## 統計工学課題

情報経営システム工学分野 B3

**学籍番号** : 24336488

**氏名** : 本間 三暉

## 課題 1

## 問題 1 記述統計

以下の表1はある地域に生息するオスのカブトムシの体長と角長の測定結果である。

**表 1:** カブトムシ調査結果

個体 No	体長 (mm)	角長 (mm)
1	33.2	14.5
2	38.1	15.1
3	42.6	21.5
4	45.3	22.4
5	49.2	16.3
6	50.6	22.5
7	51.7	21.8
8	54.8	24.9
9	57.1	25.4
10	60.2	15.6
11	61.8	25.1
12	64.3	45.2
13	68.5	45.3
14	68.9	32.6
15	69.1	32.4
16	69.3	35.4
17	75.5	41.6
18	75.7	42.6
19	77.6	41.4
20	78.1	45.6
21	78.5	46.2
22	79.2	45.8
23	79.3	45.4
24	84.8	47.9
25	87.6	48.6
26	88.1	50.7
27	89.1	49.3
28	89.6	52.3
29	97.5	51.4
30	98.8	51.2

## (1) 体長、角長をそれぞれ解析せよ (度数分布表、ヒストグラム、平均、分散、標準偏差を求め、結果について考察する)。

体長と角長の度数分布表をそれぞれ表2と表3に示す.

表 2: 体長の度数分布表

以上	未満	階級値	度数	累積度数	相対度数	累積相対度数
30	40	35	2	2	0.067	0.067
40	50	45	3	5	0.100	0.167
50	60	55	4	9	0.133	0.300
60	70	65	7	16	0.233	0.533
70	80	75	7	23	0.233	0.767
80	90	85	5	28	0.167	0.933
90	100	95	2	30	0.067	1.000

表 3: 角長の度数分布表

以上	未満	階級値	度数	累積度数	相対度数	累積相対度数
10	20	15	4	4	0.133	0.133
20	30	25	7	11	0.233	0.367
30	40	35	3	14	0.100	0.467
40	50	45	12	26	0.400	0.867
50	60	55	4	30	0.133	1.000

体長と角長のヒストグラムをそれぞれ図 1(a), 図 1(b) に示す.

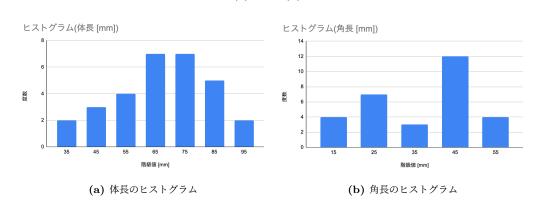


図 1: カブトムシのヒストグラム

表 4 に体長と角長の平均,分散,標準偏差を示す.データは,ある地域に生息する押すのカブトムシの測定結果であり,標本調査なので,分散と標準偏差は不偏分散と普遍標準偏差を求めた.

表 4: 体長と角長の統計量

	体長 (mm)	角長 (mm)
平均	68.80	35.87
分散	301.74	164.81
標準偏差	17.37	12.84

ヒストグラムから、体長は 65[mm] から 75[mm] を中心に大きな山となっていて、角長は 25[mm] と 45[mm] の 2 つの山に分布していることがわかる。また、データから角長より体長の値のほうが広範囲に分布していることがわかる。体長と角長のデータより、体長の小さい個体は角長が小さく、体長の大きい個体は角長が大きいことが予想できる。また、体長が 1 つの山に固まっているのに対し、角長が 2 つの山になっていることと、特定地域のオスの個体の標本であることから、成長段階の違う個体を測定していることが考えられるが、これらのデータだけでは断言することはできない。

## (2) 体長と角長の関係について調べよ (散布図、共分散、相関係数を求め、結果について考察する) 図 2 に体長と角長の散布図を示す.



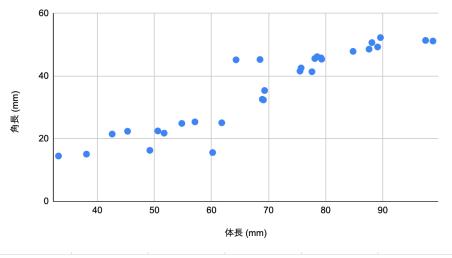


図 2: 散布図

表5に共分散と相関係数を示す.

表 5: 共分散と相関係数

項目	値
共分散	206.80
相関係数	0.93

まず、相関係数が 0.93 なので、体長と角長には強い正の相関があることがわかる。また、図 2 より、グラフ左下と右上の 2 グループに分かれているように見えることから、(1) で考察した内容と被る部分もあるが、体長の小さい個体は角長が小さく、体長の大きい個体は角長が大きいことが分かり、成長段階の違う個体を測定しているためこのような結果が現れるのではないかと考えられる。

#### メモ

## 共分散の定義

共分散の定義は式(1)のように定義される。これは2変量がxy平面でどのように散らばっているかを 表す.

$$S_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$
(1)

#### 相関係数の定義

対応のある 2変量 x, y について、式 (2) を x と y の相関係数という.

$$r = \frac{S_{xy}}{s_x \cdot s_y} \tag{2}$$

## 問題2 確率

コインを 3 回投げる試行で、

事象 A:1 回目に表がでる

事象 B:3 回とも同じ側が出る

とするとき、次の確率を求めなさい。

事象 A と B についてそれぞれの確率 P(A), P(B) の値を式 (3) と式 (4) に示す.

$$P(A) = \frac{1}{2} \tag{3}$$

$$P(A) = \frac{1}{2}$$

$$P(B) = \frac{2}{2^3} = \frac{2}{8} = \frac{1}{4}$$
(3)

