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

- 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 → 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

- 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
♡1	$\frac{1}{52}$
♡2	$\frac{1}{52}$
⋮	⋮
♠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 岩薙林 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

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

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

Ω が有限集合で, 等しく確からしい, なら,

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

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

ここで,

定義 (特徴関数)

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

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.

連続型確率変数

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　L02(2023-04-17 Mon)

最終更新: Time-stamp: "2023-04-19 Wed 11:40 JST hig"

今日の目標

- 離散型確率変数の確率が計算できる 岩薙林 確率・統計 §3.2
- 離散型確率変数の標本抽出ができる
- 連続型確率変数とは何か説明できる 岩薙林 確率・統計 §3.3
- 連続型確率変数の母期待値母平均値母分散が計算で

L01-Q1

Quiz 解答: 離散的な確率変数の母平均・母分散・母標準偏差

1. 母期待値 $\mathrm{E}[\mathrm{e}^X] = \frac{4}{12}\cdot \mathrm{e}^{-1} + \frac{5}{12}\cdot \mathrm{e}^0 + \frac{3}{12}\cdot \mathrm{e}^2$.
2. 母平均値 $\mathrm{E}[X] = \frac{4}{12}\cdot(-1) + \frac{5}{12}\cdot 0 + \frac{3}{12}\cdot 2 = \frac{1}{6} (= \mu)$.
3. 母分散
   $\mathrm{V}[X] = \mathrm{E}[(X-\mu)^2] = \frac{4}{12}\cdot(-1-\frac{1}{6})^2 + \frac{5}{12}\cdot(0-\frac{1}{6})^2 + \frac{3}{12}(2-\frac{1}{6})^2 = \frac{47}{36}$.

な別解 $\mathrm{E}[X^2] = (-1)^2\cdot\frac{4}{12} + 0^2\cdot\frac{5}{12} + 2^2\frac{3}{12} = \frac{4}{3}$. $\mathrm{V}[X] \overset{\mathrm{V1}}{=} \mathrm{E}[X^2] - \mu^2 = \frac{47}{36}$.

計算を楽にする母期待値の性質

命題 (母期待値の性質 高校 数学 B)

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

$$\mathrm{E}[1] = 1 \tag{E1}$$

$$\mathrm{E}[g_1(X) + g_2(X)] = \sum_x (g_1(x) + g_2(x)) \times p(x) = \mathrm{E}[g_1(X)] + \mathrm{E}[g_2(X)]$$

(E2, 岩薩林 確率・統計 (3.5))

$$\mathrm{E}[a \cdot g(X)] = \sum_x (a \cdot g(x)) \times p(x) = a\mathrm{E}[g(X)] \tag{E3}$$

$$\mathrm{E}[aX + b] \overset{\mathrm{E2}}{=} \mathrm{E}[aX] + \mathrm{E}[b] \overset{\mathrm{E3}}{=} a\mathrm{E}[X] + b\mathrm{E}[1] \overset{\mathrm{E1}}{=} a\mathrm{E}[X] + b$$ (E4, 岩薩林 確率・統計 (3.6))

定義 (k 次のモーメント 岩薩林 確率・統計 p.53)

$\mathrm{E}[X^k]$ を X の k 次のモーメントという $(k = 0, 1, 2, \dots)$.

k 次のモーメントを部品として計算しておき, 他の母期待値は性質から導くのが楽.

母分散の性質 岩薩林 確率・統計 (3.8)(3.9)p.55

母分散の性質 高校 数学 B 岩薩林 確率・統計 (3.9)p.55

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

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

$$\mathrm{V}[aX+b] = a^2\mathrm{V}[X]. \qquad \text{(V2,(3.9))}$$

証明 V1 $\mu = \mathrm{E}[X]$ とする.

$$\mathrm{V}[X] = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2 - 2\mu X + \mu^2] \overset{\mathrm{E2,E3}}{=}$$

$$\mathrm{E}[X^2] - 2\mu\mathrm{E}[X] + \mu^2\mathrm{E}[1] \overset{\mathrm{E1}}{=} \mathrm{E}[X^2] - 2\mu^2 + \mu^2.$$

証明 V2 $\mu = \mathrm{E}[X]$ と, $\mathrm{E}[Y] = \mu_Y = a\mu + b$ とする.

$$\begin{aligned}\mathrm{V}[Y] =& \mathrm{E}[(Y-\mu_Y)^2] \\ =& \mathrm{E}[(aX+b) - (a\mu+b))^2] \\ =& a^2\mathrm{E}[(X-\mu)^2] = a^2\mathrm{V}[X].\end{aligned}$$

ここまで来たよ

1. 離散型確率変数
   - 事象と確率
   - 標本抽出と Python の scipy による確率変数の扱い
2. 連続型確率変数
   - 確率密度関数

確率変数にまつわる確率

定義 (確率関数 岩薩林 確率・統計 (3.1) **)**

x の関数 $p(x) = P(X = x)$ を離散型確率変数 X の確率関数という.

定義 (累積分布関数 岩薩林 確率・統計 (3.2) **)**

x の関数 $F(x) = P(X \leq x)$ を確率変数 X の累積分布関数という.

命題 (累積分布関数の性質 岩薩林 確率・統計 pp.50,51 **)**

(C1) $F(x)$ は広義単調増加関数
(C2) $F(-\infty) = 0, F(+\infty) = 1.$
(C3) $P(a < X \leq b) = F(b) - F(a).$
(C4) $F(x) = \sum_{x' \leq x} p(x').$

小文字 $x \in \mathbb{R}$ はふつうの変数 (実数), 大文字 X は確率変数.

累積分布関数

確率変数 X の確率関数を次とする.

$$p(x) = \begin{cases} 0.1 & (x=0) \\ 0.2 & (x=1) \\ 0.3 & (x=2) \\ 0.1 & (x=3) \\ 0.1 & (x=4) \\ 0 & (x=5) \\ 0.2 & (x=6) \\ 0 & (他) \end{cases} =$$

p_k	x_k
0.1	(x=0)
0.2	(x=1)
0.3	(x=2)
0.1	(x=3)
0.1	(x=4)
0	(x=5)
0.2	(x=6)

ここまで来たよ

1. 離散型確率変数
   - 事象と確率
   - 標本抽出と Python の scipy による確率変数の扱い
2. 連続型確率変数
   - 確率密度関数

確率変数の標本抽出

岩薩林 確率・統計 §5 を先取り

定義 (標本, 標本抽出)

確率変数 X の値を得る試行を n 回行って得られる x の長さ n の列を,
サイズ (size)n の標本 (sample) という.
標本を得る操作を標本抽出 (sampling) という.

