

2次元正規分布

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 II L11(2016-07-07 Thu)

最終更新: Time-stamp: "2016-07-07 Thu 13:20 JST hig"

今日の目標

- 2次元正規分布の定義が説明できる.
- 2次元正規分布の確率密度関数から母平均値と共分散行列が求められる. その逆.



<http://hig3.net>

L10-Q1

Quiz 解答:分散分析

水準の数 $\ell = 3$, 繰り返しの数 $r = 5$.

$$\bar{y}_{1\bullet} = 10, \bar{y}_{2\bullet} = 15, \bar{y}_{3\bullet} = 23, \bar{y}_{\bullet\bullet} = 16.$$

級間平方和 (級間変動)

$$S_A = \sum_j \sum_i (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 = 5 \times \sum_i (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 = 430.$$

$$\text{残差平方和 (誤差変動)} S_E = \sum_i \sum_j (y_{ij} - \bar{y}_{i\bullet})^2 = 106.$$

$$\text{全平方和 (全変動)} S_T = \sum_i \sum_j (y_{ij} - \bar{y}_{\bullet\bullet})^2 = 430 + 106 = 536.$$

分散分析表は次の通り.

要因	平方和	自由度	平均平方	F_0
級間 A	430	$3 - 1 = 2$	$430/2 = 215$	$215/8.833 = 24.34$
残差 E	106	$14 - 2 = 12$	$106/12 = 8.833$	
全 T	536	$15 - 1 = 14$		

$24.34 > F_{0.05}(2, 12) = 3.885$ より, 全水準の母平均値が等しいという帰無仮説は棄却される.

ここまで来たよ

3 分散分析・2次元正規分布

4 2次元正規分布

- 2変量の連続型確率変数
- 2次元正規分布

復習:2 変量の離散的確率変数の同時分布

同時分布

確率統計☆演習 I(2016)L01

$$P(X = x, Y = y) = f_{XY}(x, y)$$

表で書いたほうが見やすい.

$y \backslash x$	158	160	165
45	3/8	0	1/12
50	1/8	1/3	1/12

$y \backslash x$	x_1	x_2	x_3
y_1	$f_{XY}(x_1, y_1)$	$f_{XY}(x_2, y_1)$	$f_{XY}(x_3, y_1)$
y_2	$f_{XY}(x_1, y_2)$	$f_{XY}(x_2, y_2)$	$f_{XY}(x_3, y_2)$

2 変量の離散型確率変数の母期待値

$$E[\phi(X, Y)] = \sum_{i=1}^a \sum_{j=1}^b \phi(x_i, y_j) f_{XY}(x_i, y_j)$$

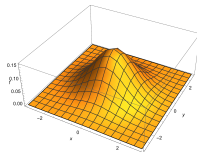
2 変量の連続型確率変数の同時分布

確率密度関数 (2 変数関数)

$$f_{XY}(x, y)$$

同時分布

確率統計☆演習 I(2015)L09



2 変量の連続型確率変数の母期待値

$$E[\phi(X, Y)] = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \phi(x, y) f_{XY}(x, y)$$

2 変量の連続型確率変数の確率 (母比率)

$$\begin{aligned} P(a \leq X < b, c \leq Y < d) &= E[\mathbf{1}_{[a \leq x < b, c \leq y < d]}(X, Y)] \\ &= \int_a^b dx \int_c^d dy f_{XY}(x, y). \quad \text{体積} \end{aligned}$$

ここまで来たよ

3 分散分析・2次元正規分布

4 2次元正規分布

- 2変量の連続型確率変数
- 2次元正規分布

復習:1 変数の正規分布

標準正規分布の確率密度関数

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$X = aZ + b$ を考える。 確率統計☆演習 I(2015)L08

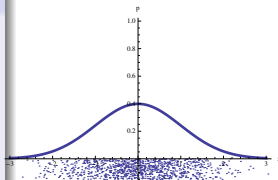
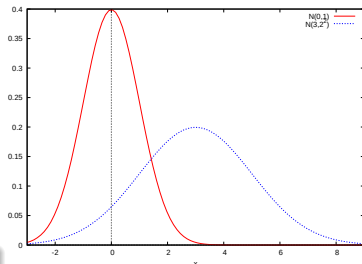
確率密度関数は、 z のところに $z = \frac{x-b}{a} = \frac{x-\mu}{\sigma}$ を代入すればいいので、

正規分布 $N(\mu, \sigma^2)$ の確率密度関数

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

パラメタ μ (= 実は $E[X]$),
 σ^2 (= 実は $V[X]$).