## (1) $P(A \cup B)$

 $P(A \cup B)$  は事象 A か B のどちらか一方は必ず起こる確率である. 事象 A は一回目に表が出ること, 事象 B は三回とも同じ側が出ることなので、式 (5) に示すように、これらの確率を足し合わせ、重複す る部分を引いた値が解となる.

$$P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{4} - \frac{1}{8} = \frac{5}{8}$$
 (5)

(2)  $P(\overline{A})$ 

$$P(\overline{A}) = 1 - P(A) = 1 - \frac{1}{2} = \frac{1}{2}$$
 (6)

(3)  $P(\overline{B})$ 

$$P(\overline{B}) = 1 - P(B) = 1 - \frac{1}{4} = \frac{3}{4} \tag{7}$$

## 問題3 確率

袋に当たりくじ 1 枚とハズレくじ 5 枚が入っており、2 人で交互に 1 枚ずつくじを取り出すとする。引いたくじは戻さないとき、先に引き始めた者と、後から引き始めた者とではどちらが得か?

まず、n 回目にくじを引いたときあたりを引く確率を  $P_n$  とすると、今回あり得る事象の確率は式 (8) から式 (13) の式の内どれかで表せる.

$$P_1 = \frac{1}{6} \tag{8}$$

$$P_2 = \frac{5}{6} \times \frac{1}{5} = \frac{1}{6} \tag{9}$$

$$P_3 = \frac{5}{6} \times \frac{4}{5} \times \frac{1}{4} = \frac{1}{6} \tag{10}$$

$$P_4 = \frac{5}{6} \times \frac{4}{5} \times \frac{3}{4} \times \frac{1}{3} = \frac{1}{6} \tag{11}$$

$$P_5 = \frac{5}{6} \times \frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{6}$$
 (12)

$$P_6 = \frac{5}{6} \times \frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} \times \frac{1}{2} \times \frac{1}{1} = \frac{1}{6}$$
 (13)

つまり、それぞれのくじが引かれるかどうかが同様に確からしい場合、いつ引いても確率は変わらない。そのため、引く回数が同じ場合どちらも同じ確率であたりを引く可能性があるため、引く順番による差は無いと言える。

余談だが、これはくじを引く前に考えているためすべて同じ確率となるが、くじを引いた結果があらかじめ分かってしまうと確率が変わってしまう.

#### メモ

#### 排反事象

事象 A,B が背反  $\rightarrow A \cup B = \phi$ 

 $P(A \cap B) = P(A) + P(B) - P(A \cup B)$  のうち、 $P(A \cup B) = 0$  となる.

## 問題 4 確率と条件付き確率

## サイコロを2回振るとき、次の確率を求めなさい。

サイコロを 2 回振る場合の目の組み合わせは  $6 \times 6 = 36$  通りである.

## (1) 目の和が偶数である確率

和が偶数となる目の組み合わせは,偶数 + 偶数か,奇数 + 奇数の場合である.これを事象 A とすると,事象 A が起こり得るサイコロの組み合わせは式 (14),その時の確率は式 (15) と表せる.

$$n(A) = (3 \times 3) + (3 \times 3) = 18 \text{ in } 0$$
(14)

$$P(A) = \frac{n(A)}{6 \times 6} = \frac{18}{36} = \frac{1}{2} \tag{15}$$

#### (2) 目の和が 3 の倍数である確率

サイコロの出る目を3で割ったときのあまりとしてあり得る組み合わせは0, 1, 2 の3 通りとなる. その時、和が3 の倍数となるあまりの組み合わせは(0,0), (1,2), (2,1) の3 通りである.

また,2つのサイコロのあまりの組み合わせは $3 \times 3 = 9$  通りであるので,3 の倍数となるあまりの組み合わせとなる確率は  $\frac{3}{6} = \frac{1}{3}$  となる.

したがって、2つのサイコロの目の和が3の倍数になる事象をBとしたときの確率は、 $P(B)=\frac{1}{3}$ となる。

#### (3) 目の和が偶数であるとき、目の和が3の倍数でもある確率

事象 A と B は独立なので、式 (16) で表せる.

$$P_A(B) = \frac{P(A \cap B)}{P(A)} = \frac{P(B) \times P(A)}{P(A)} = P(B) = \frac{1}{3}$$
 (16)

## (4) 目の和が3の倍数であるとき、目の和が偶数でもある確率

事象 A と B は独立なので、式 (17) で表せる.

$$P_B(A) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A) = \frac{1}{3}$$
 (17)

## 問題 5 条件付き確率

あるお店で次のようなゲームが行われている。3 つの箱があり、うち1 つの箱には豪華商品が入っており、ほか2 つの箱は空である。客はこの中から1 つの箱を選び、もし商品が入っていたらもらうことができる。ただし、箱を1 つ選んだ後、残りの2 つのうち空の箱を店員が開け、選ぶ箱を変更するかどうかを尋ねてくる(残り2 つの両方が空の場合、無作為に開ける)。変更したほうがよいかどうか、答えよ。

この問題は、モンティ・ホール問題を模した問題であると考えられる。モンティ・ホール問題は選ぶ箱は変更した方が変える前の2倍当たりやすくなることで知られているが、計算で求めようと思う。

それぞれの箱を A,B,C としたとき,プレイヤーは A の箱を選び,司会者が B の箱を開けたと仮定する.この状況を E としたとき,E の状況下で箱 A,B,C が当たりであるという事象をそれぞれ事象 A,B,C とすると, $P(A) = P(B) = P(C) = \frac{1}{3}$  であり,求める確率は式 (18) で表せる.

$$P_E(A) = \frac{P(A \cap E)}{P(E)} \tag{18}$$

 $P(A \cap B) = P(A)P_A(E)$  について、A が正解であれば、B と C のどちらも外れとなり、B が開けられる確率は  $\frac{1}{2}$  であるから、 $P_A(E) = \frac{1}{2}$  となる.よって、 $P(A \cap E) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$  となる.

