

数理統計学 問題演習

Mathematical Statistics: Problems

本問題集で取り上げている問題は、

尾畑伸明・荒木由布子 共著『Python で学ぶ確率統計』共立出版, 2023 年

の章末問題です。解説動画では、通常の解法に加えて、【Python 利用】とある問題については、Python のコード例を説明しています。

問題 4.1

1 変量データ x_1, x_2, \dots, x_n を標準化したものを z_1, z_2, \dots, z_n とするとき, $\bar{z} = 0$ と $s_z^2 = 1$ を示せ. ただし, 1 変量データ x_1, x_2, \dots, x_n には揺らぎがあるものとする.

問題 4.2

生データの平均値と, 生データから度数分布表を作って得られる度数データから計算した平均値との差は階級幅の $1/2$ を越えないことを示せ.

問題 4.3

平均値, メディアン, モードの間には一般に成り立つ大小関係がないことを具体例で確認せよ.

問題 4.4

2変量のデータ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ は $s_x > 0, s_y > 0$ を満たすものとする.

- (1) すべての i に対して $y_i = ax_i + b$ が成り立つような定数 $a > 0$ と b が存在するとき, 相関係数は $r_{xy} = 1$ となることを証明せよ.
- (2) 逆に, 相関係数が $r_{xy} = 1$ を満たすとき, すべての i に対して $y_i = ax_i + b$ が成り立つような定数 $a > 0$ と b が存在することを示し, a と b を求めよ.

問題 4.5

【Python 利用】次の表はある試験結果(100点満点)の生データである。

34	78	16	36	54	42	75	78	100	41	64	38	73	81	24
22	52	58	38	94	28	43	66	54	14	8	73	18	32	36
16	49	65	37	73	36	100	60	22	58	73	84	77	65	68
66	81	56	75	66	69	28	64	38	46	38	76	55	81	69
54	46	36	87	80	63	49	85	26	48	56				

- (1) データのサイズ, 最大値, 最小値, メディアンを求めよ.
- (2) 平均値と標準偏差を求めよ.
- (3) 適当な階級を設定して, 度数分布表とヒストグラムを示せ.
- (4) 度数分布表をもとにして, 平均値と標準偏差を求めよ.

なお, 上記の数値データ (Problem4.5.csv) は次のサイトから入手可能である:

<https://www.kyoritsu-pub.co.jp/book/b10033631.html>

問題 4.6

【Python 利用】次の表はあるコースの受講者の中間試験と期末試験の結果である。

中間試験	44	68	32	63	82	88	37	51	65	60	45	78	50
期末試験	53	76	56	68	86	79	64	61	71	56	68	64	42

- (1) 散布図を描け。
- (2) 中間試験, 期末試験それぞれの平均値と標準偏差を求めよ。
- (3) 中間試験と期末試験の相関係数を求めよ。

なお, 上記の数値データ (Problem4_6.csv) は次のサイトから入手可能である:

<https://www.kyoritsu-pub.co.jp/book/b10033631.html>

問題 5.1

トランプのカード 52 枚から 5 枚を選ぶとき, 同じ数字のカードが 4 枚含まれる事象を A とし, 同じ数字のカードが丁度 2 枚含まれ, ほかには同じ数字のカードがない事象を B とする. ただし, 数字には J, Q, K も含める. 事象 A, B が起こる確率をそれぞれ求めよ.

(補注) ポーカーの役で言えば, 事象 A はフォーカード, 事象 B はワンペアである. 意欲があれば, ポーカーの役すべてについて確率を求めよ. (実際のポーカーではカードを取り替えるが, ここでは, 配られた状態だけを考える.)

問題 5.2

棒をランダムに折って 2 本の断片に分割するとき、長い方の断片の長さが短いほうの 3 倍以上になる確率を求めよ。

問題 5.3

地図帳で目的地を探すとき, いつも目的地が地図の周辺にあって不便な思いをしないだろうか. 縦 30cm, 横 40cm の長方形の地図で, 目的地が地図の周辺 5cm の範囲に見つかる確率を求めよ.

問題 5.4

(分配法則) 事象 A, B, C に対して, 次が成り立つことを示せ.

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

問題 5.5

(ド・モルガンの法則) 事象 A, B に対して, 次が成り立つことを示せ.

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c.$$

問題 5.6

コイン 3 枚を同時に投げるとき, 表を向いたコインの枚数を X とする. X の分布関数を求め, そのグラフの概形を示せ.