標本は ‘毎回違った結果’ になる.
データ分析 では標本を扱っていた.

scipy による確率変数の標本抽出

https://colab.research.google.com

```
1  import numpy as np # ライブラリの読込
2  from scipy import stats
3
4  xk = (0,  1 ,  2  , 3  ,4   , 5  ,6)    # x_1, x_2,..,x_k
5  pk = (0.1, 0.2, 0.3, 0.1, 0.1, 0.0, 0.2) # p_1, p_2,..,p_k
6  rvx = stats.rv_discrete(values=(xk, pk)) # 離散型確率変数 X を定義
7  sample=rvx.rvs(size=30) # サイズ30の標本抽出(毎回違った結果になる)
8
9  # 母ナントカ. 答はいつも同じ
10 rvx.mean()# 母平均値
11 rvx.var() # 母分散
12 rvx.std() # 母標準偏差
13
14 rvx.cdf(x)# 累積分布関数
15
16 # データ分析 でやってた標本平均値
17 ## 標本抽出のたびに結果が変わる
18 sample.mean() # 標本平均値
19 sample.var()  # 不偏標本平均値
20 sample.std()  # 不偏標本標準偏差
```

ここまで来たよ

1. 離散型確率変数
   - 事象と確率
   - 標本抽出と Python の scipy による確率変数の扱い
2. 連続型確率変数
   - 確率密度関数

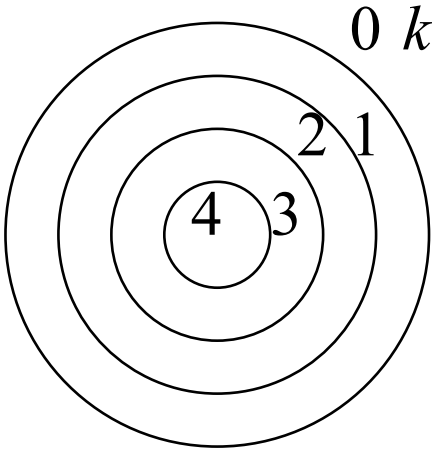
離散型確率変数の例

ジャンボ宝くじの賞金 X 円は離散型確率変数.
確率関数

$$p(x) = \begin{cases} \frac{889}{1000} & (x = 0) \\ \frac{1}{10} & (x = 300) \\ \frac{1}{100} & (x = 3000) \\ \frac{1}{1000} & (x = 10000) \\ 0 & (他) \end{cases}$$

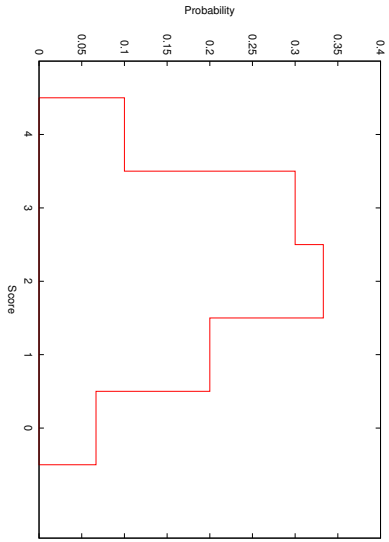
あるプレイヤーのダーツの得点確率

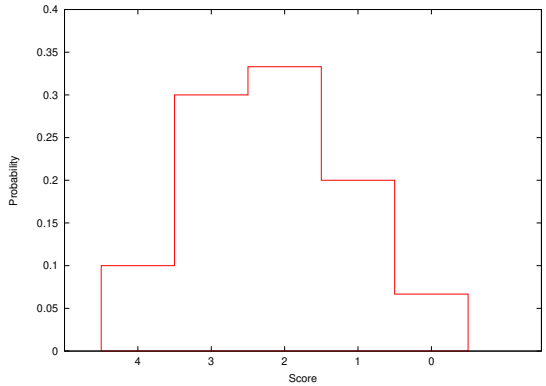
得点: 的の真ん中から順に $4, 3, 2, 1, 0$ 点



離散型確率分布

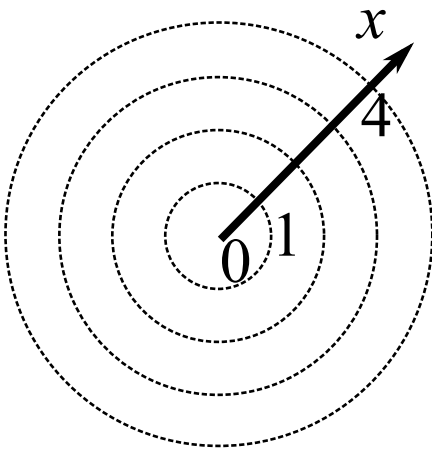
得点 s	確率関数 $f(s)$
4	0.1
3	0.3
2	0.3333
1	0.2
0	0.0667

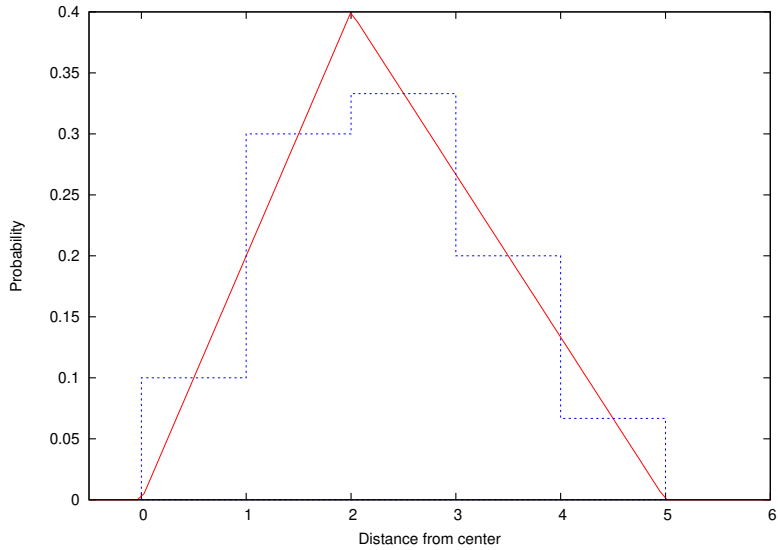




中心から x cm にあてる確率 岩薩林 確率・統計 §4.1

的の真ん中からの距離 x cm, 得点 $s = 4 - x$ 点 (実数).





$x = 0.5$cm と 0.9 cm への当たりやすさは違う. $x = 1.0$cm を境に急に変わるわけじゃない. これを表現したい.

$\rightsquigarrow$ 点数の出やすさは x のある関数 $p(x)$ で表される!