 $P(B \cap E) = P_B(E)P(B)$  について、B が正解の場合、E は起こらない(B のドアは開かない)ので、 $P_B(E) = 0$ . よって、 $P(B \cap E) = 0$  となる.

 $P(C\cap E)=P_C(E)P(C)$  について,C が正解であれば司会者は必ず B を開ける必要があるので, $P_C(E)=1$  であるから, $P_C(E)=1$  となる.よって, $P(C\cap E)=\frac{1}{3}\times 1=\frac{1}{3}$  つまり,

$$P(E) = P(A \cap E) + P(B \cap E) + P(C \cap E) \tag{19}$$

$$=\frac{1}{6}+0+\frac{1}{3}=\frac{1}{2}\tag{20}$$

よって,

$$P_E(A) = \frac{P(A \cap E)}{P(E)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$
 (21)

$$P_E(B) = \frac{P(B \cap E)}{P(E)} = 0 \tag{22}$$

$$P_E(C) = 1 - P_E(A) - P_E(B) = 1 - \frac{1}{3} - 0 = \frac{2}{3}$$
(23)

となる. E(最初に A を選んで B が開けられた) という条件下では、A が正解の確率が  $\frac{1}{3}$ 、B が正解の 確率が0, C が正解の確率が $\frac{2}{3}$  なので、箱を変えたほうが当たる確率が2 倍になるので変えたほうが良 いと言える.

## 問題6 ベイズ確率

X病は3000人に1人の割合で発症する病気であり、この病気に対する第1次血液検査 の精度は図3の通りである。

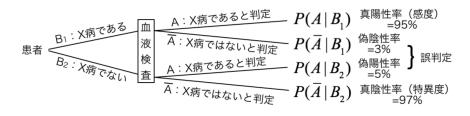


図3: X 病に対する第1次血液検査の精度

A: X 病であると判定されること (結果が陽性)

 $B_1$ : 実際に X 病であること  $B_2$ : 実際に X 病でないこと

各事象を上記のように定義する.

## (1) Y さんが 1 次検査の結果、X 病であると判定されたとき、本当に X 病にかかっている確率を求め

Y さんが 1 次検査の結果,X 病であると判定されたとき、,本当に X 病にかかっている確率を式 (27) に示す.

$$P_A(B_1) = \frac{P_{B_1}(A)P(B_1)}{P(A)} \tag{24}$$

$$= \frac{P_{B_1}(A)P(B_1)}{P_{B_1}(A)P(B_1) + P_{B_2}(A)P(B_2)}$$
 (25)

$$= \frac{P_{B_1}(A)P(B_1)}{P_{B_1}(A)P(B_1) + P_{B_2}(A)P(B_2)}$$

$$= \frac{0.95 \times \frac{1}{3000}}{0.95 \times \frac{1}{3000} + (1 - 0.95) \times \frac{3000 - 1}{3000}}$$
(25)

$$= 6.0 \times 10^{-4} \tag{27}$$

## (2) Y さんが 1 次検査の結果、X 病でないと判定されたにも関わらず、実際は X 病にかかっている確 率を求めなさい。

Y さんが 1 次検査の結果、X 病でないと判定されたにも関わらず、実際は X 病にかかっている確率 を式 (31) に示す.

$$P_{\overline{A}}(B_1) = \frac{P_{B1}(\overline{A})P(B_1)}{P(\overline{A})} \tag{28}$$

$$=\frac{P_{B_1}(\overline{A})P(B_1)}{P_{B_1}(\overline{A})P(B_1) + P_{B_2}(\overline{A})P(B_2)}$$

$$\tag{29}$$

$$= \frac{P_{B_1}(\overline{A})P(B_1)}{P_{B_1}(\overline{A})P(B_1) + P_{B_2}(\overline{A})P(B_2)}$$

$$= \frac{(1 - 0.97) \times \frac{1}{3000}}{(1 - 0.97) \times \frac{1}{3000} + 0.97 \times \frac{3000 - 1}{3000}}$$
(30)

$$=1.03\times10^{-5}$$
 (31)

#### メモ

#### 期待値の定義

確率変数の値を代表する平均(正確には、確率の重み付き平均)を期待値という、確率論でいう期待値 とは、得られる値の根拠のない期待ではなく、客観的な予想値のことである。離散型の定義を式(32)、 連続型の定義を式(33)に示す.

$$E(X) = \sum_{x} x f(x) \tag{32}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$
 (33)

#### 分散の定義

集中やばらつきを表すのに、期待値 E(X) からのずれの量 X-E(X) を考える. この時に、期待値を  $\mu = E(X)$  と表す.  $X - \mu$  はそれ自体確率変数なので、目安としてその平均を考えると、式 (34) のよ うになる.

$$E(X - \mu) = E(X) - \mu = \mu - \mu = 0 \tag{34}$$

となり、打ち消し合うため、0となる. そこで、二乗を考える. 式 (35) で示されるように、分散は次 のように定義される.

$$V(X) = E\{(X - \mu)^2\} \tag{35}$$

離散型および連続型の分散の定義をそれぞれ式(36)と式(37)に示す.

$$V(X) = \sum_{x} (x - \mu)^2 f(x)$$
 (36)

$$V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$
(37)

実際に計算する時は、式(38)を用いる方が便利である.

$$V(X) = E(X^{2}) - 2\mu E(X) + \mu^{2}$$

$$= E(X^{2}) - \mu^{2}$$

$$= E(X^{2}) - (E(X))^{2}$$
(38)

## 課題 2 確率分布と一様分布

指数分布と一様分布について、累積分布関数、期待値、分散をそれぞれ求めなさい。

指数分布を (1) に、一様分布を (2) に示す.

