多次元の確率変数

樋口さぶろお https://hig3.net

龍谷大学大学院 理工学研究科 数理情報学専攻

理論物理学特論　L01 (2022-09-21 Wed)

最終更新: Time-stamp: "2022-09-21 Wed 12:48 JST hig"

今日の目標

- 1 次元/多次元の確率変数の確率密度関数, 累積分布関数, 分位数を説明できる
- Google Colab で `scipy.stats.分布().pdf(x)`, `.cdf(x)`, `.ppf(p)`, `.rvs(n)` を使える.

ここまで来たよ

科目ののり

計算統計学 (の手法のいくつかは物理から来た)
講義概要 → シラバスブートストラップ法やマルコフ連鎖モンテカルロ法など, 計算機による大量の計算を前提とした統計学の手法群がある. これらは計算統計と総称される. 広く見れば機械学習はこれらの手法群と地続きな場所にある. いくつかの手法について説明し, 実際に計算機を使って試す機会を設ける. このような手法群は特に, 情報科学, 人文科学, 社会科学, 自然科学 (特に実験物理, 計算物理) で取得される大量のデータを扱う際に有用である.
到達目標 → シラバス

- 計算統計学の手法のいくつかについて, 原理を説明することができる
- 計算統計学の手法のいくつかを, 既存のライブラリを使用して, 適切なパラメタを選択し, データに適用することができる
  - ベイズ推定
  - ブートストラップ
  - マルコフ連鎖モンテカルロ法
  - 時系列分析

成績評価

平常点=15 回 x 大小 Quiz 提出

欠席届

Chat で

利用するシステム

毎回 PC 持参でお願いします.

- Microsoft Teams
- 独自 LearnMoodle `https//learn.hig3.net` (Google account Ryukoku で).
- Google Colab 高橋先生の資料 `https://www-tlab.math.ryukoku.ac.jp/wiki/?Data/2022/ex02#colab`
- `https://hig3.net` → 理論物理学特論 (配布資料).

参考書

岩佐-薩摩-林 理工系の数理 確率・統計, 裳華房 (2018)

ここまで来たよ

- はじめに
  - この授業どんなのり?
- 1 多次元の確率変数
  - 1 変数の連続型確率変数
  - 2 変数の連続型確率変数

確率密度関数と累積分布関数 岩薩林 確率・統計 §4.1

X: 実数に値をとる確率変数 (確率空間から実数への関数)
確率 $P(X$ の条件$)$ が考えられる (大部分の条件に対して)

定義 (確率密度関数 (pdf=probability density function))

確率密度関数 $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ があって, 次のように書ける.

$$P(a < X \leq b) = \int_a^b f(x)\,\mathrm{d}x.$$

定義 (母期待値)

X: 連続型確率変数, $g(x)$: 関数に対して, $g(X)$ の母期待値

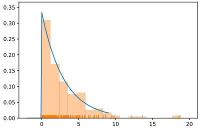
$$\mathrm{E}[g(X)] := \int_{-\infty}^{+\infty} g(x) f(x) \mathrm{d}x.$$

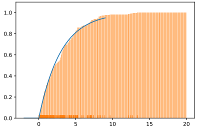
累積分布関数

定義 (累積分布関数 (cdf=cumulative density function))

$$F(x) := \int_{-\infty}^{x} f(x')\,\mathrm{d}x'.$$

指数分布の確率密度関数と累積分布関数





命題 (累積分布関数の性質)

- $F(x)$ は広義単調増加.
- $P(X \le b) = F(b)$.
- $P(a < X \le b) = F(b) - F(a)$
- $F(-\infty) = 0, \quad F(+\infty) = 1.$

経験累積分布関数

定義 (経験累積分布関数 (empirical cumulative density function))

X の標本 $\{x_i\}_{i=1,\ldots,N}$ に対して，

$$\hat{F}_N(x) := \frac{\|\{i|x_i \leq x\}\|}{N}$$

確率密度関数/ヒストグラムのほうが直観的だが (?)，累積分布関数のほうが数学的に扱いやすい

母ナントカ	標本ナントカ
確率密度関数	ヒストグラム
累積分布関数	'標本累積分布関数'=経験累積分布関数

ヒストグラムは階級によるし，$N \to +\infty$ で母に似ない．

経験累積は $N \to +\infty$ で，ある意味，母に収束する．

例: 指数分布 岩薩林 確率・統計 p.80

例 (指数分布)

連続型確率変数 X がパラメタ $\lambda > 0$ に対して確率密度関数

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & (x > 0) \\ 0 & (\text{他}) \end{cases}$$

を持つとき, X はパラメタ λ の指数分布 (パラメタ $n = 1, \lambda$ のガンマ分布) にしたがうという (分布の幅のパラメタ $a = 1/\lambda$ で指定することもある).

命題 (指数分布のモーメント)

$$\mathrm{E}[X^k] = k!\lambda^{-k}, \quad \mathrm{V}[X] = \lambda^{-2}.$$

L01-Q1

Quiz(指数分布の確率)

連続型確率変数 X はパラメタ $\lambda = 2$ の指数分布にしたがう.

1. 確率 $P(X \leq 3)$ を小数で求めよう.
2. 累積分布関数 $F(x)$ を求めよう.
3. 累積分布関数 $F(x)$ の逆関数 $x = F^{-1}(p)$ を求めよう.
4. 確率 $P(3 < X \leq 4)$ を小数で求めよう.
5. 確率 $P(X \geq 4)$ を小数で求めよう.
6. 確率 $P(X \leq x) = 3/4$ となる x を小数で求めよう.

