離散型確率変数

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 I L05(2015-10-16 Fri)

最終更新: Time-stamp: "2015-10-16 Fri 09:56 JST hig"

今日の目標

- 離散確率分布が与えられたときに, 母期待値・母平均値・母分散・母標準偏差が計算できる
- 離散確率分布が与えられたときに, 事象の確率が計算できる

http://hig3.net

L04 p.6 の標準得点の例の訂正

例 $n = 5$

i	1	2	3	4	5	平均値	標準偏差
データ x_i	15	13	12	11	9	12	2
標準得点 z_i	1.50	0.5	0	-0.5	-1.50	0	1

L04-Q1

Quiz 解答:平均値・分散・標準偏差の換算

1.6m, 0.0025m^2, 0.05m.

L04-Q3

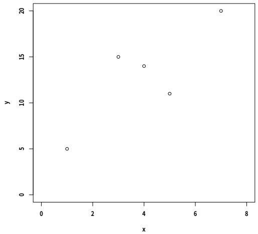
Quiz 解答:標準得点と偏差値　平均値 $\overline{x} = 90$, 分散 $s_x^2 = 4$, 標準偏差 $s_x = 2$.

標準得点 $z = (87 - 90)/2 = -1.5$.

偏差値 $w = (-1.5) \times 10 + 50 = 35$.

L04-Q4

Quiz 解答:クロス集計表



$y \backslash x$	0 以上 2 未満	2 以上 4 未満	4 以上 6 未満	6 以上 8 未満	計
0 以上 5 未満					1
5 以上 10 未満	1				1
10 以上 15 未満		1	1		2
15 以上 20 未満		1			1
20 以上 25 未満				1	1
計	1	2	1	1	5

L04-Q5

Quiz 解答:共分散

$\overline{x} = 4,\ s_x^2 = 4,\ s_x = 2.$
$\overline{y} = 13,\ s_x^2 = 122/5 = 24.4,\ s_y = \sqrt{122/5} = 4.94.$
共分散 $C = \frac{1}{5}[(1-4)(5-13) + (3-4)(15-13) + (4-4)(14-13) + (5-4)(11-13) + (7-4)(20-13)] = 41/5 = 8.2.$
相関係数 $r = \frac{41/5}{2\cdot\sqrt{122/5}} = 0.83.$

L04-Q7

Quiz 解答:共分散と相関係数

1. x の平均値は $\overline{x} = 18$cm, y の平均値は $\overline{y} = 4$g.
   共分散は
   $$C_{xy} = \frac{16}{5}\mathrm{cm}\cdot\mathrm{g}.$$
2. x の分散は $s_x^2 = 9$ cm^2, y の分散は $s_y^2 = 4$ g^2. よって,
   $$r = \frac{16/5\mathrm{cm}\cdot\mathrm{g}}{\sqrt{9\mathrm{cm}^2}\sqrt{4\mathrm{g}^2}} = \frac{8}{15}.$$

コース全体の計画

1. データの表現=記述統計 (データの分布)
2. データの表現=記述統計 (データの代表値)
3. データの表現=記述統計 (データのばらつきを表す量, 箱ひげ図)
4. データの表現=記述統計 (2 変量の共分散と相関係数)
5. 確率論 (離散値確率変数) **今日**
6. データの表現=記述統計 (回帰分析, 統計ソフトウェアの使用) **撮影**
7. 確率論 (連続値確率変数)
8. プチテスト
9. 確率論 (同時分布, 独立, 正規分布)
10. 確率論 (大数の法則と中心極限定理)
11. データからの確率分布の推定=推測統計 (母平均値の点推定)
12. データからの確率分布の推定=推測統計 (母平均値の区間推定)
13. データからの確率分布の推定=推測統計 (母平均値に関する検定)
14. データからの確率分布の推定=推測統計 (母分散に関する推定と検定)
15. データからの確率分布の推定=推測統計 (母比率に関する推定と検定)

ここまで来たよ

3. 2 変量データ
4. 離散型確率変数
   - 確率分布
   - 母期待値・母平均値・母分散・母標準偏差

高校数学でありがちな設定

コインを 1 回投げる

結果	確率
表	$\frac{1}{2}$
裏	$\frac{1}{2}$
計	1

前回までの話 (記述統計) との関係.
$\{$ 表, 裏 $\} = \{$ 高橋みなみ, 渡辺麻友, $\cdots\}$ ではない. とりあえず無関係な別の話だと思って.