## (1) 指数分布

指数分布とは確率密度関数の分布のことで、確率密度関数は次のように定義される.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \ge 0) \\ 0 & (x < 0) \end{cases}$$

これは,  $f(x) \ge 0$  で,

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \lambda e^{-\lambda x} dx = 1$$

と規格化されている.

累積分布関数は,x < 0のとき,

$$F(x) = \int_{-\infty}^{x} f(u)du = \int_{-\infty}^{x} 0 \cdot du = 0$$

 $x \ge 0$  のとき,

$$F(x) = \int_0^x f(u)du = \int_0^x \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}$$

期待値は,

$$\begin{split} E(x) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{0}^{\infty} x \cdot \lambda e^{-\lambda x} dx \\ &= \left[ x \cdot (-e^{-\lambda x}) \right]_{0}^{\infty} - \int_{0}^{\infty} 1 \cdot (-e^{-\lambda x}) dx \\ &= - \left[ x e^{-\lambda x} \right]_{0}^{\infty} - \left[ -\frac{1}{\lambda} e^{-\lambda x} \right]_{0}^{\infty} \\ &= (0 - 0) - \left( 0 - \frac{1}{\lambda} \right) \\ &= \frac{1}{\lambda} \end{split}$$

分散は,

$$\begin{split} E(X^2) &= \int_0^\infty x^2 \cdot \lambda e^{-\lambda x} dx \\ &= \left[ x^2 \cdot (-e^{-\lambda x}) \right]_0^\infty - \int_0^\infty 2x \cdot (-e^{-\lambda x}) dx \\ &= 0 - \left[ 2x \cdot \left( \frac{1}{\lambda} e^{-\lambda x} \right) \right]_0^\infty - \int_0^\infty 2 \cdot \frac{1}{\lambda} e^{-\lambda x} dx \\ &= 0 - 0 - \left[ -\frac{2}{\lambda^2} e^{-\lambda x} \right]_0^\infty \\ &= 0 - \left( -\frac{2}{\lambda^2} \right) \\ &= \frac{2}{\lambda^2} \end{split}$$

$$V(X^2) = E(X^2) - (E(X))^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

## (2) 一様分布

一様分布とは、累積分布関数の分布のことで、累積分布関数は次のように定義される.

$$F(x) = \begin{cases} 0 & (x < 0) \\ x & (0 \le x \le 1) \\ 1 & (x > 1) \end{cases}$$

期待値は次のように計算される.

$$\begin{split} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{0}^{1} x \cdot 1 dx \\ &= \left[ \frac{x^2}{2} \right]_{0}^{1} \\ &= \frac{1}{2} \end{split}$$

よって、区間 [0,1] の中心となっている.

分散は次のように計算される.

$$E(X^{2}) = \int_{0}^{1} x^{2} \cdot 1 dx$$

$$= \left[\frac{x^{3}}{3}\right]_{0}^{1}$$

$$= \frac{1}{3}$$

$$V(X) = E(X^{2}) - (E(X))^{2}$$

$$= \frac{1}{3} - \left(\frac{1}{2}\right)^{2}$$

$$= \frac{1}{3} - \frac{1}{4}$$

$$= \frac{4}{12} - \frac{3}{12}$$

$$= \frac{1}{12}$$

## 課題3 指数分布に従う確率変数のモーメント

指数分布

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \ge 0) \\ 0 & (x < 0) \end{cases} \quad (\text{til } \lambda > 0)$$

に従う確率変数 X に対して、モーメント母関数は

$$M_X(t) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx \tag{39}$$

$$=\lambda \int_0^\infty e^{(t-\lambda)x} dx \tag{40}$$

$$= \frac{\lambda}{\lambda - t} \quad (t < \lambda \ \mathcal{O} \ \mathcal{E}) \tag{41}$$

となる。

# 式 (41) を繰り返し微分し、1 次、2 次、3 次、4 次、r 次のモーメントをそれぞれ求めよ。さらにこれらを用いて、期待値、分散、歪度、尖度を計算せよ。

1 次のモーメントを (1), 2 次を (2), 3 次を (3), 4 次を (4), r 次を (5) に示す. また, 期待値を (6), 分散を (7), 歪度を (8), 尖度を (9) に示す.

## (1) 1次のモーメント

1次モーメントは,

$$M_X'(t) = \frac{\lambda}{(\lambda - t)^2}$$

また, t=0 の 1 次モーメントは,

$$\mu_1 = M_X'(0) = \frac{1}{\lambda}$$

(2) 2次のモーメント

2次モーメントは,

$$M_X''(t) = \frac{2\lambda}{(\lambda - t)^3}$$

また, t=0 の 2 次モーメントは,

$$\mu_2 = M_X''(0) = \frac{2}{\lambda^2}$$

(3) 3次のモーメント

3次モーメントは,

$$M_X^{(3)}(t) = \frac{6\lambda}{(\lambda - t)^4}$$

また, t=0 の 3 次モーメントは,

$$\mu_3 = M_X^{(3)}(0) = \frac{6}{\lambda^3}$$

(4) 4次のモーメント

4次モーメントは,

$$M_X^{(4)}(t) = \frac{24\lambda}{(\lambda - t)^5}$$

また, t=0 の 4 次モーメントは,

$$\mu_4 = M_X^{(4)}(0) = \frac{24}{\lambda^4}$$

(5) r 次のモーメント

r次モーメントは、

$$M_X^{(r)}(t) = \frac{r! \cdot \lambda}{(\lambda - t)^{r+1}}$$

また, t=0 の r 次モーメントは,

$$\mu_r = M_X^{(r)}(0) = \frac{r!}{\lambda^r}$$

(6) 期待値

$$E(X) = \mu_1 = \frac{1}{\lambda} \tag{42}$$

(7) 分散

$$V(X) = \sigma^2 = \mu_2 - \mu_1^2 = \frac{1}{\lambda^2}$$
(43)

