

1 解説

信頼区間と仮説検定について解説する。

1.1 信頼区間

推定の問題では、期待値 μ の値を標本平均 \bar{X}_n で不偏に推定することができる。しかし、推定値として 1 つの値 \bar{X}_n が得られるよりも、どのあたりに本当の期待値 μ がありそうかという範囲を提示するほうが便利な場合もある。このための方法が信頼区間である。

1.2 仮説検定

仮説検定とは「仮説がデータから支持されるかどうかを判断するための統計的手法」である。仮説検定は科学的な推論を必要とするときによく使われる。とくに、因果関係が複雑で原理的な法則から結果を予測することが困難な状況 (医学・薬学分野または心理実験などの人文科学の領域) では、データから仮説の正否を判断をする上で仮説検定は必須の方法である。仮説検定の方法は、信頼区間を構成する方法と類似点がある。

2 パーセント点

信頼区間や仮説検定のために必要になる「パーセント点」を解説する。

標準正規分布のパーセント点： $X \sim N(0, 1)$ のとき

$$P(X \leq z_\alpha) = 1 - \alpha \text{ を満たす点 } z_\alpha \longrightarrow \text{標準正規分布の 上側 } \alpha \text{ 点 (100}\alpha\% \text{ 点)}$$

$$P(|X| \leq z_\alpha^*) = 1 - \alpha \text{ を満たす点 } z_\alpha^* \longrightarrow \text{標準正規分布の 両側 } \alpha \text{ 点 (100}\alpha\% \text{ 点)}$$

正規分布以外に t -分布のパーセント点も必要にあることがある。

t -分布： 確率変数 X, Y_1, \dots, Y_n はすべて独立に標準正規分布 $N(0, 1)$ にしたがうとする。このとき確率変数 T_n を

$$T_n = \sqrt{n} \cdot \frac{X}{\sqrt{Y_1^2 + \dots + Y_n^2}} = \frac{X}{\sqrt{\frac{Y_1^2 + \dots + Y_n^2}{n}}}$$

と定める。 T_n がしたがう分布のことを 自由度 n の t -分布 とよぶ。確率変数 T が自由度 n の t -分布にしたがうとき

$$T \sim t_n$$

と書く¹。

t -分布のパーセント点： $X \sim t_n$ (自由度 n の t -分布) のとき

$$P(X \leq t_{n,\alpha}) = 1 - \alpha \text{ を満たす点 } t_{n,\alpha} \longrightarrow \text{自由度 } n \text{ の } t\text{-分布の 上側 } \alpha \text{ 点 (100}\alpha\% \text{ 点)}$$

$$P(|X| \leq t_{n,\alpha}^*) = 1 - \alpha \text{ を満たす点 } t_{n,\alpha}^* \longrightarrow \text{自由度 } n \text{ の } t\text{-分布の 両側 } \alpha \text{ 点 (100}\alpha\% \text{ 点)}.$$

¹ t -分布の確率密度関数 $f(x; n)$ は $f(x; n) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$, $x \in \mathbb{R}$ で与えられる。ここで $\Gamma(k)$ はガンマ関数 ($\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$) である。 $n \rightarrow \infty$ のとき $f(x; n)$ は標準正規分布の密度関数に収束する。

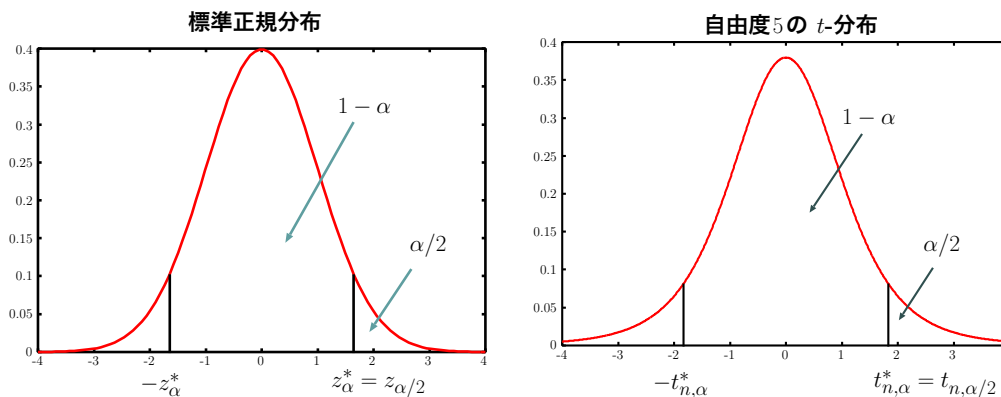


図 1: 標準正規分布と t -分布のパーセント点. t -分布の自由度が大きいつき, ほとんど標準正規分布と同じ.

