

5. 推定と検定

母集団分布の母数を推定する方法と仮説検定の方法を解説する。まず、母数を一つの値で推定する点推定について、推定精度としての標準誤差を説明する。また、母数が区間に存在することを推定する信頼区間も取り扱う。後半は統計的仮説検定について述べる。検定法の基本的な考え方と正規分布および二項確率についての検定法を解説する。

5.1. 点推定

先に述べた統計量は対応する母数の推定値である。このように母数を一つの値およびベクトルで推定する場合を点推定(point estimation)という。母数の推定値は次のように定義される。

定義 5.1. 標本 X_1, X_2, \dots, X_n に対して、統計量 $t(X_1, X_2, \dots, X_n)$ で母数 θ を推定するとき、 $t(X_1, X_2, \dots, X_n)$ を母数 θ の推定量(estimator)といい、その値を推定値(estimate)と呼ぶ。□
推定量は抽出する標本毎に偶然に変動する確率変数であり、その推定精度を考える必要がある。 $t(X_1, X_2, \dots, X_n)$ を母数 θ の推定量とすると、確率 1 で

$$t(X_1, X_2, \dots, X_n) \neq \theta$$

である。標本平均と標本分散はそれぞれ母集団(分布)の平均と分散の推定値である。正規分布 $N(100, 25)$ から、標本数 10 と 100 の無作為抽出実験を 10 回行い、標本平均と分散を求めた結果を下の表にまとめた。

表 5.1. 標本数 10 の標本からの標本平均と分散

実験	1	2	3	4	5	6	7	8	9	10
平均	99.69	99.24	103.00	98.85	100.77	99.10	99.90	99.74	100.30	100.36
分散	37.54	40.20	29.18	19.03	30.30	36.82	41.60	14.41	36.15	15.23

表 5.2. 標本数 100 の標本からの標本平均と分散

実験	1	2	3	4	5	6	7	8	9	10
平均	99.49	100.34	100.27	100.03	99.98	99.42	100.11	99.99	100.48	99.83
分散	29.47	28.67	28.29	25.02	28.91	25.84	24.86	21.13	23.83	28.96

標本平均や標本分散は標本を抽出するごとに偶然変動するが、変動の大きさは標本が増えるに従って小さくなる。この性質は前章で述べた大数の法則である。推定量の良さの基準として次の 2 つの性質は基本的である。

(i) 一致性(consistency)

$t(X_1, X_2, \dots, X_n)$ を母数 θ の推定量とすると、 $n \rightarrow \infty$ に対して

$$t(X_1, X_2, \dots, X_n) \rightarrow \theta \quad (\text{確率収束})$$

であるとき、推定量は一致性をもつという。また、その推定量は θ の一致推定量 (consistent estimator) という。

(ii) 不偏性 (unbiasedness)

推定量 $t(X_1, X_2, \dots, X_n)$ が

$$E\{t(X_1, X_2, \dots, X_n)\} = \theta$$

のとき、この性質を推定量の不偏性という。また、その推定量は不偏推定量 (unbiased estimator) である。

標本平均 \bar{X} は母平均 μ の推定量であり、大数の法則から一致性を持つことが分かる。また、 $E(\bar{X}) = \mu$ であるから、標本平均は不偏推定量でもある。次に偏差平方和

$$\sum_{i=1}^N (X_i - \bar{X})^2 = \sum_{i=1}^N X_i^2 - n\bar{X}^2$$

を考える。ここで、

$$E(X_i^2) = \sigma^2 + \mu^2, \quad E(\bar{X}^2) = \sigma^2/n + \mu^2$$

であるから、この期待値は $(n-1)\sigma^2$ である。従って

$$E(S^2) = (n-1)\sigma^2/n, \quad E(U^2) = \sigma^2$$

を得る。このことから、 S^2 は σ^2 の不偏推定量では無いが、 U^2 は不偏推定量である。両者が一致推定量であることは大数の法則を用いれば分かる。標本数が大きいとき S^2 は不偏推定量に近づき、このような性質を漸近不偏性という。このように、我々が通常用いている統計量は解釈が容易で、かつ数学的にも良好な性質をもっている。