## (8) 歪度

$$\alpha_3 = \frac{E((X - \mu)^3)}{\sigma^3} \tag{44}$$

$$=\frac{E(X^3) - 3\mu E(X^2) + 2\mu^3}{(\sigma^2)^{3/2}} \tag{45}$$

$$= \frac{E(X^3) - 3\mu E(X^2) + 2\mu^3}{(\sigma^2)^{3/2}}$$

$$= \frac{\mu_3 - 3\mu \cdot \mu_2 + 2\mu^3}{(\sigma^2)^{3/2}}$$
(45)

$$=2\tag{47}$$

## (9) 尖度

$$\alpha_4 - 3 = \frac{E((X - \mu)^4)}{\sigma^4} - 3 \tag{48}$$

$$= \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{(\sigma^2)^2} - 3$$

$$= \frac{\mu_4 - 4\mu \cdot \mu_3 + 6\mu^2 \cdot \mu_2 - 3\mu^4}{(\sigma^2)^2} - 3$$
(50)

$$=\frac{\mu_4 - 4\mu \cdot \mu_3 + 6\mu^2 \cdot \mu_2 - 3\mu^4}{(\sigma^2)^2} - 3\tag{50}$$

$$=9-3\tag{51}$$

$$=6\tag{52}$$

#### メモ

## 歪度(わいど)

非対称性の指標として使われる.

$$\alpha_3 = \frac{E((X - \mu)^3)}{\sigma^3}$$

が用いられる.  $lpha_3>0$  ならば右の裾が長く,  $lpha_3<0$  ならば左の裾が長い.  $|lpha_3|$  がその程度を表す.  $eta_3$ と表記することがある.

$$E((X - \mu)^3) = E(X^3) - 3\mu E(X^2) + 3\mu^2 E(X) - \mu^3$$
$$= E(X^3) - 3\mu E(X^2) + 2\mu^3$$

## 尖度(せんど)

確率分布の尖り(=平均近くに集中し、かつ裾が重い)の程度を表す指標として使われる.

$$\alpha_4 = \frac{E((X - \mu)^4)}{\sigma^4}$$

と表記することがある.

普通は,正規分布の  $\alpha_4=3$  と比較して, $\alpha_4-3$  を X の確率分布の尖度(超過係数)という.  $\alpha - 3 > 0$  ならば正規分布より尖っている,  $\alpha - 3 < 0$  ならば正規分布より丸く鈍い形をしている. 計算には以下を用いることができる.

$$E((X - \mu)^4) = E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 4\mu^3 E(X) + \mu^4$$
$$= E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4$$

4次関数は小さな変化に敏感.中心近くと遠くのコントラストの激しさを表す.中心付近でその小さい値にもかかわらず十分大きな密度がありかつ,遠くからの影響が少しでもあれば尖度は大きくなる.

## チェビシェフの不等式

いかなる確率変数についても、期待値と分散がわかっていれば、確率の値が不等式の形で現れることを示す.

$$P(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}, \quad (k > 0)$$

課題 4 ある確率変数 X について E(X)=1,  $V(X)=\frac{1}{3}$  であることがわかっている。このとき、 $0\leq X\leq 2$  となる確率に見当をつけよ。

$$\mu = 1, \quad \sigma = \frac{1}{\sqrt{3}}$$

$$P(0 \le X \le 2) = P(|X - \mu| < \sqrt{3}\sigma) > 1 - \left(\frac{1}{\sqrt{3}}\right)^2 = \frac{2}{3}$$

となり、確率は、 $\frac{2}{3}$ 以上であることがわかる.

## 課題 5

問題1 正規分布に従う確率変数の期待値と分散を証明せよ。

ヒント: 置換積分を用いる。特に、分散の計算には部分積分、べき関数と指数関数の比の極限、ガウス積分も用いる。なお、ガウス積分は式 (53) である。

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}} \quad (a > 0)$$
 (53)

式 (33) より,

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \tag{54}$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{55}$$

 $\frac{x-\mu}{\sigma}=t$  と積分変数を置換する.

$$x = \sigma t + \mu \tag{56}$$

$$\frac{(x-\mu)^2}{2\sigma^2} = \frac{t^2}{2} \tag{57}$$

$$\frac{dx}{dt} = \sigma \tag{58}$$

ガウス積分の計算結果を使うと,

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}} \quad (a > 0)$$

$$\tag{59}$$

$$E(X) = 0 + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\left(-\frac{t^2}{2}\right)} dt \tag{60}$$

$$=\frac{\mu}{\sqrt{2\pi}}\cdot\sqrt{2\pi}\tag{61}$$

$$=\mu \tag{62}$$

分散は,

$$V(X) = E(X - E(X))^{2} = E(X - \mu)^{2}$$
(63)

$$= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \tag{64}$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$
 (65)

 $\frac{x-\mu}{\sigma} = t \ \texttt{L} \, \texttt{J} \, \texttt{SL},$ 

$$x = \sigma t + \mu \tag{66}$$

$$\frac{(x-\mu)^2}{2\sigma^2} = \frac{t^2}{2}$$

$$\frac{dx}{dt} = \sigma$$
(67)

$$\frac{dx}{dt} = \sigma \tag{68}$$

置換積分すると,

$$V(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)^2 e^{\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)} dx \tag{69}$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \sigma^2 t^2 e^{\left(-\frac{t^2}{2}\right)\sigma} dt \tag{70}$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{\left(-\frac{t^2}{2}\right)} dt \tag{71}$$