標準正規分布, t -分布の密度関数は原点に関して対称なので

$$z_{\alpha}^* = z_{\alpha/2}, \quad t_{n,\alpha}^* = t_{n,\alpha/2}$$

が成り立つ (図 2 参照). 正規分布と t -分布のパーセント点を以下に示す. 自由度 n が大きいつきには z_{α}^* と $t_{n,\alpha}^*$ はほぼ同じ値をとる.

表 1: 正規分布と t -分布のパーセント点.

| α | $t_{10,\alpha}^*$ | $t_{20,\alpha}^*$ | $t_{100,\alpha}^*$ | z_{α}^* |
|----------|-------------------|-------------------|--------------------|----------------|
| 0.01 | 3.169 | 2.845 | 2.626 | 2.576 |
| 0.05 | 2.228 | 2.086 | 1.984 | 1.960 |
| 0.1 | 1.812 | 1.725 | 1.660 | 1.645 |

3 信頼区間

データが独立に同一の分布にしたがうとする:

$$X_1, \dots, X_n \sim_{i.i.d.} f(x; \theta)$$

このとき, パラメータ θ を次のように区間で推定する:

「 $1 - \alpha$ 以上の確率で $\theta \in [T_1, T_2]$ が成り立つ」

ここで $[T_1, T_2]$ は閉区間である. また α は通常 0.01 や 0.05 などの小さい値であり, データを解析する人が, どの程度の信頼性を確保したいかを考慮して決める. 上のような区間 $[T_1, T_2]$ を, 信頼水準 $100(1 - \alpha)\%$ の信頼区間という. 以下ではとくにデータが正規分布にしたがう場合に, 期待値の信頼区間を構成する方法について説明する.

3.1 正規分布のもとでの信頼区間の構成法 – 分散が既知のとき –

正規分布 $N(\mu, \sigma^2)$ から独立な n 個のデータが得られているとする:

$$X_1, \dots, X_n \sim_{i.i.d.} N(\mu, \sigma^2)$$

ここで分散 σ^2 は既知であると仮定する (つまり信頼区間を作るときに σ^2 の値を用いてよい). このとき期待値 μ の信頼区間を構成する. 標本平均を $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ とすると $\bar{X}_n \sim N(\mu, \sigma^2/n)$ なので

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim N(0, 1)$$

が成り立つ. したがって, 標準正規分布の両側パーセント点 z_α^* を用いると

$$\begin{aligned} P\left\{ \left| \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \right| \leq z_\alpha^* \right\} &= 1 - \alpha \\ \iff P\left\{ \bar{X}_n - z_\alpha^* \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X}_n + z_\alpha^* \sqrt{\frac{\sigma^2}{n}} \right\} &= 1 - \alpha \end{aligned}$$

が成り立つ. 分散 σ^2 が既知のとき, 期待値 μ の信頼水準 $100(1 - \alpha)\%$ の信頼区間は

$$\left[\bar{X}_n - z_\alpha^* \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_\alpha^* \sqrt{\frac{\sigma^2}{n}} \right]$$

となる.