例 5.1. 確率変数 X と Y の標本分散をそれぞれ U_{XX} , U_{YY} で、また共分散を U_{XY} とすれば、標本相関係数は

$$r = \frac{U_{XY}}{\sqrt{U_{XX}} \sqrt{U_{YY}}}$$

で示される。この統計量は不偏性を持たないが、 U_{XX} , U_{YY} , U_{XY} が一致推定量であるので、相関係数の一致推定量である。標本数が大きいときは、標本相関係数は

$$r = \frac{U_{XY}}{\sigma_X \sigma_Y}$$

で近似されるので、漸近的に不偏推定量である。

5.2. 推定量の標準誤差

母数 θ の不偏推定量を $t(X_1, X_2, \dots, X_n)$ とするとき、 $t(X_1, X_2, \dots, X_n)$ の標準偏差を標準誤差 (standard error, SE) という。すなわち、

$$SE = \sqrt{\text{Var}(t(X_1, X_2, \dots, X_n))}$$

である。例えば、標本平均 \bar{X} の標準誤差は $SE = \sigma / \sqrt{n}$ で、また不偏分散 U^2 の分散が $\text{Var}(U^2) = \frac{2}{n-1} \sigma^4$ であるから標準誤差は $SE = \sqrt{\frac{2}{n-1}} \sigma^2$ である。実際の解析では標準誤差はその推定値で

求めることになる。すなわち、前者は $\hat{SE} = \frac{U}{\sqrt{n}}$ 、後者は $\hat{SE} = \sqrt{\frac{2}{n-1}} U^2$ である。

例 5.2. 次のデータは男子学生の身長データのデータである。母集団が正規分布するものとして、データから平均と分散を推定する。

データ : 168.0, 168.5, 172.5, 182.5, 164.0, 160.5, 167.5, 174.5, 162.5, 173.0, 168.5, 173.5
164.5, 177.5, 160.0, 172.0, 168.0, 170.5, 166.5, 168.5, 175.0, 157.0, 171.0, 176.5 (cm)

このデータから

標本平均 $\bar{x} = 169.3$ (cm) ($\hat{SE} = 1.2$), 標本分散 $U^2 = 36.3$ (cm²) ($\hat{SE} = 10.7$)

を得る。

円周率 π の推定実験を行う。一辺が 2 の正方形内から無作為に点を n 個抽出し、そのうち内接する半径 1 の円に入る点の個数 n_0 の相対度数の 4 倍、すなわち $\frac{4n_0}{n}$ で円周率を推定する

(モンテカルロ法)。図 5.1 の実験で推定値 3.15 ($SE = 0.0818$) を得る。図 5.2 はこの実験を 10 回反復した結果である。推定値は標本抽出を行う度に変動する様子が示されている。