連続型確率変数 連続型確率分布

連続型 確率密度関数 $f(x)$ (x は実数)

離散型 確率関数 $p(x)$ (x は整数またはとびとびの値)

連続型確率変数

定義 (連続型確率変数)

連続型確率変数 X とは, 実数値をとり, 確率が確率密度関数 $f(x)$ で指定されるもの.

離散的

得点 x	確率 $p(x)$
0	0.1
1	0.3
$\vdots$	
x	$p(x)$

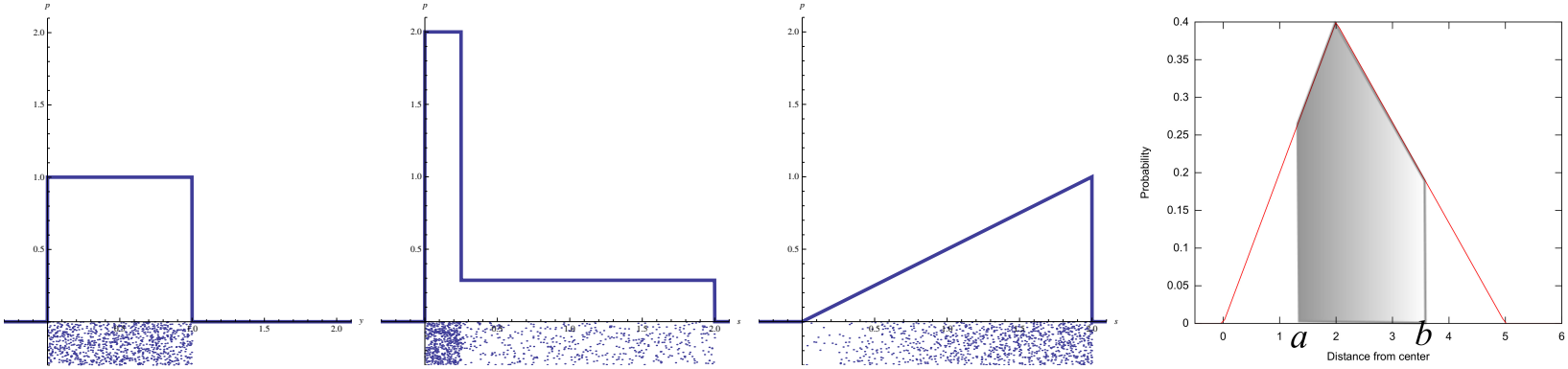
連続的

- $f(x)$ が大きいほど, その値 x がでやすい
- $0 \leq f(x)$ である. $f(x) \leq 1$ とは限らない.

物理・工学系では $p(x)$ と書いたら確率密度関数 $f(x)$ を意味することも

計算科学☆演習 II

確率密度関数の例



横軸下の細かい点が, 標本 (縦方向の位置はランダムで意味なし)

定義 (累積分布関数 岩薩林 確率・統計 (4.4))

$$F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' \qquad \text{(C4 相当)}$$

(C1),(C2) を満たす.

命題 (確率密度関数と確率 岩薩林 確率・統計 (4.1))

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x \quad \text{(下側面積)} \qquad \text{(C3)}$$

連続型確率変数で, $P(x = a) = ?$
ちょうど距離 $x = a$ cm となる確率は? $\rightsquigarrow$ 0.

$$1 = P(\text{全事象}) = P(-\infty < X < +\infty) = \int_{-\infty}^{+\infty} 1 \, \mathrm{d}x.$$

連続型確率変数の母期待値 岩薩林 確率・統計 §4.2

定義 (母期待値 岩薩林 確率・統計 (4.8))

$$\text{離散型確率変数} \quad \mathrm{E}[g(X)] = \sum_x g(x) \cdot p(x)$$

$$\text{連続型確率変数} \quad \mathrm{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x)\, \mathrm{d}x$$

- 離散型と同じ定義: 母平均値 $\mu = \mathrm{E}[X]$, 母分散 $\mathrm{V}[X] = \mathrm{E}[(X-\mu)^2]$
- 離散型と同じ公式 $\mathsf{E}n$, $\mathsf{V}n$ が成立

定義 (k 次のモーメント ($k = 0, 1, 2, 3, \cdots$) 岩薩林 確率・統計 (4.9))

$$\mathrm{E}[X^k] = \int_{-\infty}^{+\infty} x^k \cdot f(x)\, \mathrm{d}x$$

L02-Q1

Quiz(連続型確率変数)

次の確率密度関数 $f(x)$ を持つ確率変数 X を考える. 次を求めよう.

$$f(x) = \begin{cases} 8x & (0 \leq x < \frac{1}{2}) \\ 0 & (\text{それ以外}) \end{cases}$$

1. 母期待値 $\mathrm{E}[X^k]$ $(k = 0, 1, 2, \ldots)$
2. 母平均値 $\mathrm{E}[X]$
3. 母分散 $\mathrm{V}[X]$
4. 母期待値 $\mathrm{E}[(2X+3)^2]$
5. 母分散 $\mathrm{V}[2X+3]$
6. 確率 $P(-\frac{1}{4} < X \leq \frac{1}{4})$.

岩薩林 確率・統計 例題 4.2,4.3

連続型確率変数の母期待値と分位数

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　L03(2023-04-24 Mon)

最終更新: Time-stamp: "2023-04-24 Mon 11:29 JST hig"

今日の目標

- 連続型確率変数の確率 岩薩林 確率・統計 §4.1 ・分位数 岩薩林 確率・統計 なし を計算できる
- 連続型確率変数の母期待値母平均値母分散が計算できる 岩薩林 確率・統計 §4.2

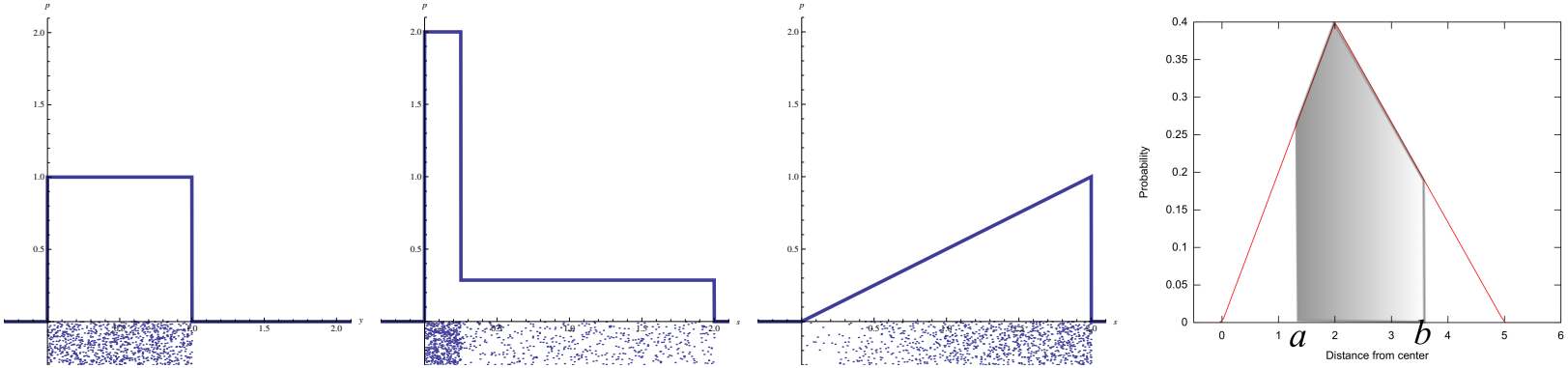
L02-Q1 $P(-\frac{1}{4} < X \leq \frac{1}{4}) = \int_{-\frac{1}{4}}^{+\frac{1}{4}} f(x)\,\mathrm{d}x = \int_{-\frac{1}{4}}^{0} 0\,\mathrm{d}x + \int_{0}^{+\frac{1}{4}} 8x\,\mathrm{d}x = \frac{1}{4}.$

