2$ とします。大きさ n の各部分集合を等確率で選ぶ単純無作為非復元抽出では

$$V[\bar{X}] = \frac{\sigma_N^2}{n} \frac{N-n}{N-1}.$$

$n = N$ なら分散は 0 です。 n が N に比べて十分小さい場合には、独立モデルの式 σ_N^2/n に近くなります。

8 練習問題

問 1 | 推定量と推定値 観測前の $\bar{X}$ 、観測後の $\bar{x} = 16.5$ 、母平均 μ を、「母数」「推定量」「推定値」に分類しなさい。また、 $\bar{x} = 16.5$ から $\mu = 16.5$ と断定できない理由を説明しなさい。

問 2 | 計算を通して区別する 12, 15, 18, 21 について、(a) 標本平均、(b) 偏差平方和、(c) 分母 n の標本分散、(d) 不偏分散の値、(e) 標準誤差の推定値を求めなさい。(f) 自由度が 3 となる理由を説明しなさい。

問 3 | 全列挙による確認 2 節の母集団から 2 回独立に復元抽出します。(a) $V[\bar{X}]$ を分布表から求め、 σ^2/n と一致することを確認しなさい。(b) 不偏分散が真の母分散と一致する確率を求めなさい。(c) それでも不偏と呼べる理由を説明しなさい。

問 4 | 標本数と標準誤差 母標準偏差が 12 のとき、(a) $n = 36, 144$ に対する標本平均の標準誤差を求めなさい。(b) 標準誤差を 1 以下にするための最小の標本数を求めなさい。(c) これは実際の誤差の絶対値が必ず 1 以下となることを意味しますか。

問 5 | 中心が既知の場合 母平均が $\mu = 10$ と分かっている母集団から、標本 8, 10, 15 が得られました。(a) 標本平均を使う不偏分散の値、(b) 既知の母平均を使う $n^{-1} \sum (x_i - 10)^2$ を求めなさい。(c) どちらも不偏な推定量の値であるのに、答えが異なる理由を説明しなさい。

問 6 | 正しいか、誤りか 理由も述べなさい。

- (a) 不偏推定量の値は、毎回、真の母数に近い。
- (b) 無作為抽出なら、標本は必ず独立同分布である。
- (c) 標本数を 4 倍にすると、母集団の標準偏差も半分になる。
- (d) $E[S^2] = \sigma^2$ なら、 $E[S] = \sigma$ である。

問 7 | AI の評価データ 全利用者の満足度を知るため、毎日利用する人だけにアンケートを送り、10 万件の回答を集めました。(a) 回答数が多くても残る問題を説明しなさい。(b) 調査設計の改善を 2 つ提案しなさい。

問 8 | 発展：不偏性と精度 $0 \leq a \leq 1$ とし、母平均の推定量 $T_a = aX_1 + (1-a)X_2$ を考えます。 X_1, X_2 は独立同分布で、分散は $\sigma^2 > 0$ です。(a) すべての a で不偏であることを示しなさい。(b) 分散を最小にする a を求めなさい。

授業・自習での使い方

基本確認には問 1~4、誤解の確認には問 6~7 を使えます。問 5・8 は、公式の背景をもう一歩理解するための問題です。次ページから解答です。

9 解答と解説：問 1～5

問 1

$\bar{X}$ は推定量、 $\bar{x} = 16.5$ は推定値、 μ は母数です。標本平均には標本抽出によるばらつきがあるので、1 回の推定値が真の値と一致するとは限りません。

問 2

(a) $\bar{x} = 16.5$ 。(b) 偏差平方和は $20.25 + 2.25 + 2.25 + 20.25 = 45$ 。(c) $s_4^2 = 45/4 = 11.25$ 。(d) $s^2 = 45/3 = 15$ 。(e) $s = \sqrt{15}$ なので、標準誤差の推定値は

$$\widehat{\text{SE}}(\bar{X}) = \frac{s}{\sqrt{4}} = \frac{\sqrt{15}}{2} \simeq 1.94.$$