確率統計☆演習 I(2015)L08



2次元正規分布 (のうち X, Y が独立な簡単なケース)

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \times \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}.$$

$$E[X] = \mu_X, \quad E[Y] = \mu_Y, \quad V[X] = \sigma_X^2, \quad V[Y] = \sigma_Y^2,$$

$$\text{母共分散 } C_{XY} = (\text{独立なので}) = 0,$$

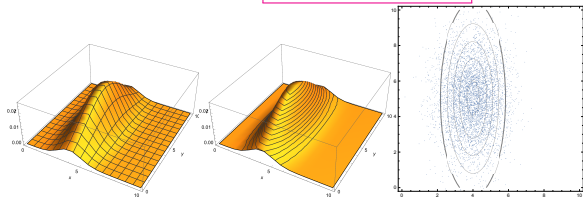
$$\text{母相関係数 } r_{XY} = \frac{\text{Cov}[X, Y]}{\sqrt{V[X]}\sqrt{V[Y]}} = 0$$

2 変量データ

確率統計☆演習 I(2015)L04

同時分布

確率統計☆演習 I(2015)L09



等高線は楕円.

ここで使った/使う「公式」

モーメント母関数 $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. から簡単に示せる.

$f(x; \mu, \sigma^2)$: 1 次元正規分布の確率密度関数.

$$\int_{-\infty}^{+\infty} x^0 f(x; \mu, \sigma^2) dx = \int_{-\infty}^{+\infty} x^0 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$$

$$\int_{-\infty}^{+\infty} x^1 f(x; \mu, \sigma^2) dx = \int_{-\infty}^{+\infty} x^1 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu.$$

$$\int_{-\infty}^{+\infty} x^2 f(x; \mu, \sigma^2) dx = \int_{-\infty}^{+\infty} x^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2 + \mu^2.$$

$$\int_{-\infty}^{+\infty} (x - \mu)^2 f(x; \mu, \sigma^2) dx = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2$$

L11-Q1

Quiz(2 次元正規分布)

次の 2 変数確率密度関数は 2 次元正規分布を定める.

$$f(x, y) = C \cdot e^{-x^2 - 4x - 2y^2 + 12y - 5}.$$

- ① X, Y の母平均値, 母分散, 母共分散を求めよう.
- ② $E[1] = 1$ が満たされるように定数 C を定めよう.

2 次形式の標準化 (線形代数 I)

L11-Q2

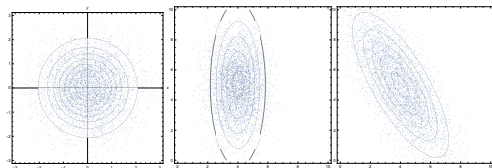
Quiz(2 次元正規分布)

次の 2 変数確率密度関数は 2 次元正規分布を定める.

$$f(x, y) = C \cdot e^{-4x^2 - \frac{1}{6}y^2 + 2y}$$

- ① X, Y の母平均値, 母分散, 母共分散を求めよう.
- ② $E[1] = 1$ が満たされるように定数 C を定めよう.

等高線が傾いた楕円の分布も作りたい!



考え方 1

等高線が円

$$D = x^2 + y^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix}$$

から出発して, 回転と拡大縮小する.

$$\begin{aligned} D &= \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1/\sigma_X & 0 \\ 0 & 1/\sigma_Y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \\ &\quad \times \begin{pmatrix} 1/\sigma_X & 0 \\ 0 & 1/\sigma_Y \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \end{aligned}$$

考え方 2 指数関数の肩に xy の項も入れておけばいい.

2 次元正規分布

2 次元正規分布, 2 変量正規分布

2 次元正規分布 $N(\boldsymbol{\mu}, V)$ の確率密度関数は, 確率変数を $\mathbf{X} = {}^t(X, Y)$ とするとき,

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{2/2} \sqrt{\det V}} e^{-\frac{1}{2} {}^t(\mathbf{x} - \boldsymbol{\mu}) V^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

ただし, パラメタは,

$$\text{母平均値 (ベクトル)} \boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} (= \text{実は} \begin{pmatrix} E[X] \\ E[Y] \end{pmatrix})$$

$$\text{母共分散行列 } V = \begin{pmatrix} \sigma_X^2 & C_{XY} \\ C_{XY} & \sigma_Y^2 \end{pmatrix} (= \text{実は} \begin{pmatrix} V[X] & \text{Cov}[X, Y] \\ \text{Cov}[X, Y] & V[Y] \end{pmatrix})$$

母共分散行列 V は大文字だが, 確率変数ではなく, パラメタ (正値対称行列). σ^2 が 2 次元になったもの. V の固有値は 2 個とも正である必要. パラメタ V が母分散行列になることはわりに簡単に証明できる.

