離散型確率変数

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

確率統計 I L01(2023-04-10 Mon)

最終更新: Time-stamp: "2023-04-10 Mon 11:20 JST hig"

今日の目標

- 授業の到達目標/合格条件を説明できる
- 離散型確率変数とは何か説明できる 岩薩林 確率・統計 §3.1
- 離散型確率変数の確率, 母平均値, 母分散, 母期待値が計算できる 岩薩林 確率・統計 §3.2

ここまで来たよ

1. はじめに
   - この授業どんなのり?
2. 離散型確率変数
   - 離散型確率変数
   - 母期待値・母平均値・母分散・母標準偏差
   - 事象と確率

学習目標

講義概要 → シラバス
現実世界の現象を理解し、数理モデルとの関係を明らかにするためには、観察・実験により取得したデータを整理・解析することが必要である。限ら れたデータから数理モデルのパラメタを推測する推測統計と、必要な確率論を説明する。
到達目標 → シラバス

- 確率論:1 変数、2 変数の離散型、連続型確率変数の期待値や確率の計算ができる → 1Q 各回のトライアル
- 推測統計: 実験・観察により取得したデータから数理モデルのパラメタを推測して、根拠とともに他者に説明ができる → 2Q 各回のトライアル

きょうは 1 変数離散型

確率統計 I を履修してはいけない理由

次のどれも響かない人は履修しないことを奨めます.

- データサイエンスプログラムの前提科目 (データ分析, 確率統計 I, 多変量解析及び実習, 機械学習 I,II, 確率統計 II,III, 確率モデル及び実習)
- 数学の教員免許の必修科目
  - 高校 数学 I(データの分析)=毎年センター試験に出題, 高校 数学 A(場合の数と確率), 高校 数学 B(確率分布と統計的推測)(選択)
  - 高校の高校 数学 I に統計的仮説検定が来る
- いま, データサイエンス, 統計が熱い!
- いま, 機械学習, 人工知能 (AI) が熱い! 線形代数と確率が必要
- 統計は科学技術の言葉 $\rightsquigarrow$ 数理卒は当然期待されてる
- 科目に合格したら統計検定 2 級の 6 合目くらいまで来た?

こんなことに答えます

1. 「データ分析」で伏線はりまくったけど，どこで回収するの? 推測統計 (確率統計)↔ 記述統計 (データ分析)
2. (ゲーム運営) このガチャの確率の設定で，プレイヤーのレベルってどのくらいあがる?
3. YouTube から猫の動画を見つけるアルゴリズム，こう改良して，100 個の入力画像で試したら，判定精度が 3 個分あがった．これたまたま? 10000 個でやり直すべき?
4. 秋元 P は日向坂に櫻坂より身長高いメンバーをいれてる説を唱えたけどみんな信じてくれない…どうやって説得する?

(再掲) 到達目標 → シラバス

- 確率論:1 変数、2 変数の離散型、連続型確率変数の期待値や確率の計算ができる → 1Q 各回のトライアル
- 推測統計: 実験・観察により取得したデータから数理モデルのパラメタを推測して、根拠とともに他者に説明ができる → 2Q 各回のトライアル

ピーナッツカウント (成績計算) ののり

成績計算
難しくないけどとにかく注文の多い科目です…
科目の成績 100 ピーナッツは

- 25 ピーナッツ: 平常点 毎週の Web の練習問題, チーム課題, 授業時間内の活動, それほどたいへんじゃないレポートなど
- 75 ピーナッツ: 小テスト 毎週 9:15 からのトライアル=5 ピーナッツ x15 回

~~(任意の受験で) 統計検定 3 級に合格した人は,~~
~~小テスト Q1,Q2 各 30 ピーナッツ (以上) とみなします.~~ → 2023 年度は換算は行いません.

75 ピーナッツ: 小テストのうち, 1Q(40) 2Q(35) どちらかが, 50%未満のときは無条件に科目を不合格とします (2 つの到達目標のうち 1 個しか達成していないから)

欠席届 毎回出席を前提に進めます. 欠席に事前連絡は不要. やむを得ず欠席して, ピーナッツ的に考慮されたい場合は事後 2 週間以内に hig3Moodle に届けてください

週サイクルののり

説明—時間内 (チーム) 課題—週内 Web 練習問題—(翌週) トライアル
今週を例に.