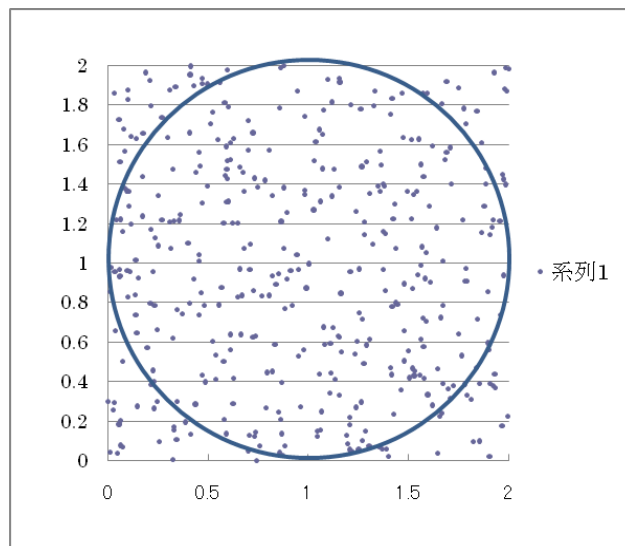


図 5.1 $n=400$ の場合の実験

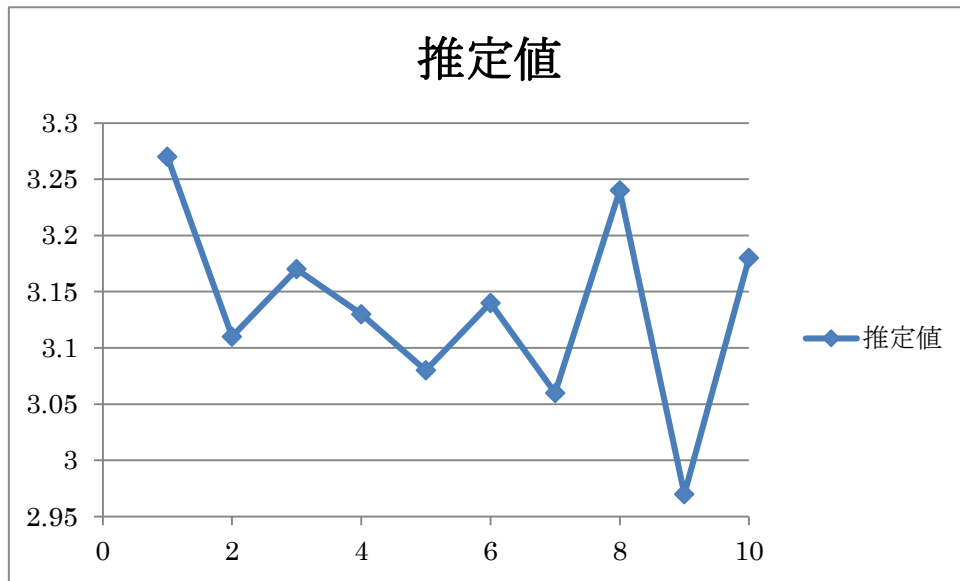


図 5.2 $n=400$ の実験を 10 回反復した結果

標準誤差は推定量の推定精度を表すので、推定精度を予め決定して、その精度に合わせた標本数を決定することが可能である。

例 5.3. 男子学生の身長標準偏差を高々 $\sigma=6.5\text{cm}$ とするとき、平均の推定で標準誤差を 0.5cm 以下にするための標本数を求める。標本数を n とするとき

$$SE = \sigma / \sqrt{n} = 6.5 / \sqrt{n} \leq 0.5$$

を解いて、 $n \geq 169$ を得る。

例 5.4. 成功確率 p のベルヌイ試行で、 p の推定を考える。ただし、 $0.2 < p < 0.8$ とする。実験回数 n としたときの標本平均の標準誤差は

$$SE = \sqrt{p(1-p)/n} \leq 1/(2\sqrt{n})$$

である。標準誤差を 0.05 以下にするための標本数は

$$1/(2\sqrt{n}) \leq 0.05$$

を解いて、 $n \geq 100$ を得る。

一般に標準誤差は漸近的な方法によって算出する場合が多い。その理論は次節で説明する推定量の漸近正規性に基づく。

問 5.1. 次のデータは男子学生 20 名に対する胸囲のデータ (単位 cm) である。平均とその標準誤差を求めよ。

81.0, 93.6, 85.0, 86.1, 95.3, 91.1, 79.4, 84.4, 84.0, 87.7,

85.6, 90.5, 82.0, 85.3, 91.9, 87.7, 85.5, 81.1, 87.1, 978

5.3. 信頼区間

母数の点推定に対して、母数が区間に入ることを推定する。この推定法を区間推定(interval estimation)という。

定義 5.2. 母数 θ をもつ分布 (母集団) からの無作為標本を x_1, x_2, \dots, x_n とするとき、統計量 $T_L(x_1, x_2, \dots, x_n), T_U(x_1, x_2, \dots, x_n)$ が

$$T_L(x_1, x_2, \dots, x_n) < T_U(x_1, x_2, \dots, x_n),$$

$$P(T_L(x_1, x_2, \dots, x_n) < \theta < T_U(x_1, x_2, \dots, x_n) | \theta) = 1 - \alpha$$

を満たすとき、区間 $(T_L(x_1, x_2, \dots, x_n), T_U(x_1, x_2, \dots, x_n))$ を信頼係数 $1 - \alpha$ の信頼区間(confidence interval)、または $100(1 - \alpha)\%$ 信頼区間という。

注意 定義上で信頼区間の信頼係数は自由に決めることができるが、通常は $\alpha = 0.05, 0.01$ すなわち 95% と 99% の信頼区間が用いられる。