アイドル作成ゲームで, 新しいメンバーをスカウトする ボタンを押したら, CPU 内部でサイコロが振られて (=確率) 身長体重が決まって…を 77 回繰りかえしたら, 77 個からなる 2 変量データができた, みたいな関係.
推測統計まで行ったときに明らかになります.

高校数学でありがちな問題

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

x	確率 f_x
$\vdots$	0
-1	0
0	$\frac{1}{10} = \frac{1}{{}_5\mathrm{C}_3}$
1	$\frac{6}{10} = 2 \cdot 3/{}_5\mathrm{C}_3$
2	$\frac{3}{10} = 1 \cdot 3/{}_5\mathrm{C}_3$
3	0
$\vdots$	0
計	1

言葉

X は離散型確率変数 離散型 $\approx$ 整数値

確率分布 (確率関数)

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

確率分布の性質 $0 \le f_x \le 1$. $\sum_x f_x = 1$.

事象と確率

X:離散型確率変数

ここでは, 離散型確率変数 1 個という限られた範囲で確率論を展開しています.

本来は, 事象が基本で, そこを定義域とする関数として確率変数を後から考えます.

事象 集合 $A = \{x \in \mathbb{Z} |$ 条件 $a(x)$ が成立 $\}$ のこと.

- $\{-2, -1, 0, 1, 2\} = \{x \in \mathbb{Z} | x^2 \leq 4\}$.
- $\{3\} = \{x \in \mathbb{Z} | x = 3\}$
- $\{2, 3, 5, 7, 11, \ldots\} = \{x \in \mathbb{Z} | x$ は素数 $\}$

全事象 $U = \mathbb{Z}$.
空事象 $\emptyset$
基本事象 $A = \{x_1\}$. それ以上分けられない

以下は当面高校の知識で

補事象 $A^c = U \setminus A$.　A が起きなかったという事象.
和事象 $A \cup B$ または,
積事象 $A \cap B$ かつ,
排反事象 「A, B が排反事象」 $\Leftrightarrow A \cap B = \emptyset$. 同時に起きない

事象の確率

確率分布 f_x を持つ確率変数 X に対して,
「事象 A の確率」 $=$ 「条件 $a(X)$ が成立する確率」
$=P(A)=P(a(X))$

- $P(\{-2,-1,0,1,2\}) = P(X^2 \leq 4) = (X^2$ が 4 以下になる確率$)$
- $P(X=3) = (X=3$ となる確率$)$
- $P(X$ は素数$) = (X$ が素数となる確率$)$

性質

$$P(X = x_1) = f_{x_1}.$$

$0 \leq P(A) \leq 1$ $\qquad$ $0 \leq f_x \leq 1$

$P(U) = 1$ $\qquad$ $\sum_x f_x = 1$

$P(\emptyset) = 0$

$P(A^c) = 1 - P(A)$

A,B が排反事象のとき $P(A \cup B) = P(A) + P(B)$

ここまで来たよ

3 2 変量データ

4 離散型確率変数

- 確率分布
- 母期待値・母平均値・母分散・母標準偏差

母期待値

関数 $\phi(x)$ の母期待値

確率変数 X が確率分布 $f_x = \cdots$ に従うとき

$$\mathrm{E}[\phi(X)] = \sum_x f_x \times \phi(x)$$

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

性質

$\mathrm{E}[1] = 1$. ($\phi(x) = 1$ と $\sum_x f_x = 1$ から)

特に名前のついた量

- 母平均値 $m = \mathrm{E}[X]$.　($\phi(x) = x$ってこと)
- 母分散 $= V[X] = \mathrm{E}[(X - m)^2]$.　($\phi(x) = (x - m)^2$ってこと)
- 母標準偏差 $= \sqrt{V[X]}$

L05-Q1

Quiz(離散的な確率変数の母平均・母分散・母標準偏差)

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.