Example (検算)

$E[X] = \mu_X, E[Y] = \mu_Y, V[X] = \sigma_X^2, V[Y] = \sigma_Y^2, \text{Cov}[X, Y] = 0$ のときに独立の場合の式に戻ることをチェックしよう.

Example (2 次元正規分布の確率密度関数)

$$V = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix},$$

$$\mu = \begin{pmatrix} -4 \\ 5 \end{pmatrix}$$

のとき, f を (行列やベクトルを使わずに) 具体的に書こう.

超便利なファミリー: 条件付き分布, 周辺分布は 1 次元正規分布

母相関係数 r , 共分散行列 V . $c = \frac{9\sqrt{21}}{10}$

$r = -0.90$

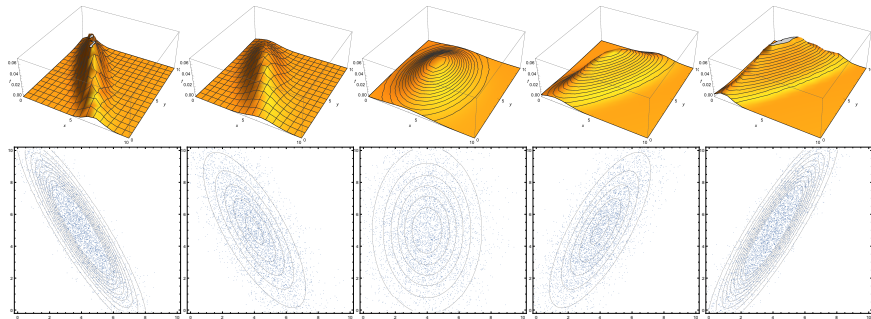
-0.76

0

$+0.76$

$+0.90$

$$V = \begin{pmatrix} 3 & -c \\ -c & 7 \end{pmatrix}, \begin{pmatrix} 3 & -2\sqrt{3} \\ -2\sqrt{3} & 7 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix}, \begin{pmatrix} 3 & +2\sqrt{3} \\ +2\sqrt{3} & 7 \end{pmatrix}, \begin{pmatrix} 3 & +c \\ +c & 7 \end{pmatrix}$$



2 変量データは標本をやったようなもの 確率統計☆演習 I(2015)L04

L11-Q3

Quiz(2 次元正規分布)

次の 2 変数確率密度関数は 2 次元正規分布を定める.

$$f(x, y) = C \cdot e^{-x^2 + 4xy - 7y^2}.$$

- ① 母共分散行列を求めよう.
- ② $E[1] = 1$ が満たされるように定数 C を定めよう.

2 次形式の標準化 (線形代数 I)

L11-Q4

Quiz(2 次元正規分布)

次の 2 変数確率密度関数は 2 次元正規分布を定める.

$$f(x, y) = C \cdot e^{-2x^2 - xy - \frac{1}{2}y^2}.$$

- ① 母共分散行列を求めよう.
- ② $E[1] = 1$ が満たされるように定数 C を定めよう.

2 次形式の標準化 (線形代数 I)

お知らせ

- 予習問題と同じタイミングで、「学期途中のリフレクションレポート」をやりましょう. 100 ピーナッツ以外の 3 ピーナッツ.
- 確率統計☆演習 I と同じセッティングで予習問題をやりましょう.
<http://hig3.net> → RaMMoodle
<https://el.math.ryukoku.ac.jp/moodle/> → 確率統計☆演習 II(2016)
- チューター/Math ラウンジ 月火水木昼 1-614



<https://manaba.ryukoku.ac.jp>
マイページの下の方に
manaba 出席カード提出

瀬田龍大生調査プロジェクト

何回かの授業にまたがって、チーム別で、問題 (RQ=Research Question) をたて、調査し、検定して答をだします。

制約

- 指定の検定で答えられるような問題で。
- 母集団=瀬田学舎の龍大生。したがって問題は「瀬田学舎の龍大生の…は…か?」のようになるでしょう。
- 標本=確率統計☆演習 II 参加者。どこかの回で Web で調査します。回答は匿名、問別の無回答の選択肢あり

今日のタスク

- ① シートを受け取る
- ② 教員からのコメント (manaba で見たのと同じ) を確認する
- ③ コメントに応じて 1-4 を修正する
- ④ 紙を提出