(f) 偏差の和が 0 なので、3 個を決めれば残る 1 個が決まります。自由度は $4 - 1 = 3$ です。なお、偏差の和が 0 という事実だけで不偏性が証明されたわけではなく、厳密な期待値の計算は 5 節の通りです。

問 3

(a) 母平均は 1 なので

$$V[\bar{X}] = \frac{1}{4}(0-1)^2 + \frac{1}{2}(1-1)^2 + \frac{1}{4}(2-1)^2 = \frac{1}{2} = \frac{\sigma^2}{n}.$$

(b) S^2 は 0 か 2 で、 $\sigma^2 = 1$ には一致しないので確率は 0 です。(c) $E[S^2] = (0 + 2 + 2 + 0)/4 = 1$ だからです。不偏性は、可能な標本全体について確率で重み付けした平均の性質です。

問 4

(a) $12/\sqrt{36} = 2$ 、 $12/\sqrt{144} = 1$ 。(b) $12/\sqrt{n} \leq 1$ は $n \geq 144$ と同値なので、最小は 144 です。(c) 意味しません。標準誤差は標本平均の分布の広がりであり、誤差の最大値ではありません。例えば正規母集団では、平均から 1 標準誤差を超えて離れる確率は約 32% あります。

問 5

(a) $\bar{x} = 11$ なので

$$s^2 = \frac{(8-11)^2 + (10-11)^2 + (15-11)^2}{2} = \frac{26}{2} = 13.$$

(b) $\{(8-10)^2 + (10-10)^2 + (15-10)^2\}/3 = 29/3 \simeq 9.67$ 。(c) 不偏性は繰り返し抽出に関する期待値の性質です。異なる推定量が、同じ標本に対して同じ値を取ることは要求しません。

10 解答と解説：問 6～8

問 6

- (a) 誤り。不偏でも大きく外れることがあります。不偏性に加え、分散などでばらつきも評価します。
- (b) 誤り。有限母集団からの非復元抽出は反例です。前に誰を選んだかによって、次に選べる個体が変わります。
- (c) 誤り。同じ母集団なら σ は変わりません。半分になるのは、独立同分布モデルでの標本平均の標準誤差 $\sigma/\sqrt{n}$ です。
- (d) 誤り。2 節の例では、 $E[S^2] = 1$ ですが $E[S] = 1/\sqrt{2} \neq 1$ です。

問 7

(a) 日常的に利用する人だけでは、利用頻度の低い人の意見を取りこぼします。さらに、アンケートへの回答者と非回答者で満足度が異なる可能性もあります。10 万件あっても、全利用者を代表する保証はありません。

(b) 例えば、次のように改善できます。

- 対象期間と「全利用者」の範囲を定め、その名簿から無作為に抽出する。利用頻度別に層を作って抽出する方法も考えられる。
- 回答しやすい方法や再案内を用意し、非回答の状況を確認する。抽出確率や層の構成比が異なる場合は、それに合った重み付けを検討する。

重み付けだけで、観測していない要因による非回答の偏りまで必ず解消できるわけではありません。

問 8

(a) 期待値の線形性により $E[T_a] = a\mu + (1-a)\mu = \mu$ 。(b) 独立性により

$$\begin{aligned} V[T_a] &= \{a^2 + (1-a)^2\}\sigma^2 \\ &= \left\{2\left(a - \frac{1}{2}\right)^2 + \frac{1}{2}\right\}\sigma^2. \end{aligned}$$

よって $a = 1/2$ で最小となり、最小値は $\sigma^2/2$ です。このとき $T_a = (X_1 + X_2)/2$ 、すなわち標本平均になります。

最後に、自分の言葉で説明してみる

- $\bar{X}$ と $\bar{x}$ は、どのように違うか。
- なぜ不偏であっても、推定が外れることがあるのか。
- σ と $\sigma/\sqrt{n}$ は、それぞれ何の散らばりか。
- 分母 $n-1$ は、何を補正しているのか。
- データを増やすだけでは直らない問題には、何があるか。