ここでガウス積分の計算結果を用いると,

$$V(X) = \sigma^2$$

## 問題 2 対数密度関数の期待値と分散を証明せよ。

 $Y \sim N(\mu, \sigma^2)$  で, $X = e^Y$  の確率分布を求める. $Y = \log(X)$  となり,確率密度関数 f(x) は,

$$f(x) = \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \frac{1}{x} dx \tag{72}$$

$$= \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} dx \tag{73}$$

期待値は、 $\log x=t$  と置換すると、 $x=e^t$ 、 $\frac{dx}{dt}=e^t$  となる。また、積分範囲は  $(0\to\infty)\to (-\infty\to\infty)$  となる。

$$\begin{split} E(X) &= \int_0^\infty x \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} e^t dt \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} \left(t - (\mu + \sigma^2)\right)^2 + \frac{1}{2} (2\mu + \sigma^2)} dt \\ &= e^{\frac{1}{2} (2\mu + \sigma^2)} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} \left(t - (\mu + \sigma^2)\right)^2} dt \\ &= e^{\mu + \frac{\sigma^2}{2}} \end{split}$$

分散は、 $E(X^2) - (E(X))^2$  で求められるので、

$$\begin{split} E(X^2) &= \int_0^\infty x^2 \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^\infty e^{2u} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u - \mu)^2}{2\sigma^2}} du \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} \left(u - (\mu + \sigma^2)\right)^2 + 2(\mu + \sigma^2)} du \\ &= e^{2(\mu + \sigma^2)} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} \left(u - (\mu + \sigma^2)\right)^2} du \\ &= e^{2\mu + 2\sigma^2} \end{split}$$

$$V(X) = E(X^{2}) - (E(X))^{2}$$

$$= e^{2\mu + 2\sigma^{2}} - \left(e^{\mu + \frac{\sigma^{2}}{2}}\right)^{2}$$

$$= e^{2\mu + \sigma^{2}} - e^{2\mu + 2\sigma^{2}}$$

問題 3 確率変数 X が標準正規分布 N(0,1) に従うとき、X はガンマ分布  $Ga(\frac{1}{2},\frac{1}{2})$  に従う。これを証明せよ。

ヒント: 確率変数 Y=X の累積分布関数を考え、これを微分することで密度関数を求める。

標準正規分布 N(0,1) に従う確率変数 X の確率密度関数は以下のように与えられる.

$$f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

ここで、 $X^2$  の確率密度関数  $f_{X^2}(x)$  を求めるには、変数変換を使用する.  $Y=X^2$  とおくと、Y の密度関数は以下のように求められる.

$$f_Y(y) = f_X(\sqrt{y}) \left| \frac{d}{dy}(\sqrt{y}) \right| + f_X(-\sqrt{y}) \left| \frac{d}{dy}(-\sqrt{y}) \right|$$

Y が X の正の値と負の値から得られることを考慮に入れる.  $f_X(x)$  は偶関数なので,  $f_X(\sqrt{y})=f_X(-\sqrt{y})$  よって, Y の確率密度関数は次のようになる.

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}} + \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}}$$

これを  $\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$  の形式に変換するために、確率密度関数は次のようになる.

$$f(y; k, \theta) = \frac{y^{k-1}e^{-\frac{y}{\theta}}}{\theta^k \Gamma(k)}$$

ここで、 $k=\frac{1}{2}, \theta=\frac{1}{2}$ を代入すると

$$f(y; \frac{1}{2}, \frac{1}{2}) = \frac{y^{\frac{1}{2} - 1} e^{-y}}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})} = \frac{1}{\sqrt{\pi y}} e^{-y} = \frac{2}{\sqrt{2\pi y}} e^{-2(\frac{y}{2})}$$

よって、 $X^2$  はガンマ分布  $Ga\left(\frac{1}{2},\frac{1}{2}\right)$  に従う.

# 問題 4 ベータ分布 Be(lpha,eta) の期待値と分散を証明せよ。 なお式 (74) に示すベータ 関数の積分公式を用いる。

$$\int_0^1 x^m (1-x)^n dx = \frac{m! \, n!}{(m+n+1)!} \tag{74}$$

ベータ関数は.

$$B(p,q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$$

また,期待値は,

$$E(X) = \int_0^1 x \frac{1}{B(\alpha, \beta)} x^{\alpha - 1} (1 - x)^{\beta - 1} dx$$

$$= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{(\alpha + 1) - 1} (1 - x)^{\beta - 1} dx$$

$$= \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)}$$

$$= \frac{\alpha}{\alpha + \beta} \frac{B(\alpha, \beta)}{B(\alpha, \beta)}$$

$$= \frac{\alpha}{\alpha + \beta}$$
(75)

分散は、 $E(X^2) - (E(X))^2$  で求められるので、

$$E(X^{2}) = \int_{0}^{1} x^{2} \frac{1}{B(\alpha, \beta)} x^{\alpha - 1} (1 - x)^{\beta - 1} dx$$
 (76)

$$= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{(\alpha+2)-1} (1-x)^{\beta-1} dx$$
 (77)

$$=\frac{B(\alpha+2,\beta)}{B(\alpha,\beta)}\tag{78}$$

$$= \frac{\alpha+1}{\alpha+\beta+1} \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \tag{79}$$

$$= \frac{\alpha+1}{\alpha+\beta+1} \frac{\alpha}{\alpha+\beta} \frac{B(\alpha,\beta)}{B(\alpha,\beta)}$$
(80)

$$= \frac{\alpha^2 + \beta}{(\alpha + \beta)(\alpha + \beta + 1)} \tag{81}$$

$$V(X) = E(X^{2}) - (E(X))^{2}$$
(82)

$$= \frac{\alpha^2 + \beta}{\alpha + \beta + 1} - \left(\frac{\alpha}{\alpha + \beta}\right)^2 \tag{83}$$

$$= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{84}$$

## 課題 6

問題 1 式 85 に示すような同時確率分布をもつ X,Y の周辺確率分布 (グラフの概形も描け)、期待値、分散および X,Y の相関係数を求めよ。

$$f(x,y) = \begin{cases} 6(x-y) & (0 \le y < x \le 1) \\ 0 & \text{otherwise} \end{cases}$$
 (85)