分位数

四分位数 (quartile), 中央値 (median) の一般化.

定義 (分位数 (quantile,percentile))

F が単射とする. 連続型確率変数の p 分位数 $(0 \leq p \leq 1)$ とは,

$$x = F^{-1}(p).$$

すなわち, 与えられた p に対して,

$$P(X \leq x) = p$$

となる x のこと.

F が単射でないときも, $x = \inf\{x | P(X \leq x) = p\}$ などと定めればあいまいさはない. があまり気にかけてはいない.

Python の scipy.stats での確率変数の扱い

loc(ation)=分布の位置, scale=分布の幅
指数分布の確率分布名は expon(loc=0,scale=$1/lambda$)

```
from scipy import stats
# 確率密度関数(連続型用) probability density function
stats.確率分布名(loc=位置, scale=幅).pdf(x)
# 確率(質量)関数(離散型用) probability mass function
stats.確率分布名(loc=位置, scale=幅).pmf(x)
# 累積分布関数 cummulative distribution function
stats.確率分布名(loc=位置, scale=幅).cdf(x)
# 分位数=累積分布関数の逆関数 percent point function
stats.確率分布名(loc=位置, scale=幅).ppf(x)
# サンプル random variable sample
stats.確率分布名(loc=位置, scale=幅).rvs(size=サンプルサイズ)
確率分布にしたがうランダムな数を サンプルサイズ個 取得する.
# 母平均値 mean
stats.確率分布名(loc=位置, scale=幅).mean()
# 母分散 variance
stats.確率分布名(loc=位置, scale=幅).var()
# 母標準偏差 standard deviation
stats.確率分布名(loc=位置, scale=幅).std()
# モーメント xのk乗の母期待値
stats.確率分布名(loc=位置, scale=幅).moment(k)
# xの一般の関数の母期待値
stats.確率分布名(loc=位置, scale=幅).expect(lambda x: f(x))
```

sypy.stats パッケージ, 名前空間
(連続型) 確率分布オブジェクト=uniform, expon, norm, $\cdots$.
(離散型) 確率分布オブジェクト=binom, poisson, $\cdots$.
.pdf, .cdf メンバー関数

ここまで来たよ

1. はじめに
   - この授業どんなのり?
2. 多次元の確率変数
   - 1 変数の連続型確率変数
   - 2 変数の連続型確率変数

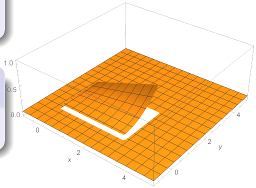
2 次元の連続型確率変数 岩薩林 確率・統計 §4.6

X, Y: 連続型確率変数, $D \subset \mathbb{R}^2$ の, ある性質を満たす部分集合.

定義 (確率密度関数)

次の性質を持つ関数 $f: \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ を確率密度関数という.

$$P((X, Y) \in D) = \iint_D f(x, y) \mathrm{d}x \mathrm{d}y.$$



定義 (同時累積分布関数)

$F(x, y) := P(X \leq x, Y \leq y).$

便利な分位数は定義できない…

定義 (母期待値)

X, Y: 連続型確率変数, $g(x, y)$: 関数に対して, $g(X, Y)$ の母期待値

$$\mathrm{E}[g(X, Y)] := \iint_{\mathbb{R}^2} g(x, y) f(x, y) \mathrm{d}x \, \mathrm{d}y.$$

確率変数の独立性

定義 (独立)

X, Y が独立とは, $f(x,y) = f_{\mathrm{X}}(x) \times f_{\mathrm{Y}}(y)$ と書けること.

命題 (周辺分布)

X, Y が独立なとき,

- $f_{\mathrm{X}}(x), f_{\mathrm{Y}}(y)$ は周辺分布の確率密度関数と同じ
- $$P(a < X \le b \text{かつ} c < Y \le d) = \int_a^b \mathrm{d}x \int_c^d \mathrm{d}y f(x,y)$$
  $$= \int_a^b f_{\mathrm{X}}(x)\mathrm{d}x \times \int_c^d f_{\mathrm{Y}}(y)\mathrm{d}y$$
  $$= P(a < X \le b) \times P(c < Y \le d)$$

L01-Q2

Quiz(2 次元の確率密度関数)

連続型確率変数 X, Y の確率密度関数は次で与えられる.

$$f(x,y) = \begin{cases} 6\mathrm{e}^{-2x-3y} & (x > 0, y > 0) \\ 0 & (\text{他}) \end{cases}$$

1. 確率 $P(X \leq 1 \text{または} Y \leq 5)$ を求めよう.
2. 確率 $P(Y \leq 5X)$ を求めよう.
3. 母期待値 $\mathrm{E}[3 + \mathrm{e}^{x+y}]$ を求めよう.

L01-Q3

Quiz(独立と限らない多次元の連続型確率変数)

連続型確率変数 X, Y の同時確率密度関数は

$$f(x,y) = \begin{cases} \frac{1}{24}(2+xy^2) & (0 \leq x \leq 3, 0 \leq y \leq 2) \\ 0 & (\text{他}) \end{cases}$$

で与えられる.

1. 母期待値 $\mathrm{E}[X^3Y]$ を求めよう.
2. 確率 $P(X \leq 2$ かつ $Y \leq 1)$ を求めよう.