例 1 (データ数と信頼区間の幅の関係). $X_1, \dots, X_n \sim_{i.i.d.} N(\mu, 1)$ のとき, μ について信頼水準 95% の信頼区間を構成する. 図 2 のグラフは観測数がそれぞれ 10 個のときと 30 個のときのプロットである. データ数が多いほうが信頼区間の幅は狭くなっていることが見てとれる. \square

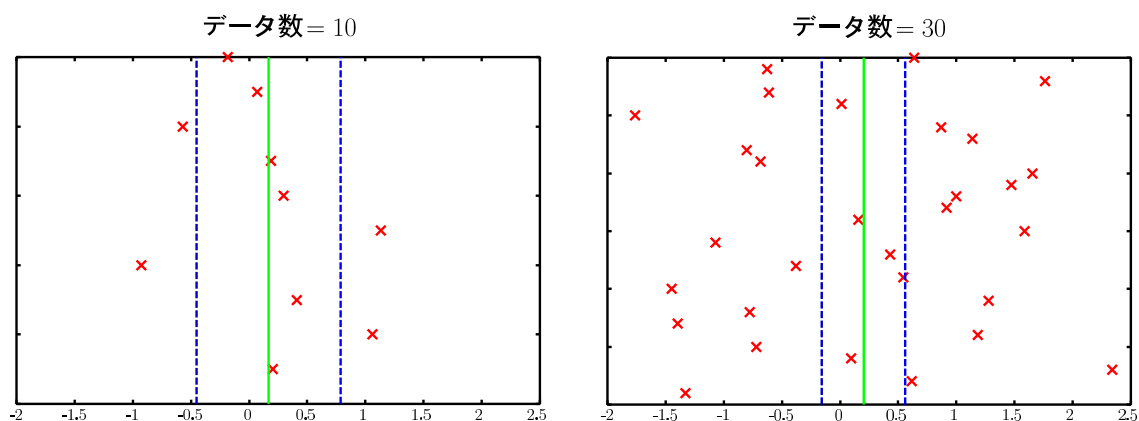


図 2: 観測値と 95% 信頼区間のプロット

3.2 正規分布のもとでの信頼区間の構成法 – 分散が未知のとき –

前節と同様に

$$X_1, \dots, X_n \sim_{i.i.d.} N(\mu, \sigma^2)$$

とする. 期待値 μ と分散 σ^2 がともに未知のときに μ の信頼区間を構成する. 期待値の分散の不偏推定量として

$$\text{標本平均} : \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad \text{標本分散} : S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

を考える. このとき

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S_n^2}} \sim t_{n-1} \quad (\text{自由度 } n-1 \text{ の } t\text{-分布})$$

が成り立つ (詳細は省略. 統計の教科書などを参照のこと). したがって μ の $100(1-\alpha)\%$ 信頼区間は次のように構成できる:

$$P\left\{\left|\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{S^2}}\right| \leq t_{n-1, \alpha}^*\right\} = 1 - \alpha$$

$$\iff P\left\{\bar{X}_n - t_{n-1, \alpha}^* \sqrt{\frac{S_n^2}{n}} \leq \mu \leq \bar{X}_n + t_{n-1, \alpha}^* \sqrt{\frac{S_n^2}{n}}\right\} = 1 - \alpha.$$

以上より, 分散 σ^2 が未知のとき, μ の信頼水準 $100(1-\alpha)\%$ の信頼区間は

$$\left[\bar{X}_n - t_{n-1, \alpha}^* \sqrt{\frac{S_n^2}{n}}, \bar{X}_n + t_{n-1, \alpha}^* \sqrt{\frac{S_n^2}{n}}\right]$$

で与えられる.

正規分布のもとでの期待値の信頼区間の構成 (まとめ)

$X_1, \dots, X_n \sim i.i.d. N(\mu, \sigma^2)$ とする. σ^2 が既知か未知かに応じて, 期待値 μ の信頼区間はそれぞれ以下ようになる.

$$\sigma^2 \text{が既知のとき} \implies P\left(\mu \in \left[\bar{X}_n - z_{\alpha}^* \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{\alpha}^* \sqrt{\frac{\sigma^2}{n}}\right]\right) = 1 - \alpha$$

$$\sigma^2 \text{が未知のとき} \implies P\left(\mu \in \left[\bar{X}_n - t_{n-1, \alpha}^* \sqrt{\frac{S^2}{n}}, \bar{X}_n + t_{n-1, \alpha}^* \sqrt{\frac{S^2}{n}}\right]\right) = 1 - \alpha$$

