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(2025-04-14 Mon)

最終更新: Time-stamp: "2025-04-14 Mon 17:52 JST hig"

今日の目標

- 科目の目標/合格条件を説明できる, Moodle 使える
- 久保川 統計学入門 §4.1 離散型確率変数とは何か説明できる
- 久保川 統計学入門 §4.2 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

ここまで来たよ

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 trial
- 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 trial

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, 新課程高校 数学 A(場合の数と確率), 新課程高校 数学 C(確率分布と統計的推測)
  - 高校の新課程高校 数学 I に統計的仮説検定が来てる
- いま, データサイエンス, 統計が熱い!
- いま, 生成 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 trial
- 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 trial

ピーナツカウント (成績計算) ののり

成績計算
ムズくないけどだるい科目です
科目の成績 100 ピーナツは

- 25 ピーナツ: 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 ピーナツ: 小テスト 毎週 9:15 からの trial=5 ピーナツ x15 回

75 ピーナツ: 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, ピーナツ的に考慮されたい場合は事後 2 週間以内に LearnMoodle 欠席届から届けてください

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)trial
今週を例に.

- 2025-04-14 月 対面授業 来週の trial を予告, チーム課題で練習
- 2025-04-15 火 9:20 ごろ –2025-04-18 金 20:00 LearnMoodle 練習問題
- 2025-04-17 木昼 オフィスアワー 1-539
- 2025-04-21 月 09:15(10 分間くらい) trial 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).
- → LearnMoodle https://learn.hig3.net (Google でログイン) → 確率統計 I
  - → Teams

教科書 必須です. 久保川 統計学入門 (3.2) は式 (3.2) ということ.
久保川達也 公式と例題で学ぶ 統計学入門, 共立出版 (2024)
演習解答・正誤表

https://hig3.net

§1,2: 記述統計. データ分析 相当.
§3.1, §3.2: 大事だけど今日圧縮してやります
§3.3, §3.4: 必要な時に戻ってきます
§4–9,12.1,12.2: とばすところもあるけどこの科目でカバー

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はず.
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 Google Colab や動画を使います. 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
- へや: 1-507
- Web ページ. `https://hig3.net`
- オフィスアワー 前木昼, 1-539 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

- はじめに
  - この授業どんなのり?

1. 離散型確率変数
   - 事象と確率
   - 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 久保川 統計学入門 §3.1

確率は集合位相の言葉で語られる. 集合位相 (2 年後期)

定義

全事象 Ω =(トランプのカード全体の集合)

事象 $A = \{x \in \Omega | a(x)\} \subset \Omega$ (例えば赤のカード全体からなる部分集合)

基本事象 $x \in \Omega$ のとき, $A = \{x\}$ のような元 1 個の集合 ($\{\heartsuit 1\}$)

試行 基本事象のひとつを選ぶこと (トランプからカード 1 枚引くこと)

全事象 $\Omega \subset \Omega$

空事象 $\emptyset \subset \Omega$

補事象 $A^c = \Omega \setminus A$. A が起きないという事象.

積事象 $A \cap B$ ‘かつ’.

和事象 $A \cup B$ ‘または’.

排反事象「A, B が排反事象」$\Leftrightarrow A \cap B = \emptyset$. 同時に起きない.

定義 (確率の公理 久保川 統計学入門 定義 3.1)

確率 P は，Ω の部分集合を，0 以上 1 以下の実数に対応させる関数で，次をみたす．

- (P1) すべての $A \subset \Omega$ に対して $0 \leq P(A) \leq 1$.
- (P2) $P(\Omega) = 1$.
- (P3) $A \cap B = \emptyset$ (排反) なら，$P(A \cap B) = P(A) + P(B)$.

「事象 A の確率」$=P(A)$ を，P(条件) のように書くことがある．
全事象 $\Omega =$(トランプのカード全体の集合) のとき，

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

事象と確率

公理から導ける公式

全事象 $\Omega \subset \Omega$. $P(\Omega) = 1$.

空事象 $\emptyset \subset \Omega$. $P(\emptyset) = 0$.

補事象 $A^c = \Omega \setminus A$. A が起きないという事象. $P(A^c) = 1 - P(A)$.

積事象 $A \cap B$ ‘かつ’.

和事象 $A \cup B$ ‘または’.