ここまで来たよ

2. 連続型確率変数の確率

3. 連続型確率変数の母期待値と分位数
   - 累積分布関数と分位数関数
   - 連続型確率変数の母期待値
   - 指数分布

復習: 連続型確率変数の確率密度関数



横軸下の細かい点が, 標本 (縦方向の位置はランダムで意味なし)

定義 (累積分布関数 岩薩林 確率・統計 (4.4))

$$F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' \qquad \text{(C4 相当)}$$

確率統計 I(2023)LL02 | 岩薩林 確率・統計 pp.50,51 (C1),(C2) を満たす.

命題 (確率密度関数と確率 岩薩林 確率・統計 (4.1))

$$P(c < X \leq d) = F(d) - F(c) = \int_{c}^{d} f(x)\ \mathrm{d}x \quad \text{(下側面積)} \qquad \text{(C3)}$$

連続型確率変数の分位数関数

定義 (分位数関数 岩薩林 確率・統計 なし)

確率変数 X の 累積分布関数の, 区間 $[0,1]$ で定義された逆関数 $x = F^{-1}(y)$ を分位数関数, $F^{-1}(q)$ を, q-分位数 という.

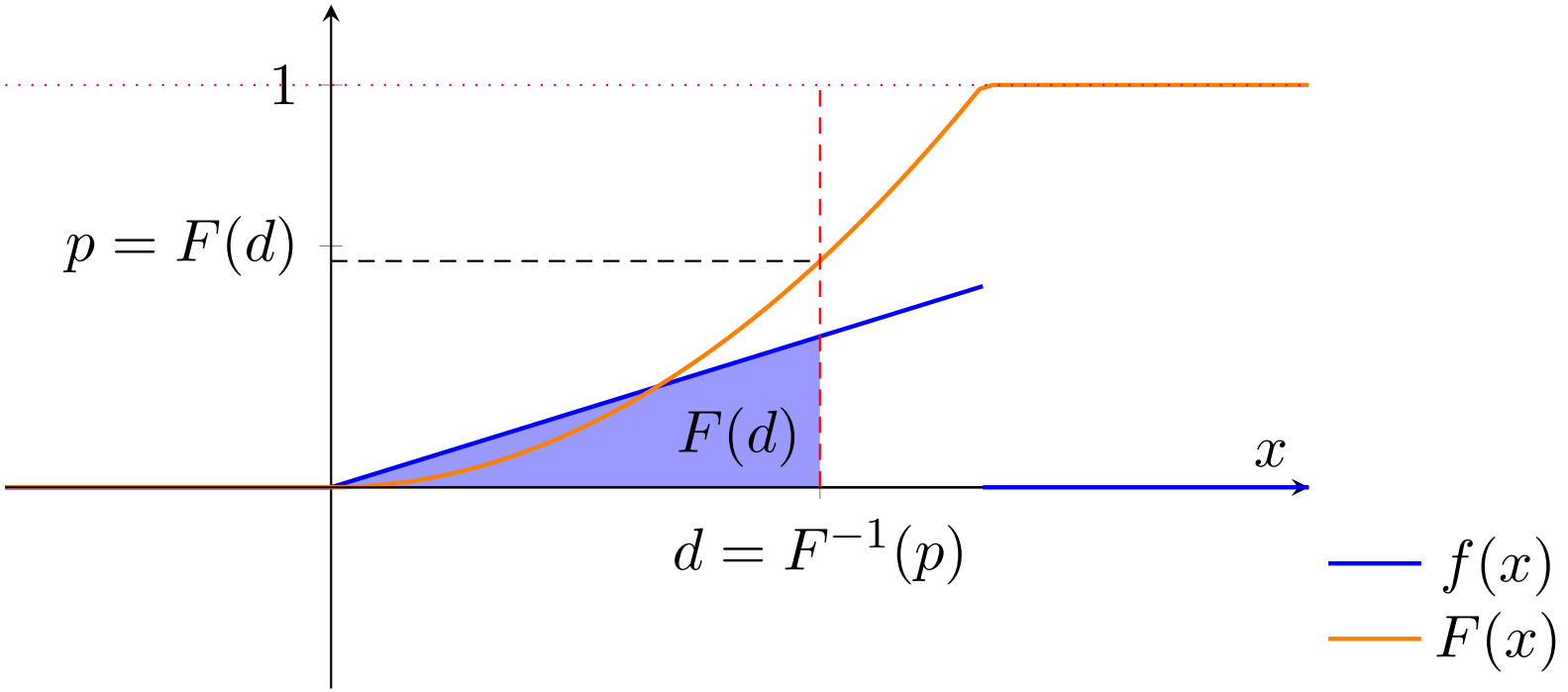
意味

q-分位数とは, 確率 $P(X \leq d)$ が q に等しくなるような境い目 $x = d$.
$F^{-1}(y)$ は, 確率 $y = q$ を与えると, そうなる境い目を返してくれる関数.

言い訳

$F(x)$ が狭義単調増加でない限り逆関数 F^{-1} は定義できない. 逆像 $F^{-1}(y)$ は集合になるので $\inf F^{-1}(y)$ とするのが正確. ただし, $F^{-1}(0)$ は極限 $\lim_{\epsilon \to +0} F^{-1}(+\epsilon)$ とする.
'確率 $P(X \leq d)$ が q に等しくなる**最小ぎりぎりの**境い目 $x = d$' のこと

```
1  rvx.ppf(q) # ppf=percent point function 分位数関数(の類義語)
```



例 (中央値・四分位数)

$F^{-1}(\frac{1}{2})$ は分布の (母) 中央値, $F^{-1}(\frac{1}{4}), F^{-1}(\frac{3}{4})$ は分布の (母) 四分位数.