3.3 区間推定のその他の例

区間 $[0, \theta]$ 上の一様分布から n 個データ X_1, X_2, \dots, X_n が独立に得られている. このとき θ の信頼区間を構成する. θ の不偏推定量は

$$T = \frac{n+1}{n} \max_{1 \leq i \leq n} X_i$$

なので, これを利用する. T の分布関数 $P(T \leq t)$ を計算すると,

$$P(T \leq t) = \left(\frac{n}{\theta(n+1)}\right)^n t^n, \quad 0 \leq t \leq \frac{n+1}{n}\theta$$

となる. ここで $u \in \mathbb{R}$ を

$$P(u \leq T \leq \frac{n+1}{n}\theta) = 1 - \alpha$$

となるように定めると

$$u = (1 - \alpha)^{1/n} \frac{\theta(n+1)}{n}$$

となる. したがって

$$P\left(\left(1 - \alpha\right)^{1/n} \frac{\theta(n+1)}{n} \leq T \leq \frac{n+1}{n}\theta\right) = 1 - \alpha$$

$$\iff P\left(\frac{n}{n+1}T \leq \theta \leq (1 - \alpha)^{-1/n} \frac{n}{n+1}T\right) = 1 - \alpha$$

となるので, 閉区間

$$\left[\frac{n}{n+1}T, (1 - \alpha)^{-1/n} \frac{n}{n+1}T\right]$$

は θ の $100(1-\alpha)\%$ 信頼区間になっている.

4 仮説検定

仮説検定とは、ある仮説が正しいかどうかをデータから検証するための統計的方法である。

- 応用例

- 実験結果から理論が正しいかどうかの検証：メンデルの実験など
- 薬の認可：薬が効くか効かないか、副作用が重いか軽いかなど

- 用語の準備

- 帰無仮説：検証したい仮説 (H_0 と書く)
- 対立仮説： H_0 が棄却されたときに採択される仮説 (H_1 と書く)

例：コイン投げ、確率 p で表が出る。仮説を

$$H_0 : p = 1/2, \quad H_1 : p \neq 1/2$$

とすれば、コインが公平かどうかの検定をすることになる。また仮説を

$$H_0 : p = 1/2, \quad H_1 : p > 1/2$$

と設定すると、 $p \geq 1/2$ という事前知識がある場合に、表が出やすいかどうかの検定をすることになる。

4.1 仮説検定の考え方

帰無仮説 H_0 のもとで観測データ X_1, \dots, X_n が得られる確率が非常に小さいなら、 H_0 は正しくないと判断して H_1 を採択する。

有意水準：データが出現しやすいかどうかを判断する基準として、有意水準 α を設定する。この α は通常、0.05 や 0.01 などの小さな値にしておく。

以上の考え方にしたがって、仮説検定の手順を以下に示す。

1. 帰無仮説 H_0 と対立仮説 H_1 を定める。
2. 有意水準 α を定める。
3. データ X_1, \dots, X_n を観測する。
4. 帰無仮説 H_0 が正しいと仮定したとき、観測データが得られる確率を計算する。
 - (a) 確率が有意水準 α 以下 $\implies H_0$ は間違いと判断して H_1 を採択。
 - (b) 確率が有意水準 α より大きい $\implies H_0$ は正しいと判断。

帰無仮説 H_0 が正しくないと判断するような観測データの集合を 棄却域 という。つまり

$$H_0 \text{ を棄却} \iff \text{観測データ } (X_1, \dots, X_n) \text{ が棄却域に属する。}$$

以下に例を示す。

例 2. $X_1, \dots, X_n \sim_{i.i.d.} N(\mu, \sigma^2)$ とする。分散は既知として、期待値 μ に関する以下の検定を行う：

$$H_0 : \mu = 1, \quad H_1 : \mu \neq 1$$

有意水準を $\alpha = 0.05$ とする. 期待値の推定量 \bar{X}_n に対して $\bar{X}_n \sim N(\mu, \sigma^2/n)$ となるので

$$\sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} \sim N(0, 1)$$

となる. したがって, もし $\mu = 1$ が正しいと仮定すると

$$P\left(\left|\sqrt{n} \cdot \frac{\bar{X}_n - 1}{\sigma}\right| \geq z_{0.05}^*\right) = 0.05 \iff P\left(|\bar{X}_n - 1| \geq z_{0.05}^* \sqrt{\frac{\sigma^2}{n}}\right) = 0.05$$

となる. 棄却域 W を

$$W = \left\{ (X_1, \dots, X_n) \mid |\bar{X}_n - 1| \geq z_{0.05}^* \sqrt{\frac{\sigma^2}{n}} \right\}$$

と定めて, 観測データが W に入れば帰無仮説 H_0 を棄却することにする. つまり $|\bar{X}_n - 1|$ が大きな値をなら仮説 $H_0: \mu = 1$ は正しくないと判断する.