問題 5.7

k を定数とする. 関数

$$f(x) = \begin{cases} \frac{x}{25} + k, & 2 \leq x \leq 7, \\ 0, & \text{その他,} \end{cases}$$

が確率密度関数になるように k の値を定めて, この分布の平均値を求めよ.

問題 5.8

(包除原理) 次の等式を証明せよ.

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) \\ - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C).$$

問題 5.9

事象 E, F が $P(E) = 1, P(F) = 0$ を満たすものとする. このとき, すべての事象 A に対して次式が成り立つことを示せ.

$$P(A \cap E) = P(A \cup F) = P(A).$$

(補注) 確率 1 の事象 E を**殆ど確実な事象**, 確率 0 の事象 F を**殆ど不可能な事象**という. 文脈によっては, E は殆ど確実に起こる (成り立つ), あるいは F は殆ど確実に起こらない (不成立である) などともいう. $P(E) = 1$ だからといって $E = \Omega$ とは限らないし, $P(F) = 0$ だからといって $F = \emptyset$ とは限らないので注意しよう. 現実問題への応用では, 確率 1 の事象は必ず起こり, 確率 0 の事象は決して起こらないと解釈される.

問題 5.10

確率変数 X の平均値を μ , 分散を σ^2 とするとき, $\mathbf{E}[(X - 2)(X + 3)]$ を μ と σ を用いて表せ.

問題 5.11

確率変数 X に対して, $\mathbf{E}[(X - a)^2]$ を最小にする実数 a を求めよ.

問題 5.12

確率変数 X が $E[X^2] = 0$ をみたせば, $E[X] = 0$ であり, $V[X] = 0$ となることを示せ.

問題 5.13

サイコロを 2 個振って出る目のうち大きい方を X , 小さい方を Y とする. ただし, 同じ目が出た場合は $X = Y$ とする. Y の平均値と分散を求めよ.

問題 5.14

長さ L の線分をランダムに 2 分割して得られる 2 つの線分のうち, 長いほうの長さを X , 短いほうを Y とする (丁度 2 等分されたときは $X = Y$ とする). X, Y それぞれの平均値と分散を求めよ.

問題 5.15

確率変数 X が非負の整数 $0, 1, 2, \dots$ にのみ値を取るとき,

$$\mathbf{E}[X] = \sum_{n=1}^{\infty} P(X \geq n)$$

が成り立つことを示せ.

問題 6.1

サイコロを 2 個振って出る目のうち大きい方を X , 小さい方を Y とする. ただし, 同じ目が出た場合は $X = Y$ とする. 次の条件付き確率を求めよ.

$$P(X \leq 4|Y = 2), \quad P(X + Y \geq 8|X \geq 5),$$

問題 6.2

壺の中に a 個の白玉と b 個の黒玉が入っている. この壺から 1 個を取り出し, 同色の玉を c 個付け加えて戻す. 壺の中には $a + b + c$ 個の玉が入っている. そこから再び 1 個を取り出すとき, その玉が白玉である確率を求めよ.

問題 6.3

(乗法定理の一般化) 事象 A, B, C に対して,

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

が成り立つことを示せ.

問題 6.4

3つの事象 A, B, C が $P(A \cap B \cap C) \neq 0$ と $P(C|A \cap B) = P(C|B)$ を満たすとき, $P(A|B \cap C) = P(A|B)$ が成り立つことを示せ.

問題 6.5

2つの壺 U_1, U_2 があり, U_1 には赤玉 4 個, 白玉 2 個, 黒玉 4 個, U_2 には赤玉 5 個, 白玉 3 個, 黒玉 2 個が入っている. まず, U_1 から 1 個の玉を取り出して U_2 に入れ, 次に, U_2 から 1 個の玉を取り出したところ黒玉であった. 初めに U_1 から取り出した玉は何色であったか, 確率で答えよ.

問題 6.6

感染率が 0.2% の地域において, 感染症の検査 T を実施する. ただし, この検査 T は感染者の 80% に陽性反応を示すが, 非感染者の 0.5% にも陽性反応が出てしまう.

- (1) この検査を受けて陽性反応が出た人が感染者である確率を求めよ.
- (2) この検査を受けて陰性反応が出た人が非感染者である確率を求めよ.