データ分析

(標本) 中央値 岩薩林 確率・統計 p.27

L03-Q1

Quiz(連続型確率変数)

次の確率密度関数 $f(x)$ を持つ連続型確率変数 X を考える．次を求めよう．

$$f(x) = \begin{cases} 8x & (0 \leq x < \frac{1}{2}) \\ 0 & (\text{それ以外}) \end{cases}$$

1. 母期待値 $\mathrm{E}[X^k]$ $(k = 0, 1, 2, \ldots)$
2. 母平均値 $\mathrm{E}[X]$
3. 母分散 $\mathrm{V}[X]$
4. 母期待値 $\mathrm{E}[(2X+3)^2]$
5. 母分散 $\mathrm{V}[2X+3]$
6. 確率 $P(-\frac{1}{4} < X \leq \frac{1}{4})$
7. 分布の中央値と $\frac{1}{4}$-分位数

ここまで来たよ

2 連続型確率変数の確率

3 連続型確率変数の母期待値と分位数
- 累積分布関数と分位数関数
- 連続型確率変数の母期待値
- 指数分布

連続型確率変数の母期待値 岩薩林 確率・統計 §4.2

定義 (母期待値 岩薩林 確率・統計 (4.9))

離散型確率変数 $\mathrm{E}[g(X)] = \sum_x g(x) \cdot p(x)$

連続型確率変数 $\mathrm{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x)\, \mathrm{d}x$

離散型・連続型共通の定義.

定義 (母平均値 岩薩林 確率・統計 (4.6) ・母分散 岩薩林 確率・統計 (4.11))

母平均値 $\mu = \mathrm{E}[X^1]$, 母分散 $\sigma^2 = \mathrm{V}[X] = \mathrm{E}[(X-\mu)^2]$.

定義 (k 次のモーメント ($k = 0, 1, 2, 3, \cdots$) 岩薩林 確率・統計 (4.10))

$$\mathrm{E}[X^k] = \int_{-\infty}^{+\infty} x^k \cdot f(x)\, \mathrm{d}x$$

復習: 母分散の性質 岩薩林 確率・統計 (3.8)(3.9)p.55

離散型・連続型共通の性質．公式 $\mathrm{E}n$, $\mathrm{V}n$.

母分散の性質 高校 数学 B 岩薩林 確率・統計 (3.9)p.55

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

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

$$\mathrm{V}[aX+b] = a^2\mathrm{V}[X]. \qquad \text{(V2,(3.9))}$$

証明 V2 $\mu = \mathrm{E}[X]$ と, $\mathrm{E}[Y] = \mu_Y = a\mu + b$ とする.

$$\begin{aligned}\mathrm{V}[Y] =& \mathrm{E}[(Y-\mu_Y)^2] \\ =& \mathrm{E}[((aX+b)-(a\mu+b))^2] \\ =& a^2\mathrm{E}[(X-\mu)^2] = a^2\mathrm{V}[X].\end{aligned}$$

L03-Q1

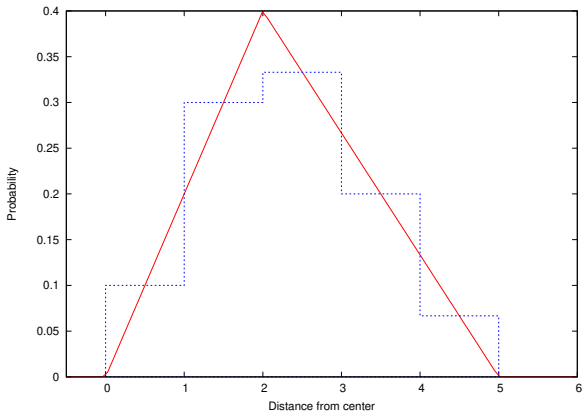
岩薩林 確率・統計 例題 4.2,4.3

連続型確率変数の母平均値と母分散の直観的意味

$\mu = \mathrm{E}[X], \sigma^2 = \mathrm{V}[X]$

その 1: 確率密度関数のグラフから

$y = 0$ と $y = f(x)$ で囲まれる形の (一様な質量 $f(x)$ のグラフ密度 1 の) 板を考える.



$\mathrm{E}[X^0]$: 板の 質量 は 1.

$\mathrm{E}[X^1]$ 板の 重心の横方向の位置 は $x = \mu$.

$\mathrm{V}[X]$ 板 の アバウトに幅 は σ,

$x = \mu$ のまわりの慣性モーメント は σ^2.

その 2: チェビシェフの不等式

母平均値から, (σ を単位に測って) 離れた値がでる確率は小さい.

その 3: 大数の弱法則

宝くじを何回も買うと, 1 回あたりの平均の賞金は,
$\mathrm{E}[$賞金$]$ に近い

ここまで来たよ

2. 連続型確率変数の確率

3. 連続型確率変数の母期待値と分位数
   - 累積分布関数と分位数関数
   - 連続型確率変数の母期待値
   - 指数分布

指数分布 岩薩林 確率・統計 p.80

定義 (指数分布)

連続型確率変数 X が 確率密度関数

$$f(x;b,a)=\begin{cases}\frac{1}{a}\mathrm{e}^{-\frac{x-b}{a}} & (\frac{x-b}{a}\geq 0)\\ 0 & (他)\end{cases}$$

を持つとき, X はパラメタ a,b の指数分布にしたがうといい (科目ローカル記号) $\mathrm{Exp}(b,a)$ とかく. (パラメタは確率変数ではない. $8x^1$ の 8 や 1 のこと.)
世の中 岩薩林 確率・統計 p.80 では, $b=0, a>0$ の場合だけを指数分布とよぶ.

性質 $\mathrm{E}[X]=b+a$, $\mathrm{V}[X]=a^2$.
意味時間的にランダムに起きる事象，例えば，「放射性元素が崩壊する」「宝くじが当たる」のような事象の，時間間隔 x のしたがう分布.