信頼区間の意味を考えるために、正規分布 $N(10, 25)$ から標本の大きさ 10 の標本抽出を 20 回反復し、各標本で平均の 95% 信頼区間を作ったものを表 5.8 にしている。この実験では 20 回中 19 回の信頼区間は平均値を含むことが分かる。第 11 回目の標本で信頼区間は平均値を含まない。平均の 95% 信頼区間は繰り返し実験すると、確率 0.95 で母平均を含むことを意味する。

表 5.3. 95% 信頼区間の 20 回の反復実験

標本	1	2	3	4	5	6	7	8	9	10
下限	9.274	6.102	6.257	6.571	4.661	4.279	7.856	9.28	10.126	6.767
上限	15.47	12.3	12.45	12.77	10.86	10.48	14.05	15.48	16.324	12.97
標本	11	12	13	14	15	16	17	18	19	20
下限	3.623	8.427	6.95	7.696	7.519	6.543	6.672	6.933	7.2665	7.627
上限	9.821	14.62	13.15	13.89	13.72	12.74	12.87	13.13	13.465	13.82

(i) 平均値の信頼区間

正規母集団 $N(\mu, \sigma^2)$ からの無作為標本を X_1, X_2, \dots, X_n とするとき、 t 統計量

$$t = \sqrt{n} (\bar{X} - \mu) / U$$

は自由度 $n-1$ の t 分布に従う。いま、 $t(n-1, \alpha/2)$ をこの t 分布の上側 $100\alpha/2\%$ 点とすれば

$$P\left(-t(n-1, \frac{\alpha}{2}) < \frac{\sqrt{n}(\bar{X} - \mu)}{U} < t(n-1, \frac{\alpha}{2})\right) = 1 - \alpha$$

である。この確率内の事象から、 $100(1-\alpha)\%$ 信頼区間は

$$\bar{X} - t(n-1, \frac{\alpha}{2}) \frac{U}{\sqrt{n}} < \mu < \bar{X} + t(n-1, \frac{\alpha}{2}) \frac{U}{\sqrt{n}} \quad (5.1)$$

で与えられる。信頼区間は次のように表現する場合もある。

$$\bar{X} \pm t(n-1, \frac{\alpha}{2}) \frac{U}{\sqrt{n}}$$

例 5.4. 男子学生 15 人を無作為抽出し、収縮期の血圧を測定し、 $\bar{x} = 123.47$ (mmHg), $U^2=153.124$ を得た。このときの 95%信頼区間を求める。血圧の分布が正規分布とすると、 $t(14,0.05)=2.145$ であるから信頼区間は(116.62,130.32)で与えられる。

問 5.2. 例 5.2 の身長データのデータから平均の 95%信頼区間を作れ。

(ii) 成功 (二項) 確率の信頼区間

成功確率 p のベルヌイ試行結果として、標本 X_1, X_2, \dots, X_n を得るとき、 p の最尤推定量 \hat{p} は標本平均であり、その漸近分布は $N(p, p(1-p)/n)$ である。このとき、

$$P(|\hat{p} - p| < q(\alpha/2) \sqrt{\hat{p}(1-\hat{p})/n}) = 1 - \alpha$$

である。ここで、分散 $p(1-p)/n$ を $\hat{p}(1-\hat{p})/n$ で近似して、次の $100(1-\alpha)\%$ 信頼区間を得る。

$$\hat{p} - q(\alpha/2) \sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + q(\alpha/2) \sqrt{\hat{p}(1-\hat{p})/n} \quad (5.2)$$

ここに、 $q(\alpha/2)$ は標準正規分布の上側確率が $\alpha/2$ である限界値、すなわち $P(Z > q(\alpha/2)) = \alpha/2$ を満たす値である。