問題 6.7

A, B, C が独立で, $P(A) = a, P(B) = b, P(C) = c$ とする. 次の確率を a, b, c を用いて表せ.

$$P(A \cap B^c), \quad P(A \cup B \cup C) \quad P(A \cup (B \cap C)), \quad P(A|B \cup C).$$

問題 6.8

$P(A|B) = P(A|B^c)$ が成り立つための条件を調べよ.

問題 6.9

2つの確率変数 X, Y は独立で, $\mathbf{E}[X] = \mu_1, \mathbf{E}[Y] = \mu_2, \mathbf{V}[X] = \sigma_1^2, \mathbf{V}[Y] = \sigma_2^2$ とするとき, 次の統計量を $\mu_1, \mu_2, \sigma_1, \sigma_2$ を用いて表せ.

$$\mathbf{E}[X + Y], \quad \mathbf{E}[XY], \quad \mathbf{V}[X + Y], \quad \mathbf{V}[XY].$$

問題 6.10

3つの確率変数 X, Y, Z について, X と Z が無相関, Y と Z が無相関であれば, $X + Y$ と Z も無相関であることを示せ.

問題 7.1

箱の中に赤玉 4 個, 白玉 8 個が入っている. この箱から無作為に玉を 1 個ずつ取り出し, 初めて赤玉が出るのに要する回数 (赤玉が取り出された回も含める) を N とする. 次の場合について N の分布を求めて, 数値的に比較せよ.

- (1) 1 回ごとに取り出された玉を箱に戻す (復元抽出).
- (2) 1 回取りだされた玉は箱に戻さない (非復元抽出).

問題 7.2

【Python 利用】成功確率が $p = 0.3$ のベルヌイ試行を繰り返すとき, 初めて成功するまでに要する試行回数 (成功した回も含めて数える) を T とする.

- (1) T の累積確率分布を図示せよ.
- (2) $P(T \geq a) \leq 0.01$ を満たす最小の a を求めよ.

問題 7.3

母関数の方法を用いて, 幾何分布とポアソン分布の平均値と分散を計算せよ.

問題 7.4

【Python 利用】二項分布 $B(100, 0.02)$ とパラメータ $\lambda = 2$ のポアソン分布が近いことを数値的に確認せよ.

問題 7.5

確率変数 X, Y はそれぞれパラメータ λ_1, λ_2 のポアソン分布に従い, 独立であるとする. このとき, $X + Y$ はパラメータ $\lambda_1 + \lambda_2$ のポアソン分布に従うことを示せ.

問題 7.6

確率変数 X が区間 $[0, 5]$ 上の一様分布に従うとき, 条件付き確率 $P(X > 1.5 | X \leq 3.5)$ を求めよ.

問題 7.7

間欠的に起こる現象があり, 現象が 1 回起こった後に次の現象が起こるまでの時間間隔 T は平均値 6 時間の指数分布に従うという. 現象が 1 回起こった後, 次の現象が起こるまでの 12 時間以上空く確率を求めよ.

問題 7.8

(指数分布の無記憶性) X を指数分布に従う確率変数とするとき, 次が成り立つことを示せ:

$$P(X \geq a + b | X \geq a) = P(X \geq b), \quad a \geq 0, \quad b \geq 0.$$

問題 7.9

【Python 利用】 $X \sim N(-2.8, 4.5^2)$ のとき, $P(X \leq a) = 0.85$ を満たす a を求めよ.

問題 7.10

【Python 利用】(偏差値) 受験者数が多数の時, 得点の分布は正規分布に近いと想定されることが多い. その分布の平均値が μ , 分散が σ^2 であるとき, 得点 X を

$$Y = 50 + 10 \cdot \frac{X - \mu}{\sigma}$$

のように変換した Y を偏差値という. 偏差値が 58 以上になる確率と偏差値が 24 以下になる確率を求めよ.

問題 7.11

【Python 利用】自由度 n の t 分布 t_n に従う確率変数 X に対して, X^2 は自由度 $(1, n)$ の F 分布 F_n^1 に従うことを乱数によるシミュレーションによって確かめよ.

問題 8.1

【Python 利用】 区間 $[0, 2]$ 上の一様分布に従う独立同分布な確率変数列 X_1, X_2, \dots に対して, 中心極限定理をシミュレーションで検証せよ.