L03-Q2

Quiz(指数分布の母平均値・母分散・確率)

パラメタ $b = 0, a = 3$ の指数分布にしたがう確率変数 $X \sim \mathrm{Exp}(0, 3)$ を考える．すなわち，X の確率密度関数は次で与えられる．

$$f(x) = \begin{cases} \frac{1}{3}\mathrm{e}^{-x/3} & (x \geq 0) \\ 0 & (\text{他}) \end{cases}$$

1. $\mathrm{E}[X]$ を求めよう．
2. $\mathrm{V}[X]$ を求めよう．
3. 累積分布関数を求めよう．確率 $P(1 < X \leq 3)$ を小数点以下第 3 位まで求めよう．
4. 分位数関数を求めよう．確率 $P(X \leq d) = \frac{1}{3}$ となるような d を小数点以下第 3 位まで求めよう．

公式: $\int x^0 \mathrm{e}^{-x} \mathrm{d}x = \mathrm{e}^{x}$, $\int x^1 \mathrm{e}^{-x} \mathrm{d}x = -(x+1)\mathrm{e}^{-x}$, $\int x^2 \mathrm{e}^{-x} \mathrm{d}x = -(x^2+2x+2)\mathrm{e}^{-x}$.

指数関数, 対数関数の値の求め方: Windows の電卓．iPhone を横向きにして関数電卓で．または, Google Colab で np.exp(数値), np.log(数値).

Python で連続型確率変数

- expon: 指数分布 (固有名詞)
- loc: 分布の (重心とは限らないけど) 位置を表す. 指数分布の b.
- scale: 分布の (標準偏差とは限らないけど) 幅を表す. 指数分布の a.

指数分布

```
1 rvx = stats.expon(loc=0,scale=2) # loc=b=0, scale=a=1/lambda=2
```

分位数関数

```
1 rvx.ppf(q) # ppf=percent point function 分位数関数 (の類義語)
```

rvx.cdf(x) の逆関数.

Google Colab 連続型確率変数-指数分布.ipynb hig3Moodle

指数分布・一様分布と確率変数の標準化

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　L04(2023-05-01 Mon)

最終更新: Time-stamp: "2023-05-01 Mon 12:00 JST hig"

今日の目標

- 指数分布 岩薩林 確率・統計 p.80, 一様分布 岩薩林 確率・統計 p.78 を説明できる
- scipy.stat の location と scale を説明できる
- 確率変数の標準化を説明できる 岩薩林 確率・統計 p.80

L03-Q1

Quiz 解答: 連続型確率変数

1. k 次のモーメントは,

$$\mathrm{E}[X^k] = \int_{-\infty}^{+\infty} x^k \cdot f(x)\,\mathrm{d}x = \int_0^{1/2} x^k 8x\,\mathrm{d}x = \frac{2^3 \cdot 2^{-k-2}}{k+2}.$$

   $\mathrm{E}[X^0] = 1$ が確認できる.
2. $\mathrm{E}[X^1] = \frac{1}{3}$.
3. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{1}{8} - (\frac{1}{3})^2 = \frac{1}{72}$.
4. $\mathrm{E}[(2X+3)^2] = 4\mathrm{E}[X^2] + 12\mathrm{E}[X^1] + 9\mathrm{E}[X^0] = 4 \times \frac{1}{8} + 12 \times \frac{1}{3} + 9 = \frac{27}{2}$.
5. $\mathrm{V}[2X+3] = 2^2\mathrm{V}[X] = \frac{1}{18}$.
6. $P(-\frac{1}{4} < X \leq \frac{1}{4}) = \int_{-\frac{1}{4}}^{+\frac{1}{4}} f(x)\,\mathrm{d}x = \int_{-\frac{1}{4}}^{0} 0\,\mathrm{d}x + \int_0^{+\frac{1}{4}} 8x\,\mathrm{d}x = \frac{1}{4}$.

7. 累積分布関数は，

$$F(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x' = \begin{cases} 0 & (x \leq 0) \\ 4x^2 & (0 < x \leq \frac{1}{2}) \\ 1 & (\frac{1}{2} < x) \end{cases}.$$

区間 $[0, 1]$ で逆関数を考えると，分位点関数は

$$F^{-1}(y) = \frac{1}{2}\sqrt{y}.$$

中央値すなわち $\frac{1}{2}$-分位数は $F^{-1}(\frac{1}{2}) = \frac{1}{2\sqrt{2}}$, $\frac{1}{4}$-分位数は $F^{-1}(\frac{1}{4}) = \frac{1}{4}$.

別解 $\int_{-\infty}^{d} f(x)\mathrm{d}x = \frac{1}{2}$ を d について解く.

ここまで来たよ

3. 連続型確率変数の母期待値と分位数

4. 指数分布・一様分布と確率変数の標準化
   - 指数分布
   - 連続型一様分布
   - 確率変数の標準化

指数分布 岩薩林 確率・統計 p.80

連続型確率変数の分布の, 名前がつくくらい有名な例

定義 (指数分布)

連続型確率変数 X が 確率密度関数

$$f(x;b,a)=\begin{cases}\frac{1}{a}\mathrm{e}^{-\frac{x-b}{a}} & (x>b)\\ 0 & (他)\end{cases}$$

を持つとき, X はパラメタ $a(>0), b$ の指数分布に従うといい, 記号 $\sim$(従う) で $X\sim \mathrm{Exp}(b,a)$ と書く (Exp は科目ローカル記号) (パラメタは確率変数ではない. $8x^1$ の 8 や 1 のこと.)
世の中 岩薩林 確率・統計 p.80 では, $b=0$ の場合だけを指数分布とよぶ.

性質 $\mathrm{E}[X]=b+a$, $\mathrm{V}[X]=a^2$.
意味時間的にランダムに起きる事象，例えば，「放射性元素が崩壊する」「宝くじが当たる」のような事象の，時間間隔 x のしたがう分布.

L04-Q1

Quiz(指数分布の母平均値・母分散・確率)

パラメタ $b = 0, a = 3$ の指数分布にしたがう確率変数 $X \sim \mathrm{Exp}(0, 3)$ を考える．すなわち，X の確率密度関数は次で与えられる．

$$f(x) = \begin{cases} \frac{1}{3}\mathrm{e}^{-x/3} & (x \geq 0) \\ 0 & (\text{他}) \end{cases}$$

1. $\mathrm{E}[X]$ を求めよう．
2. $\mathrm{V}[X]$ を求めよう．
3. 累積分布関数を求めよう．確率 $P(1 < X \leq 3)$ を小数点以下第 3 位まで求めよう．
4. 分位数関数を求めよう．確率 $P(X \leq d) = \frac{1}{3}$ となるような d を小数点以下第 3 位まで求めよう．

公式: $\int x^0 \mathrm{e}^{-x} \mathrm{d}x = \mathrm{e}^{x}$, $\int x^1 \mathrm{e}^{-x} \mathrm{d}x = -(x+1)\mathrm{e}^{-x}$, $\int x^2 \mathrm{e}^{-x} \mathrm{d}x = -(x^2+2x+2)\mathrm{e}^{-x}$.

指数関数, 対数関数の値の求め方: Windows の電卓．iPhone を横向きにして関数電卓で．または, Google Colab で np.exp(数値), np.log(数値).

