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(2021-04-07 Wed)

最終更新: Time-stamp: "2021-04-12 Mon 08:49 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

- 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

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 → テスト A
- 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 → テスト B

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 (多変量解析及び実習, 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
  - 高校の高校 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! 生成モデルの芯
- 統計は科学技術の言葉 $\rightsquigarrow$ 数理卒は当然期待されてる
- 統計検定 3 級

https://www.amazon.co.jp

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 → テスト A
- 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 → テスト B

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

- 30 ピーナッツ:平常点 毎週の練習問題–トライアル問題–リベンジ問題, 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
- 35 ピーナッツ:小テスト=テスト A (学期半ば, 学期終わりに 2 回実施して高いほうをとります)
- 35 ピーナッツ:定期試験=テスト B (学期終わり, 定期試験期間に 2 回実施して高いほうをとります)

統計検定 3 級に合格した人は, 小テスト, 期末試験をそれぞれ 35 ピーナッツとみなします.

テスト A,B どちらかが, 17.5 ピーナッツ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. 何回欠席しても定期試験参加資格を失うことはありません. 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—練習問題—トライアル問題 (—リベンジ問題)
今週を例に.

- 2021-04-07 水 対面授業 来週のトライアルを予告, 時間チーム課題で練習することも
- 2021-04-08 木 9:20(以前)–2021-04-14 水 13:30 hig3Moodle の練習問題
  - これのいずれかが 0 点の場合は, トライアル問題, リベンジ問題は加算されません
- 2021-04-14 水 13:30(10 分くらい) トライアル問題 非参照
- 2021-04-19 月ごろ トライアル採点答案返却
- – 2021-04-21 水 13:30 hig3Moodle のリベンジ問題
  - トライアル問題の満点の半分までリカバーできます T
  - トライアル問題の Trial の点数が低かった人向け

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).
- → hig3Moodlehttps://moodle.hig3.net (Google でログイン, 高橋先生とは別) → 確率統計 I
  - → Teams

https://hig3.net

教科書 必須です. 岩薩林 確率・統計

岩佐-薩摩-林 理工系の数理 確率・統計, 裳華房 (2018) テスト A,B は公式集のページのコピーを持込可.

座席指定

来週から独自の座席指定する予定. Google Form に記録して

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 y20 はノート PC とイヤフォンを毎回持参していただく予定.
Excel は龍大の Office365 で無料でインストール. 中野先生の案内

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 前期木 5, 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), 昼休みは大学院生常駐), 数理 TM-Math ラウンジ ch on Teams
- BYOD PC サポート (1 号館 4 階 1-443), オンラインサポート TM-BYOD PC サポート窓口 ch on Teams

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

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

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

高校の確率の問題と, 大学のこの科目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 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, 赤玉の個数の母期待値は?’

ここまで来たよ

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

関数 $g(x)$ の母期待値 $\mathrm{E}[g(X)]$

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

$$\mathrm{E}[g(X)] = \sum_x g(x) \times p(x)$$

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), ...

性質

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

特に名前のついた量 (「母」で データ分析 と区別) 岩薩林 確率・統計 §3.2(p.54)

- X の母平均値. $\mu \overset{\text{定義}}{=} \mathrm{E}[X]$. ($g(x) = x$ ってこと). (E0) (X の) 母期待値ともいう.
- X の 母分散は, ($\mathrm{E}[X] = \mu$ とおくとき), $\mathrm{V}[X] \overset{\text{定義}}{=} \mathrm{E}[(X - \mu)^2]$. (V0) ($g(x) = (x - \mu)^2$ の母期待値)
- X の母標準偏差 $\overset{\text{定義}}{=} \sqrt{\mathrm{V}[X]}$

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.

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

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」

$\Omega =$(トランプのカード全体の集合) のとき,

- ($\heartsuit$がでるという事象の確率) $= P(\{\heartsuit 1, \ldots, \heartsuit K\}) = P(X$ が$\heartsuit)$
- ($\heartsuit 1$ がでるという事象の確率) $= P(\{\heartsuit 1\}) = P(X$ が$\heartsuit 1)$
- (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 は $\dfrac{\text{場合の数}}{\text{すべての場合の数}}$ 高校 数学 A

$\Omega = \{0, 1, 2, 3\}$ 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

$A = \{x | a(x)\}$: 事象 (部分集合)
$a(x)$: x が A に属する条件

事象の確率

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

ここで,

特徴関数

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

例 $x \in \mathbb{Z}$ のとき,

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

計算を楽にする母分散の性質

母期待値 $\mathrm{E}[X^2]$ から母分散 $\mathrm{V}[X]$, および逆方向に計算するのに使う.

母分散の性質 高校 数学 B $\boxed{\text{岩薩林 確率・統計 (3.8)p.55}}$

$$\mathrm{V}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \qquad \text{(V1,(3.8))}$$

証明

$$\begin{aligned}
\mathrm{V}[X] = \mathrm{E}[(X-\mu)^2] &= \sum_k p(x_k)(x_k-\mu)^2 \\
&= \sum_k p(x_k)(x_k^2 - 2\mu x_k + \mu^2) \\
&= \sum_k p(x_k)(x_k^2 - 2\mu \sum_k p(x_k)x_k + \mu^2 \sum_k p(x_k) \\
&= \mathrm{E}[X^2] - 2\mu \mathrm{E}[X] + \mu^2 = \mathrm{E}[X^2] - \mathrm{E}[X]^2
\end{aligned}$$

教科書の例題

L01-Q2

岩薩林 確率・統計 例 6(p.52), 例題 3.1(p.53)

L01-Q3

岩薩林 確率・統計 第 3 章演習問題 1

数学に行きたい人に語れなくて残念なこと

- 確率の公理 岩薩林 確率・統計 §2.2
- 確率に関する基本的定理 岩薩林 確率・統計 §2.2

この授業ではやらないこと 測度 (measure), 測度空間, dx $\rightsquigarrow$ ルベーグ積分