もし σ^2 が未知なら, σ^2 を標本分散 S_n^2 で推定して t -分布のパーセント点を用いれば, 棄却域を作ることができる.

4.2 仮説検定の誤り確率

いくつかの検定方式があるときに, それらをどのように比較したらよいだろうか? たまたま偏ったデータが観測されたときには, 検定による判断は誤ってしまうこともある. 誤って判断してしまう確率が小さいほど優れた検定方式と考えることができる. 検定の棄却域を W とすると検定における誤りは以下の表のようにまとめることができる.

| | データに基づく判断 | |
|------------|-----------------------------------|-----------------------------|
| | データ $\notin W$ (H_0 を棄却しない) | データ $\in W$ (H_0 を棄却) |
| True H_0 | ○ | × (第 I 種の誤り) |
| H_1 | × (第 II 種の誤り) | ○ |

- ・ 帰無仮説 H_0 が正しいとき: もしデータが棄却域に入ると, 本来は H_0 が正しいにもかかわらず H_0 を棄却してしまうことになり, 誤った判断をすることになる. その確率は

$$\text{第 I 種の誤り} = P(H_0 \text{ が正しいときに } H_0 \text{ を棄却})$$

となる. 検定の棄却域は

$$\text{第 I 種の誤り} \leq \alpha \quad (\text{有意水準})$$

となるように構成されている.

- ・ 対立仮説 H_1 が正しいとき: もしデータが棄却域に入らないと, 帰無仮説 H_0 が誤りであるにもかかわらず H_0 を棄却しないという誤った判断をしてしまう. その確率は

$$\text{第 II 種の誤り} = P(H_1 \text{ が正しいとき } H_0 \text{ を棄却しない})$$

となる.

検定には以上の 2 通りの誤りが存在する. 仮説検定では, 有意水準 α を定めることで第 I 種の誤りを α 以下に抑えているが, 第 II 種の誤りにはとくに制約を置いていない. いくつかの検定方式を比較するときには次のように考える.

有意水準 α の検定方法が複数あるとき:
第 II 種の誤り確率が小さい検定方式 のほうが優れている

4.3 最適な検定

第 II 種の誤り確率を最も小さくするような、最適な検定方法はどのようなものか？ 仮説が簡単な場合には、答えは次の補題で与えられる。

補題 1 (ネイマン・ピアソンの補題). データ X が得られたときに、データの確率密度関数が $p_0(x)$ か $p_1(x)$ かを検定する：

$$\begin{cases} H_0 : X \sim p_0(x) & (\text{帰無仮説}) \\ H_1 : X \sim p_1(x) & (\text{対立仮説}) \end{cases}$$

有意水準 α の棄却域のうち、第 II 種の誤り確率を最も小さくする棄却域は

$$W = \left\{ x \mid \frac{p_1(x)}{p_0(x)} \geq c \right\}$$

で与えられる。ここで c は

$$\text{第 I 種の誤り確率} = \int_W p_0(x) dx = \alpha$$

を満たす値とする。 □

上で定義した $\lambda(x) = \frac{p_1(x)}{p_0(x)}$ を尤度比という。直感的には

$p_0(x)$: 仮説 H_0 のもとでのデータ x の出現しやすさ

$p_1(x)$: 仮説 H_1 のもとでのデータ x の出現しやすさ

を表す。尤度比 $\lambda(x)$ の値が大きいきには分布 p_1 からデータが出現したと判断するのが妥当であり、 H_0 を棄却することになる (図 3)。