- 2023-04-10 月 対面授業 来週のトライアルを予告, チーム課題で練習
- 2023-04-10 火 9:20 ごろ–2023-04-14 金 20:00 hig3Moodle の練習問題 2022 と違います
- 2023-04-12 火 4,14 木昼 オフィスアワー 1-507
- 2022-04-17 月 09:15(10 分間くらい) トライアル 非参照 非相談

数学系の 2 年次以上科目は週 1 コマしかない!

(週 1 コマ相当) 自分で演習問題を見つけて, 自分で解き, わからない点があったら自分から相談したり質問したりする

授業ののり (教科書やその他の準備)

https://hig3.net

- → 確率統計 I (配布資料).
- → hig3Moodle https://moodle.hig3.net (Google でログイン, 1 年生用, 高橋先生とは別) → 確率統計 I
  - → Teams チームコード 4232fje

https://hig3.net

教科書 必須です. 岩薩林 確率・統計 (3.2) は式 (3.2) ということ.

岩佐-薩摩-林 理工系の数理 確率・統計, 裳華房 (2018)

座席指定
チーム別エリア座席指定．チームはプロジェクト演習と同じ．
ノート
教科書や紙配布資料に書き込む + 自分で問題を解いた過程をノートやルーズリーフに残す，ことをお奨めします．
ノート PC ノート PC とイヤフォンを毎回持参していただきます．
担当者ののり

- なまえ: 樋口さぶろお hig@math.ryukoku.ac.jp メールはいつでも．
- へや: 1-507
- Web ページ．https://hig3.net
- オフィスアワー 前期火 4 木昼，1-507 or Teams chat a00010

相談できるところ

- Math ラウンジ (1 号館 5 階 1-536,538)，昼休みはだいたい大学院生常駐．数理 TM-Math ラウンジ ch on Teams.

ここまで来たよ

高校の確率

文章から確率を求める問題 高校 数学 A

トランプを 1 枚, 同じ確からしさで引く.

結果	確率
$\heartsuit 1$	$\frac{1}{52}$
$\heartsuit 2$	$\frac{1}{52}$
$\vdots$	$\vdots$
$\spadesuit 13$	$\frac{1}{52}$
計	1

偶数のカードのでる確率は? $\frac{24}{52}$.

離散型確率変数 岩薩林 確率・統計 §3.1

高校数学でよく見る確率の問題 高校 数学 A

袋に赤玉 2 個, 白玉 3 個がはいっている. いちどに 3 個取り出したとき, 赤玉が x 個である確率は ?

X は離散型 の確率変数 離散型 $\approx$ 整数値
事象は x の集合 $\{0, 1, 2, 3\}$ の部分集合.

x_k	確率 $p_k = p(x_k) = P(X = x_k)$
$\vdots$	0
-1	0
0	$\frac{1}{10} = \frac{1}{{}_5\mathrm{C}_3}$
1	$\frac{6}{10} = \frac{2\cdot 3}{{}_5\mathrm{C}_3}$
2	$\frac{3}{10} = \frac{1\cdot 3}{{}_5\mathrm{C}_3}$
3	0
$\vdots$	0
計	1

言葉 確率分布 (確率関数)

岩薩林 確率・統計 (3.1)p.49

$$p(x) = \begin{cases} \frac{1}{10} & (x = 0) \\ \frac{6}{10} & (x = 1) \\ \frac{3}{10} & (x = 2) \\ 0 & (\text{他}) \end{cases}$$

$p_k = p(x_k)$.

確率分布の性質

$0 \le p(x) \le 1$. $\sum_k p(x_k) = 1$.

高校と大学の, 確率の問題の違い

高校 数学 A ではこの表を作るまでを考える

高校 数学 B, 確率統計 I(*)L* ではこの表ができて与えられた後を考える. 'この表のとき, 赤玉の個数の

keyword 母期待値は?'

ここまで来たよ

- はじめに
  - この授業どんなのり?

1. 離散型確率変数
   - 離散型確率変数
   - 母期待値・母平均値・母分散・母標準偏差
   - 事象と確率

関数 $g(x)$ の母期待値 $\mathrm{E}[g(X)]$ 高校 数学 AB 岩薩林 確率・統計 §3.2(3.3)p.52

定義 (関数 $g(x)$ の母期待値)

離散型確率変数 X が確率分布 $p(x) = \cdots$ に従うとき,

$$\text{関数 } g(x) \text{ の母期待値} \quad \mathrm{E}[g(X)] = \sum_x g(x) \times p(x)$$

g は普通の関数. 例: $g(x) = x^2, \mathrm{e}^x$, (場合分けで書かれた関数), ...