問題 8.2

【Python 利用】ある会場の収容人数は 1200 名である。イベントの開催に合わせて予約を受け付けるが、これまでの経験から予約客のうち 3% が予約をキャンセルするという。キャンセルを見込んで 1230 名の予約を受け付けたとき、来客を収容しきれなくなる確率を求めよ。ただし、予約客は互いに独立に 3% の確率で予約をキャンセルする。二項分布による厳密計算と二項分布の正規分布近似による結果を比較せよ。

問題 8.3

【Python 利用】確率変数列 X_1, X_2, X_3 は独立であり, いずれも $[0, 2]$ 上の一様分布に従うものとする. このとき, $Y = X_1 + X_2 + X_3$ は次の密度関数 $f(x)$ をもつ確率変数になる:

$$f(x) = \begin{cases} \frac{1}{16} x^2, & 0 \leq x \leq 2, \\ -\frac{1}{8} (x - 3)^2 + \frac{3}{8}, & 2 \leq x \leq 4, \\ \frac{1}{16} (x - 6)^2, & 4 \leq x \leq 6, \\ 0, & \text{その他.} \end{cases}$$

このことをシミュレーションによって数値的に検証せよ.

問題 8.4

【Python 利用】ドモアブル-ラプラスの定理の主張 $B(n, p) \approx N(np, np(1-p))$ では, n を大きくするほどに近似の度合いがよくなる. このことを $B(n, p)$ に従う乱数を大量に収集して, その分布状況を調べることで確認せよ.

問題 8.5

【Python 利用】トランプのカード 52 枚からランダムに 5 枚を選んだ手札の中に含まれるエースの枚数について, その確率分布をシミュレーションを用いて調べよ.

問題 9.1

区間 $[0, 2]$ 上の一様分布に従う母集団から取り出したサンプルサイズ 2 の標本 X, Y を考える. このとき, 相乗平均 \sqrt{XY} は母平均の不偏推定量ではないことを示せ.

問題 9.2

区間 $[0, L]$ から取り出した n 個の標本 X_1, X_2, \dots, X_n の最大値を $M = \max\{X_1, X_2, \dots, X_n\}$ とおく. 平均値 $\mathbf{E}[M]$ を計算して, L の不偏推定量を求めよ.

問題 9.3

指数分布に従う母集団のパラメータ λ の最尤推定量を求めよ.

問題 9.4

【Python 利用】正規母集団 $N(\mu, \sigma^2)$ から取り出したサンプルサイズ $n = 15$ の標本から標本平均 $\bar{x} = 16.38$ を得た. 母分散 $\sigma^2 = 0.42^2$ を既知として, 母平均 μ の 95% 信頼区間を求めよ. また, 90% 信頼区間を求めよ.

問題 9.5

【Python 利用】ある地区で無作為に選ばれた 280 人に対して, ある番組の視聴を調査したところ, 視聴率は 18.4% であった. この地区の視聴率の 95%信頼区間を求めよ. また, 90%信頼区間を求めよ.

問題 9.6

【Python 利用】ある地区で番組視聴率を無作為標本から推定するとき, 90%信頼区間の幅を 1%以下にするために必要な標本数を求めよ.

問題 9.7

【Python 利用】次は正規母集団から取り出された 8 個の標本である.

136 148 153 142 145 161 141 164

- (1) 母平均の 90%信頼区間を求めよ.
- (2) 母分散の 90%信頼区間を求めよ.

問題 10.1

【Python 利用】分散 10^2 の正規母集団から取り出された 8 個の標本が,

64, 41, 62, 48, 59, 35, 65, 56

となった. この標本は正規母集団 $N(50, 10^2)$ から取り出された無作為標本といえるか検定せよ.

問題 10.2

公平なコインと表の出る確率が 0.6 に調整されたコインが紛れてしまい, 見た目では区別できない. そこで, コインを実際に 150 回振ってみたところ表が 84 回出た. このコインは公平なコインであると言えるか検定せよ. さらに, その際の第 2 種誤り確率も求めよ.

問題 10.3

ある英語の資格試験の全国平均は 69 点であった. ある特訓コースを受講した 7 名の結果は

82, 65, 84, 68, 79, 86, 71

であった. 特訓の効果は認められるか検定せよ. ただし, 受験生の得点分布は正規分布であると仮定してよい.

問題 10.4

A 組 24 名, B 組 30 名に対して学習達成度を確認するために同じ試験をしたところ, A 組の平均点は 65.6, B 組の平均点は 62.2 であった. 両組の学習達成度に有意差はあるか検定せよ. ただし, 成績は両組のとも試験の点数は正規分布に従うとしてよく, その分散は A 組 7.3^2 , B 組 8.6^2 が知られている.