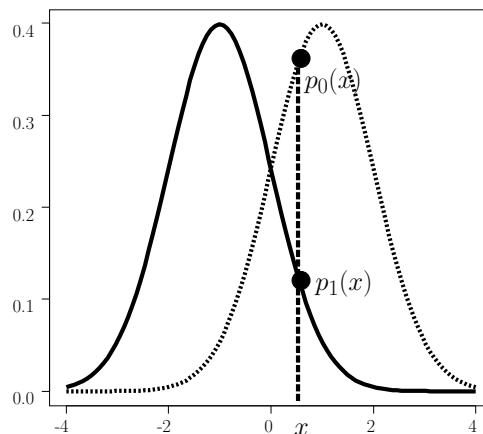


図 3: 確率密度関数 $p_0(x), p_1(x)$. 尤度比は $\lambda(x) = p_1(x)/p_0(x)$ で定義される。

例 3 (正規分布の期待値の検定). データ X_1, \dots, X_n が正規分布 $N(\mu, 1)$ から独立に得られているとする。分散の値は既知として期待値に関する次の検定を行う。

$$H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1.$$

ここで $\mu_0 < \mu_1$ とする。ネイマン・ピアソンの補題から定まる棄却域を構成する。確率密度関数で考えると上の仮説は

$$\begin{aligned} H_0 : p_0(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(x_i - \mu_0)^2 / 2} \\ H_1 : p_1(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-(x_i - \mu_1)^2 / 2} \end{aligned}$$

となる。尤度比を計算すると、

$$\lambda = \frac{p_1(x_1, \dots, x_n)}{p_0(x_1, \dots, x_n)} = \exp \left(\sum_{i=1}^n x_i(\mu_1 - \mu_0) + \frac{n(\mu_0^2 - \mu_1^2)}{2} \right)$$

となる。棄却域は

$$\begin{aligned} W &= \{(X_1, \dots, X_n) \mid \lambda(X_1, \dots, X_n) > c\} \\ &= \{\bar{X}_n \mid \bar{X}_n \cdot (\mu_1 - \mu_0) > c'\} \\ &= \{\bar{X}_n \mid \bar{X}_n > c''\} \quad (\mu_1 > \mu_0 \text{ を使って式変形}) \end{aligned}$$

と書ける。ここで c', c'' は適当な定数である。式変形していくと c' や c'' は μ_0, μ_1 に依存した複雑な式になるが、そのような複雑なことは考える必要はない。棄却域を表現する式を簡略化した後、有意水準が α になるように c'' を決めればよい。帰無仮説 H_0 のもとでは $\bar{X}_n \sim N(\mu_0, 1/n)$ より $\sqrt{n}(\bar{X}_n - \mu_0) \sim N(0, 1)$ となるので

$$P(\bar{X}_n \in W) = P(\bar{X}_n \geq c'') = P(\sqrt{n}(\bar{X}_n - \mu_0) \geq \sqrt{n}(c'' - \mu_0)) = \alpha$$

となるように c'' を定めればよいので、 $\sqrt{n}(c'' - \mu_0) = z_\alpha$ (標準正規分布の上側パーセント点) とおけばよい。したがって棄却域を

$$W = \left\{ \bar{X}_n \mid \bar{X}_n \geq \mu_0 + \frac{z_\alpha}{\sqrt{n}} \right\}$$

とすれば、有意水準 α のもとで第 II 種の誤り確率が最も小さくなる。 □

演習問題

1. 正規分布から以下のような 6 個のデータが得られた。

$$-0.4, 1.1, 1.9, 3.4, 2.1, 0.7.$$

このとき正規分布の期待値の 99% 信頼区間を計算せよ。必要となるパーセント点の値は、統計の本などで調べること。

2. 確率変数 X, Y はともに以下のような正規分布に独立にしたがうとする：

$$X \sim N(\mu, \sigma_1^2), \quad Y \sim N(\mu, \sigma_2^2).$$

このとき分散 σ_1^2, σ_2^2 は既知とする。また $Z = \frac{X+Y}{2}$ とする。区間 $[Z - c, Z + c]$ によって μ の $100(1 - \alpha)\%$ 信頼区間を構成するとき c の値を求めよ。