周辺確率密度関数  $f_X(x)$ ,  $f_Y(y)$  である.左は確率変数 X の周辺確率密度関数で,右は確率変数 Y の周辺確率密度関数である.X のほうは,x が 1 に近づくにつれて増加し,Y の方は減少している.

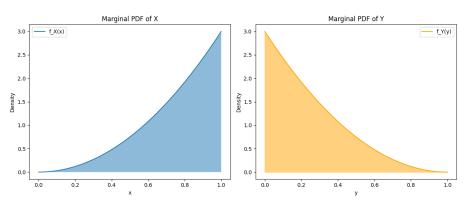


図 4: 周辺確率密度関数

y については、 $0 \le y < x$  以外で、被積分関数は 0 (x も同様) であることに注意して、

$$g(x) = \int_0^x 6(x - y)dy = 3x^2,$$
  
$$h(y) = \int_y^1 6(x - y)dx = 3(1 - y)^2$$

となる. これらの密度関数は  $\frac{1}{2}$  について線対称である. 期待値は、

$$E(X) = \int_0^1 x \cdot g(x) dx = \int_0^1 x \cdot 3x^2 dx = \frac{3}{4},$$
  
$$E(Y) = \int_0^1 y \cdot h(y) dy = \int_0^1 y \cdot 3(1-y)^2 dy = \frac{1}{4}.$$

$$E(X^{2}) = \int_{0}^{1} x^{2} \cdot 3x^{2} dx = \frac{3}{5},$$
  

$$E(Y^{2}) = \int_{0}^{1} y^{2} \cdot 3(1 - y)^{2} dy = \frac{1}{10}.$$

よって、分散は、

$$V(X) = \frac{3}{80}, \quad V(Y) = \frac{3}{80}.$$

共分散は,

$$Cov(X,Y) = \frac{1}{5} - \frac{3}{4} \cdot \frac{1}{4} = \frac{1}{80}.$$

相関係数は,

$$\rho = \frac{\frac{1}{80}}{\sqrt{\frac{3}{80}} \cdot \sqrt{\frac{3}{80}}} = \frac{1}{3}.$$

問題 2 たたみこみの直接計算によって,上で挙げた分布の再生性を証明せよ. すなわち,X,Y が独立で,(1),(2),(3) を満たすことを証明せよ.

ヒント:(1) では式 (86) に示すヴァンデルモンドのたたみこみ定理を用いると良い。また,(3) では式 (87) に示す指数部分が二次関数の一般形になる場合のガウス積分を用いると良い。

$$\sum_{x=0}^{z} \binom{n}{x} \binom{m}{z-x} = \binom{n+m}{z} \tag{86}$$

$$\int_{-\infty}^{\infty} \exp\{-\left(ax^2 + bx + c\right)\}dx = \exp\left(\frac{b^2}{4a} - c\right)\sqrt{\frac{\pi}{-a}}$$
(87)

(1) p が等しい二項分布に従うならば,X+Y も二項分布に従う

$$Bi(n,p) * Bi(m,p) = \sum_{x=0}^{z} \binom{n}{x} p^{x} q^{n-x} \cdot \binom{m}{z-x} p^{z-x} q^{m-(z-x)}$$

$$= \sum_{x=0}^{z} \binom{n}{x} \binom{m}{z-x} p^{z} q^{(n+m)-z}$$

$$= \binom{n+m}{z} p^{z} q^{(n+m)-z}$$

$$= Bi(n+m,p)$$
(88)

2行目から3行目への変換で、ヴァンデルモンドの畳み込み定理を用いた.

## (2) ポアソン分布に従うならば、X+Y もポアソン分布に従う

$$P_0(\lambda) * P_0(\mu) = \sum_{x=0}^z \frac{e^{-\lambda} \lambda^x}{x!} \cdot \frac{e^{-\mu} \mu^{z-x}}{(z-x)!}$$

$$= e^{-(\lambda+\mu)} \sum_{x=0}^z \frac{x!}{x!(z-x)!} \cdot \frac{\lambda^x \mu^{z-x}}{z!}$$

$$= e^{-(\lambda+\mu)} \cdot \frac{(\lambda+\mu)^z}{z!}$$

$$= P_0(\lambda+\mu)$$

## (3) 正規分布に従うならば,X+Y も正規分布に従う

$$N(\mu_{1}, \sigma_{1}^{2}) + N(\mu_{2}, \sigma_{2}^{2}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{1}^{2}}} e^{-\frac{(x-\mu_{1})^{2}}{2\sigma_{1}^{2}}} \cdot \frac{1}{\sqrt{2\pi\sigma_{2}^{2}}} e^{-\frac{(z-x-\mu_{2})^{2}}{2\sigma_{2}^{2}}} dx$$

$$= \frac{1}{2\pi\sigma_{1}\sigma_{2}} \int_{-\infty}^{\infty} e^{-\left[\frac{(x-\mu_{1})^{2}}{2\sigma_{1}^{2}} + \frac{(z-x-\mu_{2})^{2}}{2\sigma_{2}^{2}}\right]} dx$$
(89)