Python で連続型確率変数

- expon: 指数分布 (固有名詞)
- loc(ation): 分布の (重心とは限らないけど) 位置を表す. 指数分布の b.
- scale: 分布の (標準偏差とは限らないけど) 幅を表す. 指数分布の a.

指数分布

```
1 import scipy.stats
2 rvx = stats.expon(loc=0,scale=a) # Exp(0,a) loc=b=0, scale=a=1
3 # 連続だけにある rvx.cdf(x)の逆関数
4 rvx.ppf(q) # ppf=percent point function 分位数関数(の類義語)
5
6 # 離散にも実はあった
7 rvx.moment(4) # 4次のモーメント
8 rvx.expect(lambda x: x**2 + 3* x +1) # E[X^2+3X+1]
```

Google Colab 連続型確率変数-指数分布.ipynb hig3Moodle

指数分布の拡大縮小

確率変数 $X \sim \mathrm{Exp}(0, a)$, $Y = 2X$ とする.

1. Y の確率密度関数 $f_{\mathrm{Y}}(y)$ を求め, グラフを描こう. $f_{\mathrm{X}}(x)$ との関係は?
2. $\mathrm{E}[Y], \mathrm{V}[Y]$ を 2 つの方法で求めよう.

$$\mathrm{E}[2X] = \int_0^\infty 2x \frac{1}{a} \mathrm{e}^{-x/a} \, \mathrm{d}x \overset{y=2x}{=} \int_0^\infty y \frac{1}{a} \mathrm{e}^{-y/(2a)} \frac{1}{2} \, \mathrm{d}y = \mathrm{E}[Y]$$

ここまで来たよ

3. 連続型確率変数の母期待値と分位数

4. 指数分布・一様分布と確率変数の標準化
   - 指数分布
   - 連続型一様分布
   - 確率変数の標準化

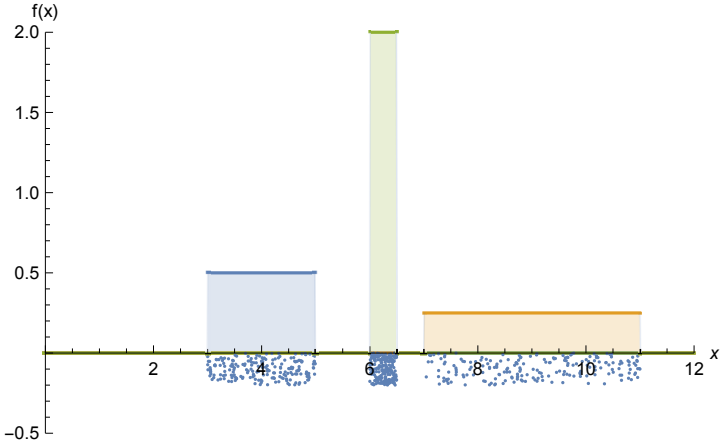
連続型一様分布 岩薙林 確率・統計 例題 4.1(p.78)

連続型一様分布は, 連続型確率変数の分布の, 名前がつくくらい有名な例.

連続型一様分布 $\mathrm{U}(c,d)$

確率変数 X の確率密度関数が次で与えられるとき, X は区間 $[c,d)$ の連続型一様分布 $\mathrm{U}(c,d)$ に従うという.

$$f(x)=\begin{cases}\frac{1}{d-c} & (c\le x\le d)\\ 0 & (\text{他})\end{cases}$$



$\mathrm{U}(4,6), \mathrm{U}(6,6.5), \mathrm{U}(7,11)$.

連続型一様分布 岩薙林 確率・統計 例題 4.1(p.78) II

連続型一様分布

```
1 import scipy.stats
2 rvx = stats.uniform(loc=c, scale=d-c) # U(c,d)
```

Google Colab 連続型確率変数-連続型一様分布.ipynb hig3Moodle

L04-Q2

Quiz(連続型一様分布)

連続型確率変数 X が連続型一様分布 $\mathrm{U}(c,d)$ にしたがう.

1. モーメント $\mathrm{E}[X^k]$ を求めよう.
2. 母平均値 $\mathrm{E}[X]$ を求めよう.
3. 母標準偏差 $\sqrt{\mathrm{V}[X]}$ を求めよう.

$\mathrm{U}(c,d)$ に対するこの結果は, 公式のように記憶して使おう.
母平均値・母標準偏差の意味とマッチしてる?

L04-Q3

Quiz(連続型一様分布の応用)

あるおんぼろキューブアイス製造マシンから立方体の氷が, 一辺 X が 18mm 以上 20mm 以下の範囲で, 同じ確からしさで決まってランダムに出てくる.

1. X のしたがう分布を答えよう.
2. キューブアイスの一辺の長さの母期待値を求めよう.
3. キューブアイスの表面積の母期待値を求めよう.
4. キューブアイスの表面積が $2000[\mathrm{mm}^2]$ 以下である確率を求めよう.
5. キューブアイスの体積の母期待値を求めよう.

ここまで来たよ

3. 連続型確率変数の母期待値と分位数

4. 指数分布・一様分布と確率変数の標準化
   - 指数分布
   - 連続型一様分布
   - 確率変数の標準化

確率変数の標準化

定義 (確率変数の標準化 岩薩林 確率・統計 例題 4.4(p.80))

任意の確率変数 X に対して, $\mu = \mathrm{E}[X], \sigma^2 = \mathrm{V}[X], \sigma > 0$ とする.
確率変数 $Z = \frac{X-\mu}{\sigma}$ を「X を標準化した確率変数」という.

$\mathrm{E}[Z] = \mathrm{E}[\frac{X-\mu}{\sigma}] = 0$,
$\mathrm{V}[Z] = \mathrm{V}[\frac{X-\mu}{\sigma}] = 1$
逆に, 標準化された Z から $X = aZ + b$ を考えると,
$\mathrm{E}[X] = b$,
$\mathrm{V}[X] = a^2$.

Z の確率密度関数 $f_{\mathrm{Z}}(z) \rightsquigarrow X$ の確率密度関数 $f_{\mathrm{X}}(x) = \frac{1}{a} \times f_{\mathrm{Z}}(\frac{x-b}{a})$.
b だけ平行移動, a だけ拡大.
$\frac{1}{a}\times$ は面積を 1 に保つもの.

グラフの平行移動と拡大縮小

$y = f(x)$ のグラフを, y 軸を中心に横 (x 軸方向) に a 倍すると $y = f(\frac{x}{a})$.
さらに, 横 (x 軸方向) に b 平行移動すると $y = f(\frac{x-b}{a})$.
$y = x, y = x^2$ を例に考えてね.
$z = \frac{x-b}{a}, x = az + b$ とおいてる感じです.