3. 区間 $[0, \theta]$ 上の一様分布から n 個データ X_1, X_2, \dots, X_n が独立に得られている。このとき θ の信頼区間を構成する。 $Y = \max_{1 \leq i \leq n} X_i$ とする。 θ の $100(1 - \alpha)\%$ 信頼区間として $Y \leq \theta \leq Y(1 + c)$ を考える。 c の値を求めよ。

4. $X_1, \dots, X_n \sim_{i.i.d.} N(\mu, 1)$ として、期待値 μ について以下の検定を行う：

$$\text{帰無仮説 } H_0 : \mu = 0, \quad \text{対立仮説 } H_1 : \mu = 1$$

このとき分散が 1 に等しいことは既知とする。

- (a) 有意水準が α の棄却域で、第 II 種の誤り確率が最小となるものを構成せよ。
 (b) 標準正規分布の分布関数を $F(x)$ として、(a) で求めた棄却域の第 II 種の誤り確率を F を用いて表せ。

5. n 個の確率変数 $X_1, X_2, \dots, X_n \sim_{i.i.d.} N(0, \sigma^2)$ とする. このとき分散について以下の検定をおこなう.

$$H_0: \sigma^2 = \sigma_0^2, \quad H_1: \sigma^2 = \sigma_1^2.$$

ただし $\sigma_0^2 < \sigma_1^2$ とする. 第 II 種の誤り確率を最小にするような棄却域 W は, 適当な定数 c を用いて

$$W = \{(X_1, \dots, X_n) \mid \sum_{i=1}^n X_i^2 \geq c\}$$

と表されることを示せ.

6. n 個の確率変数 X_1, X_2, \dots, X_n が独立に正規分布 $N(\mu, 1)$ にしたがうとする. 分散は既知とする. 期待値 μ について以下の検定

$$H_0: \mu = \mu_0 \quad H_1: \mu = \mu_0 + \frac{\varepsilon}{\sqrt{n}}$$

をおこなう. ここで $\varepsilon > 0$ とする.

(a) ネイマン・ピアソンの補題にしたがって, 有意水準 α の棄却域を構成せよ.

(b) 第 II 種の誤り確率が γ 以下になるために, ε はどのような条件を満たす必要があるか考察せよ.

7. n 次元確率変数 $X = (X_1, \dots, X_n)$ が n 次元正規分布 $N(\mu, A)$ にしたがっているとする. ここで平均ベクトル μ は

$$\mu = (c, c, \dots, c)^t \in \mathbb{R}^n$$

であり c は未知である. また分散共分散行列 A の ij 要素を a_{ij} ($i, j = 1, \dots, n$) とする. また分散共分散行列 A は既知であるとする. このとき n 次元正規分布にしたがう 1 つの観測値 $X = (X_1, \dots, X_n)$ を用いて c の信頼区間を構成する.

(a) パラメータ c を推定するために

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

を用いる. \bar{X} が c の不偏推定量になっていることを確かめよ. また \bar{X} がしたがう分布を求めよ.

(b) パラメータ c を推定量 \bar{X} で推定したとき, $100(1 - \alpha)\%$ 信頼区間を構成せよ.

8. 以下の問において標準正規分布のパーセント点が必要になるが, これはコンピュータで計算するか, 統計の本で調べよ.

(a) 正規分布 $N(\mu, 80)$ にしたがうデータを 20 個観測したとき, データの平均値は 81.2 であった. このとき 95% 信頼区間を求めよ.

(b) 正規分布 $N(\mu, 9)$ にしたがうデータを n 個観測したとき, データの平均を \bar{X}_n とおく. ほぼ $P(\bar{X}_n - 1 \leq \mu \leq \bar{X}_n + 1) = 0.90$ となるような n を求めよ.

9. データ X_1, X_2, \dots, X_n が独立に同一のポアソン分布 $Po(\lambda)$ にしたがっているとする (ポアソン分布についてはテキストなどを参照) ポアソン分布の期待値は $E[X_i] = \lambda$, 分散は $V[X_i] = \lambda$ である. パラメータ λ を $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ で推定する. このとき

$$\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\lambda}}$$

が $n \rightarrow \infty$ で標準正規分布にしたがうことを利用して, λ の $100(1 - \alpha)\%$ 信頼区間を近似的に構成せよ.