事象の確率

事象 A の確率 $\Leftrightarrow$ 条件 $a(X)$ が成立する確率

特徴関数

$$\text{関数 } \mathbf{1}_{[\mathsf{a}]}(x) = \begin{cases} 1 & (a(x) \text{ が真}) \\ 0 & (a(x) \text{ が偽}) \end{cases}$$

とすると，

$$P(A) = P(a(X)) = \mathrm{E}[\mathbf{1}_{[a]}(X)]$$

例

$$\mathbf{1}_{[X^2 \leq 4]}(x) = \begin{cases} 1 & (-2 \leq x \leq 2) \\ 0 & (\text{他}) \end{cases}$$

母平均値，母分散の性質

母平均値の性質

X: 確率変数, $a, b \in \mathbb{R}$:定数 のとき,

$$\begin{aligned} \mathrm{E}[aX+b] &= \sum_x f_x \times (ax+b) \\ &= \left(a \sum_x f_x x \right) + b = a\mathrm{E}[X] + b. \\ \mathrm{E}[\phi_1(X)+\phi_2(X)] &= \sum_x f_x \times (\phi_1(X)+\phi_2(X)) \\ &= \mathrm{E}[\phi_1(X)] + \mathrm{E}[\phi_2(X)]. \end{aligned}$$

もちろん一般には $\mathrm{E}[\phi(X)] \neq \phi(\mathrm{E}[X])$, $\mathrm{E}[X^2] \neq (\mathrm{E}[X])^2$.
これ, $\sin(x^2) \neq (\sin(x))^2$ と同じくらい当たり前+だいじ.

母分散の性質

X: 確率変数, $a, b \in \mathbb{R}$:定数 のとき,

$$\mathrm{V}[aX + b] = a^2 \mathrm{V}[X].$$

母分散の性質

$$\mathrm{V}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$$

L05-Q2

Quiz(離散的な確率変数の母平均値・母分散・母標準偏差・確率)

確率変数 X は

- 値 $X = -2$ を確率 $\frac{2}{5}$ で
- 値 $X = +1$ を確率 $\frac{3}{5}$ で

とる.

1. 母平均値 $\mathrm{E}[X]$ を求めよう.
2. 母期待値 $\mathrm{E}[2X+1]$ を求めよう
3. 母期待値 $\mathrm{E}[X^2]$ を求めよう

L05-Q3

母平均値がチーム番号であるような確率分布を作ろう.
ただし,

$$f_x = \begin{cases} C(\text{一定}) & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

型は簡単すぎるから禁止.

L05-Q4

Quiz(離散的な確率変数の母平均値・母分散・母標準偏差・確率)

確率変数 X は次の確率分布に従う.

$$f_x = \begin{cases} \frac{x}{5050} & (0 \leq x \leq 100) \\ 0 & (\text{他}) \end{cases}$$

1. 確率 $P(X \leq 50)$ を求めよう.
2. 母平均値 $\mathrm{E}[X]$ を求めよう.
3. 母分散 $\mathrm{V}[X]$ を求めよう.

今回の予習問題

今回の予習問題は「紙のレポート」的な感じ. 配点も増します.
e ラーニングシステム上の予習問題 (17:00 公開) を A4 の紙に印刷して, 過程付き解答を A4 の紙に手書きで作成して, Math ラウンジチューターに提出してください.
期限 2015-11-05 木昼 (12:45-13:30) の Math ラウンジ, ですが, この出題範囲の非参照テストの前, 2015-10-22 木昼 (12:45-13:30) の Math ラウンジまでに提出することをおすすめします.

連絡

- 統計検定 申込締切 2015-10-16 金, 受験 2015-11-29 日. 3 級 or 2 級.
- オフィスアワー月 4 木 6(1-502)
- 臨時教室変更 1-542(2015-10-23 金 2)
- プチテスト計画 (2015-11-13 金 2)

manaba 出席カード提出
https://attend.ryukoku.ac.jp