排反事象「A, B が排反事象」$\Leftrightarrow A \cap B = \emptyset$. 同時に起きない. A, B が排反事象ならば, $P(A \cup B) = P(A) + P(B)$.

命題 (同様に確からしいときの確率 高校 数学 A 久保川 統計学入門 公式 3.4)

Ω が有限集合で, 基本事象が同様に確からしい, なら, 確率は

$$P(A) = \frac{A\,の要素の個数}{\Omega\,の要素の個数} = \frac{場合の数}{すべての場合の数}$$

確率変数

コイン投げ $\Omega = \{\text{裏}, \text{表}\}$, $P(\{\text{裏}\}) = p$ を考える．
裏が出たら $X = 0$, 表が出たら $X = 1$ となるような「ランダムな」変数 X を考えたい．
この X は，全事象から実数への関数とみなせる．
$X : \{\text{裏}, \text{表}\} \to \mathbb{R}$
このような関数を確率変数という．
$X(\text{裏}) = 0, X(\text{表}) = 1$ を確率変数の実現値 という．
大胆にも，$X = 0, X = 1$ のような略記がされる．
$X = 1$ となる確率 $P(X = 1)$ は，$P(\{\text{表}\})$ に等しいと考える．
直観的な説明
数学の確率論では，上の考え方（をさらに精密化）するが，応用上は次のように思っても困らない．

- 「全事象が $\Omega = \mathbb{R}$ や $\mathbb{R}^n$ であるような事象を確率変数という」
- 「$X = 0$ となる確率, $X = 1$ となる確率…が備わっているランダムな変数 X を確率変数という」

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 久保川 統計学入門 §4.1

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 $\approx$ 実現値が可算個（例えば整数全体）全事象が整数全体と思える.

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, 確率質量関数)

久保川 統計学入門 §4.1

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

$p_k = p(x_k)$.

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, 赤玉の個数の母期待値は?’

確率 (質量) 関数と累積確率分布関数

定義 (確率 (質量) 関数と累積確率分布関数 久保川 統計学入門 定義 4.1)

確率変数 X の実現値が $x_1, x_2, \ldots$ であるとき，$p(x_i) = P(X = x_i)$ を確率 (質量) 関数 (probability mass function) という．
$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$ を累積分布関数 (cumurative distribution function) という．

確率質量関数に対して，$0 \leq p(x) \leq 1$. $\displaystyle\sum_i p(x_i) = 1$ が成り立つ．

累積分布関数の性質は来週．

久保川 統計学入門 図 4.1

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

関数 $g(x)$ の母期待値 $\mathrm{E}[g(X)]$ 高校 数学 AB 久保川 統計学入門 §4.2

定義 (関数 $g(x)$ の母期待値 久保川 統計学入門 公式 4.4)

離散型確率変数 X が確率関数 $p(x) = \cdots$ を持つとき,

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

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)

- 久保川 統計学入門 定義 4.2 X の母平均値 $\mu \overset{\text{定義}}{=} \mathrm{E}[X]$. ($g(x) = x$ ってこと). (X の) 母期待値ともいう.
- 久保川 統計学入門 定義 4.3 X の 母分散 $\mathrm{Var}(X) \overset{\text{定義}}{=} \mathrm{E}[(X-\mu)^2]$. ($g(x) = (x-\mu)^2$ の母期待値)
- X の母標準偏差 $\overset{\text{定義}}{=} \sqrt{\mathrm{Var}(X)}$

「母」平均値で データ分析 の「標本」平均値 久保川 統計学入門 §1.2 と区別

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 は次の確率関数を持つ.

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

1. 母期待値 $\mathrm{E}[3 + 12\mathrm{e}^X]$ 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.

久保川 統計学入門 問 1–4(p.100)

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.

幾何分布 久保川 統計学入門 §4.3.4

L01-Q2

Quiz(離散的な確率変数の累積分布関数)

整数に値をとる離散型確率変数 X が次の確率関数を持つ．この X のしたがう分布をパラメタ p の幾何分布 $Geo(p)$ という．

$$p(x) = \begin{cases} (1-p)^x p & (x = 0, 1, 2, \dots) \\ 0 & (\text{他}) \end{cases}$$

1. 累積分布関数 $F(x)$ を，整数 x に対して求めよう．