問題 10.5

【Python 利用】次の表は, 慢性頭痛に対して薬治療とリラックス治療の2種類の治療法の効果を比較するため, 被験者をランダムに2群に分けて, 頭痛がなくなるまでの回復時間(分)を調べたものである. ただし, 標準偏差は標本分散の平方根である. 等分散の検定で母分散は等しいことを確認した上で, 2種類の治療法の回復時間に有意差があるといえるか検定せよ.

	人数	平均回復時間	標準偏差
薬治療群	15	33.8	2.85
リラックス治療群	15	22.4	3.07

問題 10.6

【Python 利用】ある運動療法には血圧を下げる効果があると言われている。次の表は、8名の被験者に対して運動療法の前後で血圧を測定した結果である。運動療法に効果があったかどうか検定せよ。

被験者番号	1	2	3	4	5	6	7	8
療法前	137	152	165	142	130	152	142	148
療法後	135	146	158	135	139	147	145	135

問題 10.7

【Python 利用】次の表は左利きと右利きの 2 群に対してある運動能力を測定した結果である. この 2 群で運動能力に有意差があるかを順位和検定を用いて検定せよ.

左利き	2.0	0.5	1.6	2.8	1.8	9.0	1.7	3.1	1.0
右利き	0.8	0.6	1.5	1.3	0.7	0.3	1.2		

問題 10.8

【Python 利用】次の表は, ある町の成人を母集団として無作為抽出された 800 人の血液型を調べたものと, 全国の血液型の分布を示している. この町の血液型の分布は全国の血液型の分布に適合しているといえるか検定せよ.

	AB	A	B	O	合計
ある町 (人)	85	317	168	230	800
全国 (比率)	0.09	0.37	0.22	0.32	1.00

問題 10.9

【Python 利用】次の表は, 300 人の自動車所有者を年齢と過去 2 年間に起こした事故数に応じて分類したものである. 年齢と事故数の間に関係はあるだろうか. 独立性の検定によって答えよ.

	事故数 0	事故数 1 ~ 2	事故数 3
21 歳以下	8	23	14
22 ~ 26 歳	21	42	12
27 歳以上	71	90	19

問題 10.10

ある映画の客層に男女の違いはあるかを調べるために、無作為に選んだ 100 名を調べたところ、男性 44 人、女性 56 人であった。(1) 二項母集団の母比率の検定 (2) 適合度の検定、の 2 つの方法で検定せよ。

問題 11.1

与えられたデータ (x_i, y_i) に対して, x を説明変数, y を目的変数とする, 最小二乗法による線形回帰式は

$$\frac{y - \bar{y}}{s_y} = r_{xy} \frac{x - \bar{x}}{s_x}$$

を満たすことを示せ. さらに, y を説明変数, x を目的変数とする線形回帰式は上とは一般に異なることを示せ.

問題 11.2

総平方和 TSS, 残差平方和 RSS, 回帰平方和 ESS に関する等式

$$TSS = RSS + ESS$$

を示せ.

問題 11.3

【Python 利用】次の表は, スウェーデンにおけるヘラジカの生息地(北緯)と雌雄の成体量を示したものである. 雌雄別に単回帰モデルを求めて, 散布図と回帰直線を描画せよ. さらに, 回帰直線の傾き $\hat{\beta}_1$ の 95% 信頼区間を求めよ.

latitude	BM_fem	BM_mal	latitude	BM_fem	BM_mal
58.0	180.9	214.5	62.7	188.8	235.2
57.7	174.9	216.3	64.0	204.8	259.2
58.0	184.9	217.8	63.0	186.6	
57.9	177.7	214.0	63.5	188.2	
59.8	164.8	215.1	65.5	194.2	250.0
61.6	180.0	241.9	66.0	198.1	246.4
62.0	196.3	247.8	66.0	200.1	262.2

なお, 上記の数値データ (Swedish_Moose.csv) は次のサイトから入手可能である:

<https://www.kyoritsu-pub.co.jp/book/b10033631.html>

問題 11.4

母集団分布が $y = \beta_0 + \beta_1 x + \varepsilon, \varepsilon \sim N(0, \sigma^2)$ に従い, 母数 $\beta_0, \beta_1, \sigma^2$ が未知であるものとする. このとき, 単回帰モデルの回帰係数 β_0, β_1 の最尤推定量を求め, 最小二乗推定量と一致することを示せ. (重回帰モデルの場合に証明済みであるが, 単回帰モデルの最初の式に戻って導出せよ.)