例 5.5. 182 回の実験を行い、55 回成功した。このとき、成功確率の 95%信頼区間を求める。

$$\hat{p} = 55/182 = 0.302$$

であるから、次の信頼区間を得る。

$$0.151 < p < 0.453$$

次に n 回のベルヌイ試行を行って、1 回も成功しなかった場合の信頼区間を求める。いま、

$$(1-p)^n > 1-\alpha$$

を考えて、これを p について解いて、次の信頼区間を得る。

$$0 \leq p < 1-(1-\alpha)^{1/n}$$

問 5.5. 100 人をある母集団から任意に 100 人抽出し、55 人がある項目に「はい」と答えた。「はい」と答える確率の 95%信頼区間を求めよ。

5.4. 検定の考え方

仮説検定は、仮説（帰無仮説） H_0 と対立仮説 H_1 の二者択一を標本を用いて行う統計的方法である。その概要を掴むために次の例を考える。

[例 5.6] あるサイコロが正確かどうかを調べることを考える。表 6.1 はサイコロを 100 回投げた実験結果である。この場合、反復して実験を行い、もし特定の目が他より多く観察され、または少なく観察されれば、そのサイコロの正確さについて疑いをもつ。この検定の帰無仮説と対立仮説は次のようにするのが自然である。

H_0 : サイコロは正確である。すなわち

$$P(i \text{ の目が出る}) = 1/6 \quad (i=1,2,3,4,5,6).$$

H_1 : サイコロは不正確である。すなわち、少なくとも 1 つの目に対して

$$P(i \text{ の目が出る}) \neq 1/6.$$

この例で、サイコロを n 回反復し、 i の目が出る回数を n_i とするとき、観測結果に偏りがあると確率的に判断されれば、帰無仮説を棄却し(reject)、そのような偏りの傾向が見られなければ、帰無仮説を採択する(accept)ことになる。この例はカイ自乗検定の節で再度取り扱う。

表 5.4. サイコロ投げ実験

サイコロの目	1	2	3	4	5	6	計
出現回数	10	15	20	24	21	10	100

□

また、次の例は通常見られる典型的な例である。

[例 5.7] 過去数年間、同じ内容の統計学試験を行った結果、その平均は 70 点で、標準偏差は 6.5 点であった。今年は 20 人の学生が、統計学の試験を受け、その標本平均は 75 点であった。このクラスの学生は、過去の学生に対して成績が良いか否かを考える。この場合の仮説は

$$H_0: \text{平均 } \mu = 70, \quad H_1: \mu \neq 70 \quad (5.3)$$

である。また、今年は試験前に、成績向上を狙って特別補習を行っているとするれば、対立仮説は

$$H_1: \mu > 70 \quad (5.4)$$

とするのが妥当である。 □

上の例から分かるように、統計的検定では仮説と対立仮説を事前に立てることが重要で、そのためには対象とする現象を十分吟味する必要がある。検定で二者択一の判断基準として、データ情報を縮約する統計量を用いる。例 5.6 では後に述べるカイ自乗統計量(chi square statistic)、例 5.7 では標本平均、あるいは t 統計量(t -statistic)を用いて検定を行う。検定に用いる統計量を検定統計量(test statistic)という。仮説検定の理解を助けるために、例 5.7 を考える。

学生が均一で、試験成績が正規母集団 $N(\mu, \sigma^2)$ に従うとする。標本平均を \bar{X} とするとき、

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \quad (5.5)$$

の分布は標準正規分布である。この場合、仮説 $H_0: \mu = 70$ の下で $\sigma = 6.5, n = 20$, 観測値 $\bar{X} = 75$ である。ただし、過去の受験学生が多数のために、 σ を未知母数とせず、既知母数 $\sigma = 6.5$ として考える。この場合の統計量(5.5)の値は

$$z_0 = \sqrt{20} (75 - 70) / 6.5 = 3.440$$

であり、

$$P(Z > z_0) = P(Z > 3.440) = 0.0003$$

を得る。これは仮説 H_0 の下で極めてまれな観測結果であると考えられ、仮説は棄却することが妥当である。上の確率を有意確率または p 値という。仮説検定の定式化については次の節で述べる。

問 5.3. 例 5.6 で 1 の目は 10 回出現している。帰無仮説 H_0 の下で、1 の目が 10 以下になる確率を中心極限定理により近似計算し、この結果が稀なものかどうか検討せよ。

問 5.7. 例 5.7 で分散が未知とする。標本平均 $\bar{x} = 75$ 、不偏分散が $u^2 = 51$ であった。t 統計量を用いて仮説 $H_0: \mu = 70$ を検討せよ。

