

L00 この科目ののり

樋口さぶろお <https://hig3.net>

龍谷大学 先端理工学部 数理・情報科学課程

多変量解析☆演習 L00(2021-09-30 Thu)

最終更新: Time-stamp: "2021-09-29 Wed 20:26 JST hig"

今日の目標

- 科目ののりと成績評価方法を説明できる
- Google Colab で確率変数の標本抽出と可視化ができる



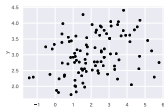
ここまで来たよ

- この科目ののり
 - この授業どんなのり?
- 復習 確率変数
 - 2次元の離散型確率分布と周辺分布
 - 2次元の連続型確率分布と周辺分布

科目の目標

時系列データ, 2次元以上の実数値データと, 問題が与えられたときに, 問題を数学・統計学的に定式化し, 統計手法を選択し, 統計ソフトウェアを使ってデータを分析し, 結論を述べることができる.

- 2次元以上の実数値データ: **多変量解析** = 確率統計☆演習 I, II の, 2次元分布の続きのようなもの. データからモデルの量を推定する (推測統計)



- 時系列データ: **時系列解析** = 確率過程 (時間による確率変数) のようなもの. データからモデルの量を推定する (推測統計)
 - ▶ 2020年の流行語大賞 (うそです) 「移動平均」



- 統計ソフトウェア: Python + Jupyter Notebook

成績評価方法

難しくないですが面倒くさい科目です.

- 平常点 実習課題 45 ピーナツ=3x15 回. 木出題. 水までにファイル提出.
- 毎回の小テスト 40 ピーナツ=8/3x15 回
 - ▶ オンライン期間は, 木出題, 水までに Web で解答. 対面未定.
- その他 15 ピーナツ x 後半に 1 回 データ分析プロジェクトの評価 (プログラムを含むレポートのようなもの)

科目と講時の進行

- 1回目木0 科目の説明, 復習, ツールの説明
- 1回目木1 確率変数の和の分布
- 1回目木2 1回目木0 に対応する実習
- 2回目木1 多次元正規分布
- 2回目木2 1回目木1 に対応する実習
- ⋮

木1はオンライン期間中は、オンデマンド、またはリアルタイム+事後オンデマンド提供

木2はリアルタイム

毎回の指示は Moodle の ex??.php と、Teams の一般、各回 ch で。

教員への連絡は Teams chat で。

匿名質問は相談と質問 ch の匿名質問で。

参考文献

多変量解析☆演習はデータサイエンスプログラムの(最初のほうの)科目
数理では データ分析, 確率統計 I, 多変量解析☆演習, 機械学習, 確率統計
I/II と続く

他課程では下の科目に続く. すみません, 科目の中を詳細に見てはいま
せん.

- 知能: 確率・統計
- 電子 確率および統計・演習
- 機械 確率・統計 I/II
- 応化 アドバンスト数学 III/IV
- 環境 データサイエンス及び演習

参考文献

- 確率統計 I のテキスト 岩薩林 確率・統計
- この科目とばっちり重なる本 永田棟方 多変量解析法入門

ここまで来たよ

- この科目ののり
 - この授業どんなのり?
- 復習 確率変数
 - 2次元の離散型確率分布と周辺分布
 - 2次元の連続型確率分布と周辺分布

岩薩林 確率・統計 §3.3, §4.6

確率統計 I(2021)L04

Quiz(2次元の離散型確率変数の同時分布と周辺分布)

2次元の離散型確率変数 X, Y の同時分布 $p(x, y)$ が次の表で与えられる

$x \backslash y$	1	2	3
4	4/32	8/32	16/32
5	3/32	1/32	0

- 1 周辺分布 $p_X(x), p_Y(y)$ は?
- 2 X, Y は独立か?
- 3 X, Y は同分布にしたがうか?

ここまで来たよ

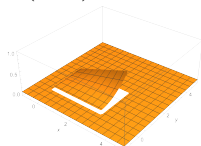
- この科目ののり
 - この授業どんなのり?
- 復習 確率変数
 - 2次元の離散型確率分布と周辺分布
 - 2次元の連続型確率分布と周辺分布

多次元の連続型確率分布

岩薩林 確率・統計 §4.6

2次元の同時確率密度関数 $f(x, y)$.

$f(x, y) \geq 0$, 大きいほど, その (x, y) が「出やすい」



確率 $P(X, Y \text{ の条件}) = E[I_{[X, Y \text{ の条件}]}(X, Y)]$
 $= \iint I_{[X, Y \text{ の条件}]}(x, y) f(x, y) dx dy = \text{その部分の体積.}$

$I_{[X, Y \text{ の条件}]}(x, y)$ は特徴関数. 条件を満たせば 1, 満たさなければ 0

周辺分布

X の周辺分布の確率密度関数 $f_X(x)$ は?

確率 $P(c \leq x < d) = \int_{-\infty}^{+\infty} I_{[c \leq x < d]}(x) f_X(x) dx$ となるべき.

左辺の定義 $\iint_{-\infty}^{+\infty} I_{[c \leq x < d]}(x, y) f(x, y) dy dx$

確率変数の周辺分布

同時分布 $p(x, y)$, $f(x, y)$ に対して,
 X, Y の周辺分布は,

$$\text{離散型 } p_X(x) = \sum_y p(x, y), \quad p_Y(y) = \sum_x p(x, y), \quad (3.11, 3.12)$$

$$\text{連続型 } f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

(4.28, 4.29)

要するに **一方を無視した分布. 小計.**

L00-Q1

Quiz(2次元の連続型確率分布の周辺分布)

連続型確率変数 X, Y の同時確率密度関数 $f(x, y)$ が次で与えられる.

$$f(x, y) = \begin{cases} \frac{1}{6} & (1 \leq x < 7, 1 \leq y < 7) \\ 0 & (\text{他}) \end{cases}$$

- ① X, Y の周辺分布 $f_X(x), f_Y(x)$ は?
- ② X, Y は独立か?
- ③ X, Y は同分布にしたがうか?

L00-Q2

Quiz(2つの連続型確率変数の周辺分布)

連続型確率変数 X, Y は同時分布

$$f(x, y) = \begin{cases} \frac{1}{18} & (0 \leq x < 6, 0 \leq y < 6 - x) \\ 0 & (\text{他}) \end{cases}$$

にしたがう.

- ① X, Y の周辺分布 $f_X(x), f_Y(x)$ は?
- ② X, Y は独立か?
- ③ X, Y は同分布にしたがうか?