ここで、指数部分について平方完成し、できた定数項をzについて平方完成すると、

## 課題7 中心極限定理を証明せよ

互いに独立な n この確率変数  $X_1, X_2, \ldots, X_n$  が平均  $\mu$ , 分散  $\sigma^2$  を持つ同一の確率分布に従うとする.  $n \to \infty$  のとき,

$$P\left(a \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq b\right) \to \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (\text{txt.}, \quad \overline{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n))$$

左辺にある、 $\overline{X}$  の標準化変数を変形すると

$$\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sqrt{n}\sigma} = \frac{Y_1 + Y_2 + \dots + Y_n}{\sqrt{n}}$$
(90)

の確率分布が、n が大きいとき正規分布に近づくことを示せばよい. ここで、

$$Y_1 = \frac{X_1 - \mu}{\sigma}, Y_2 = \frac{X_2 - \mu}{\sigma}, \dots, Y_n = \frac{X_n - \mu}{\sigma}$$

は、それぞれ  $X_1, X_2, \ldots, X_n$  の標準化変数であって,

$$E(Y_1) = E(Y_2) = \cdots = E(Y_n) = 0, V(Y_1) = V(Y_2) = \cdots = V(Y_n) = 1$$

となっている。 $X_1, X_2, \ldots, X_n$  の確率分布は全て同一であり、 $Y_1, Y_2, \ldots, Y_n$  についてもそうであるから、その一つを Y とおいてモーメント母関数を求める。1 次のモーメント  $\mu_1 = E(Y) = 0$ 、2 次のモーメント  $\mu_2 = E(Y^2) = 1$  なので、

$$M_Y(t) = 1 + \frac{t^2}{2} + \cdots$$

独立な確率変数の確率分布はモーメント母関数の積で求まるので、 $Y_1+Y_2+\cdots+Y_n\equiv T$  モーメント母関数は、

$$M_n(t) = (M_Y(t))^n = \left(1 + \frac{t^2}{2} + \cdots\right)^n$$

最後的に  $\frac{Y_1+Y_2+\cdots+Y_n}{\sqrt{n}}$  のモーメント母関数は,  $t=\frac{t}{\sqrt{n}}$  とおきかえて,

$$M_n\left(\frac{t}{\sqrt{n}}\right) = \left(M_Y\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \frac{t^2}{2n} + \cdots\right)^n \to e^{\left(\frac{t^2}{2}\right)}$$

 $e^{\left(\frac{t^2}{2}\right)}$  は標準正規分布のモーメント母関数である.

## 課題 8 1 袋 50g 入りの菓子製造工程において、製品 10 個を無作為に 抽出し、重さを測定したところ、次のようになった。

1 袋中の菓子の重さ (g):51.9, 49.5, 50.2, 50.3, 50.1, 52.1, 48.8, 48.5, 49.7, 49.6

## (1) 全製品の平均重量とその分散の不偏推定量を求めよ。

平均重量は,

$$\overline{x} = \frac{1}{10} \sum_{i} x_i = \frac{1}{10} \times 500.7 = 50.07$$
 (91)

また、分散の不偏推定量は、

$$s^{2} = \frac{1}{10 - 1} \left( \sum_{i} (x_{i}^{2}) - n\overline{x}^{2} \right) \tag{92}$$

$$= \frac{1}{10-1} \left( 25082.35 - 10 \times 50.07^2 \right) \tag{93}$$

$$=1.3668$$
 (94)

## (2) 次ページ以降の数表を用いて、母平均 $\mu$ と母分散 $\sigma^2$ の 99% 信頼区間を求めよ。

99% 信頼区間を求めるので、 $\gamma = 0.99$  つまり、 $\alpha = 1 - \gamma = 0.01$  である. 母平均  $\mu$  について、

$$\overline{x} - t_{n-1} \left(\frac{\alpha}{2}\right) \sqrt{\frac{s^2}{n}} \le \mu \le \overline{x} + t_{n-1} \left(\frac{\alpha}{2}\right) \sqrt{\frac{s^2}{n}}$$

を t 分布の数表を用いて求めれば良い.

$$t_{n-1}\left(\frac{\alpha}{2}\right) = t_9(0.005) = 3.250$$

$$t_{n-1}\left(\frac{\alpha}{2}\right)\sqrt{\frac{s^2}{n}} = 3.250 \times \sqrt{\frac{1.3668}{10}} = 1.2015$$

従って,

上側限界 = 
$$50.07 + 1.2015 = 51.2715$$

下側限界 = 
$$50.07 - 1.2015 = 48.8685$$

以上より、母平均の99%信頼区間は

$$48.87 < \mu < 51.27$$

## 母分散について

母分散  $\sigma^2$  について,

$$\frac{(n-1)s^2}{\chi_{n-1}^2(\frac{\alpha}{2})} < \sigma^2 < \frac{(n-1)s^2}{\chi_{n-1}^2(1-\frac{\alpha}{2})}$$

を  $\chi^2$  分布の数表を用いて求めれば良い.

$$\chi_{n-1}^2 \left(\frac{\alpha}{2}\right) = \chi_9^2(0.005) = 23.5894$$

$$\chi_{n-1}^2 \left(1 - \frac{\alpha}{2}\right) = \chi_9^2(0.995) = 1.7349$$

$$\frac{(n-1)s^2}{\chi_{n-1}^2 \left(\frac{\alpha}{2}\right)} = \frac{9 \times 1.3668}{23.5894} = 0.5215$$

$$\frac{(n-1)s^2}{\chi_{n-1}^2 \left(1 - \frac{\alpha}{2}\right)} = \frac{9 \times 1.3668}{1.7349} = 7.0904$$

以上より母分散の 99% 信頼区間は

$$0.52 < \sigma^2 < 7.09$$