5.5. 正規分布に関する基本的検定法

(i) 正規分布 $N(\mu, \sigma^2)$, 分散 σ^2 が既知の場合の $H_0: \mu = \mu_0$ の検定

標本 X_1, X_2, \dots, X_n が得られたとき、統計量 $Z(5.4)$ は H_0 の下で標準正規分布に従う。対立仮説 $H_1: \mu \neq \mu_0$ のとき、有意水準 α に対して棄却域は

$$W: |Z| > z(\alpha/2) \quad (5.6)$$

で与えられる。ここに、棄却限界点(critical point) $z(\alpha/2)$ は $P(|Z| > z(\alpha/2)) = \alpha$ を満たす上側 $50\alpha\%$ 点である。従って、有意水準 $\alpha = 0.05$ に対して棄却域は $|Z| > 1.96$ である。また、有意水準 $\alpha = 0.01$ に対しては、 $|Z| > 2.58$ が棄却域である。対立仮説 $H_1: \mu > \mu_0$ のときは

$$W: Z > z(\alpha) \quad (5.7)$$

が棄却域である。同様に対立仮説 $H_1: \mu < \mu_0$ のときは

$$W: Z < -z(\alpha) \quad (5.8)$$

が棄却域になる。

問 5.8. 例 5.7 を有意水準 0.05 で検定せよ。

大標本のときの近似的な検定は、この方法が基本である。標本 X_1, X_2, \dots, X_n が平均 μ 、分散 σ^2 の分布(母集団)から得られたとき、標本数が十分大きいとき、標本平均 \bar{X} は平均 μ 、分散 σ^2/n の正規分布に漸近的に従う(中心極限定理)。従って

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \quad (5.9)$$

の分布は漸近的に標準正規分布である。また、分散 σ^2 が未知のときは、不偏分散を U^2 に対し

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{U} \quad (5.10)$$

は近似的に標準正規分布に従い、このことを用いれば、平均に関する検定が行える。上の不偏分散を標本分散 S^2 で置き換えても近似的には同様である。

(ii) 正規分布 $N(\mu, \sigma^2)$, 分散 σ^2 が未知の場合の $H_0: \mu = \mu_0$ の検定

このときの検定統計量は

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{U} \quad (5.11)$$

であり、自由度 $n-1$ の t 分布に従う。有意水準 α のとき、対立仮説 $H_1: \mu \neq \mu_0$ に対する棄却域は

$$W: |t| > t(n-1, \alpha/2) \quad (5.12)$$

である。ここで $t(n-1, \alpha/2)$ は自由度 $n-1$ の t 分布の上側 $100\alpha/2\%$ 点である。対立仮説 $H_1: \mu > \mu_0$ のときは、棄却域は

$$W: t > t(n-1, \alpha)$$

であり、また対立仮説 $H_1: \mu < \mu_0$ のときは、

$$W: t < -t(n-1, \alpha)$$

となる。

[例 5.8] マウス 10 匹に特殊な栄養剤を混ぜた飼料を与えて飼育したところ、生後 1 ヶ月の体重増加が次のデータのようになった。

19, 18, 21, 22, 24, 15, 23, 20, 18, 28 (g)

マウスを通常の飼料のみで飼育した場合の発育は、1 ヶ月で 20g とされている。このとき、栄養剤の効果があつたかどうかを検定する。体重増加量は正規分布とすると仮定し、仮説 $H_0: \mu = 20$ と対立仮説 $H_1: \mu > 20$ に対して検定を行う。検定統計量は t 統計量であり、

$$t_0 = \sqrt{10} (20.8 - 20) / 3.676 = 0.688$$

を得る。 t 統計量の上側 5% 点は $t(19, 0.05) = 1.729$ で、帰無仮説は棄却されない。栄養剤の効果は無いと判断される。

問 5.4. 例 5.7 のデータが次のように与えられている。 H_0 : 平均 $\mu = 70$, H_1 : $\mu \neq 70$ を有意水準 0.05 で検定せよ。ただし、データは正規分布すると仮定する。

95 63 90 35 88 60 73 55 81 100
93 88 67 90 75 72 70 50 43 75