命題 (母期待値の性質)

$\mathrm{E}[1] = 1$. ($g(x) = 1$ と $\sum_k p(x_k) = 1$ から)

定義 (母平均値, 母分散, 母標準偏差)

- X の母平均値 $\mu \overset{\text{定義}}{=} \mathrm{E}[X]$. ($g(x) = x$ ってこと). (E0) 岩薩林 確率・統計 (3.3) (X の) 母期待値ともいう.
- X の 母分散 $\mathrm{V}[X] \overset{\text{定義}}{=} \mathrm{E}[(X-\mu)^2]$. (V0) 岩薩林 確率・統計 (3.7) ($g(x) = (x-\mu)^2$ の母期待値)
- X の母標準偏差 $\overset{\text{定義}}{=} \sqrt{\mathrm{V}[X]}$ 岩薩林 確率・統計 (3.7)

「母」平均値で データ分析 の「標本」平均値 岩薩林 確率・統計 §3.2(p.54) と区別

ここまで来たよ

- はじめに
  - この授業どんなのり?

1. 離散型確率変数
   - 離散型確率変数
   - 母期待値・母平均値・母分散・母標準偏差
   - 事象と確率

Ω =(トランプのカード全体の集合), 部分集合 $A = \{x \in \Omega | a(x)\} \subset \Omega$.

定義 (事象の確率)

「事象 A の確率」$= P(A) = P(a(X)) =$「条件 $a(X)$ が成立する確率」

- (♡がでるという事象の確率) $= P(\{\heartsuit 1, \ldots, \heartsuit K\}) = P(X$が♡$)$
- (♡1がでるという事象の確率) $= P(\{\heartsuit 1\}) = P(X$が♡1$)$
- (黒札がでるという事象の確率) $= P(\{\clubsuit 1, \ldots, \clubsuit K, \spadesuit 1, \ldots, \spadesuit K\}) = P(X$が黒札$)$

Ω が有限集合で, 等しく確からしい, なら,

$P(A) = \dfrac{\text{場合の数}}{\text{すべての場合の数}} = \dfrac{A\text{の要素の個数}}{\Omega\text{の要素の個数}}$. 高校 数学 A

例 $\Omega = \{0, 1, 2, 3\}$ のとき,

- $P(\{0, 2\}) = P(X$ は偶数$)$
- $P(\{3, 4\}) = P(X$ は 3 以上$)$
- $P(\{2\}) = P(X = 2)$

事象の確率 岩薩林 確率・統計 なし

A: 事象 (Ω の部分集合)
$a(X)$: X の条件

定義 (事象の確率)

$$P(A) = P(a(X)) = \mathrm{E}[\mathrm{I}_{[a(X)]}(X)] \quad \text{(P1)}$$

ここで,

定義 (特徴関数)

$$\text{特徴関数}\mathrm{I}_{[a(X)]}(x) = \begin{cases} 1 & (a(x)\text{ が真}) \\ 0 & (a(x)\text{ が偽}) \end{cases}$$

$\boxed{\text{例}}$ $x \in \mathbb{Z}$ のとき,

$$\mathrm{I}_{[X^2 \leq 5]}(x) = \begin{cases} 1 & (x = -2, -1, 0, 1, 2) \\ 0 & (\text{他}) \end{cases}$$

L01-Q1

Quiz(離散的な確率変数の母平均・母分散・母標準偏差)

整数に値をとる離散型確率変数 X は次の確率分布に従う.

$$f(x) = \begin{cases} \frac{4}{12} & (x = -1) \\ \frac{5}{12} & (x = 0) \\ \frac{3}{12} & (x = 2) \\ 0 & (\text{他}) \end{cases}$$

1. 母期待値 $\mathrm{E}[\mathrm{e}^X]$ を求めよう.
2. X の母平均値を求めよう.
3. X の母分散を求めよう.
4. X の母標準偏差を求めよう.
5. 事象 $X \leq 1$ の確率を求めよう.

来週: 練習問題→ Trial. 教科書備えて. データ分析 Google Colab 思い出しておいて.