連続型一様分布の標準化

$X \sim \mathrm{U}(2,8)$,

$$f_{\mathrm{X}}(x) = \begin{cases} \frac{1}{6} & (2 \leq x \leq 8) \\ 0 & (\text{他} \end{cases}$$

を標準化しよう.

$\mathrm{E}[X] = 5, \mathrm{V}[X] = (8-2)^2/12 = 3$ より, $Z = \frac{X-5}{\sqrt{3}}$

$$f_{\mathrm{Z}}(z) = \begin{cases} \frac{\sqrt{3}}{6} & (-\sqrt{3} \leq z \leq \sqrt{3}) \\ 0 & (\text{他} \end{cases}$$

標準連続型一様分布 $Z \sim \mathrm{U}(-\sqrt{3}, +\sqrt{3})$.

標準化の逆: 標準からもとのやつ I

L04-Q4

Quiz(連続型一様分布)

連続型確率変数 $Z \sim \mathrm{U}(-\sqrt{3}, \sqrt{3})$ に対して, $X = \sqrt{3}Z + 5$ を考える.

1. X の確率密度関数 $f_{\mathrm{X}}(x)$ とそのグラフを答えよう.
2. $\mathrm{E}[X]$ を求めよう.
3. $\mathrm{V}[X]$ を求めよう.

それぞれ 2 つの方法で.

復習

$X = aZ + b$ の母平均値と母分散

$$\begin{aligned}\mathrm{E}[X] =& \mathrm{E}[aZ + b] = a\mathrm{E}[Z] + b \\ \mathrm{V}[X] =& \mathrm{V}[aZ + b] = a^2\mathrm{V}[Z]\end{aligned}$$

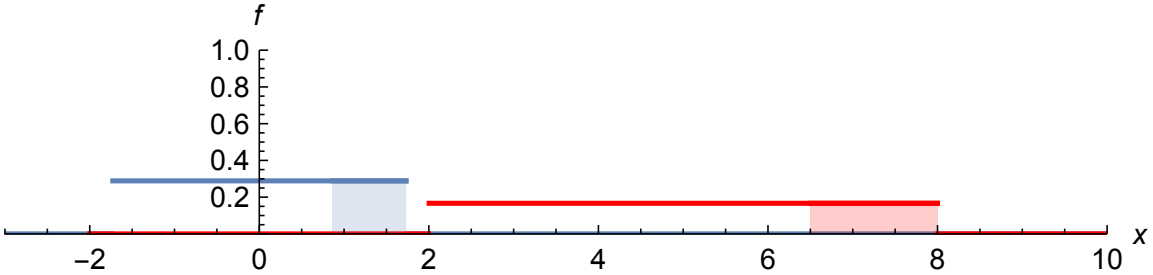
$X = aZ + b$ の意味

岩薩林 確率・統計 標準化 (p.81)

Z が 連続型一様分布 $\mathrm{U}(c, d)$ にしたがうとき,
$X = aZ + b$ は 連続型一様分布 $\mathrm{U}(ac + b, ad + b)$ にしたがう.
一般に X の確率密度関数のグラフは, Z のものを横に a 倍して b 平行移動, (縦に $1/a$ 倍)
$f_{\mathrm{X}}(x) = \frac{1}{a} \times f_{\mathrm{Z}}(\frac{x-b}{a})$

岩薩林 確率・統計 例題 4.4

確率変数 $X = \sqrt{3}Z + 5$ の確率密度関数



左から $Z \sim \mathrm{U}(-\sqrt{3}, \sqrt{3})$, $X = aZ + b = \sqrt{3}Z + 5 \sim \mathrm{U}(2, 8)$.

岩薩林 確率・統計 例題 4.4

$X = aZ + b$ のとき

$$f_{\mathrm{X}}(x) = \frac{1}{a} \times f_{\mathrm{Z}}(\frac{x-b}{a}) = \frac{1}{a} \times \begin{cases} \frac{1}{2\sqrt{3}} & (-\sqrt{3} \le \frac{x-b}{a} \le +\sqrt{3}) \Leftrightarrow (b - a\sqrt{3} \le x < b + a\sqrt{3}) \\ 0 & (\text{他}) \end{cases}$$

確率変数の標準化・標準正規分布・一般の正規分布

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　L05(2023-05-08 Mon)

最終更新: Time-stamp: "2023-05-08 Mon 11:26 JST hig"

今日の目標

- 確率変数の標準化を説明できる 岩薩林 確率・統計 p.80
- 正規分布の確率密度関数, グラフを書ける．母期待値，確率，分位数を求められる． 岩薩林 確率・統計 §4.5

L04-Q1

Quiz 解答: 指数分布の母平均値・母分散・確率

1. 部分積分で $\mathrm{E}[X] = 3$.
2. 部分積分で $\mathrm{E}[X^2] = 2 \cdot 3^2$, $\mathrm{V}[X] = 3^2$.
3. $0 \leq x$ で $F(x) = 1 - \mathrm{e}^{-x/3}$.
   $P(1 < X \leq 3) = F(3) - F(1) = \int_1^3 \frac{1}{3}\mathrm{e}^{-x/3}\mathrm{d}x = \mathrm{e}^{-1/3} - \mathrm{e}^{-1} = 0.349.$
4. $F^{-1}(y) = -3\log(1-y)$. $F^{-1}(\frac{1}{3}) = -3\log\frac{2}{3} = 1.216.$

L04-Q2

Quiz 解答: 連続型一様分布

1. $\mathrm{E}[X^k] = \frac{1}{d-c}\int_c^d x^k \,\mathrm{d}x = \frac{1}{k+1}\frac{d^{k+1}-c^{k+1}}{d-c} = \frac{1}{k+1}(d^k + d^{k-1}c + \cdots + c^k)$.
2. $\mathrm{E}[X^1] = \frac{c+d}{2}$.
3. $\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}[X^1]^2 = \frac{(d-c)^2}{12}$. $\sqrt{\mathrm{V}[X]} = \frac{d-c}{\sqrt{12}} \simeq \frac{d-c}{3.5}$.

L04-Q3
L04-Q4

Quiz 解答: 連続型一様分布

1. $X \sim \mathrm{U}(\sqrt{3}(-\sqrt{3})+5, \sqrt{3}(+\sqrt{3})+5)$, すなわち, $X \sim \mathrm{U}(2,8)$. 関数とグラフ略.
2. $\mathrm{E}[X] = \frac{2+8}{2}$, または, $\mathrm{E}[X] = \sqrt{3}\mathrm{E}[Z]+5 = \sqrt{3}\cdot 0+5$.
3. $\mathrm{V}[X] = \frac{(8-2)^2}{12} = 3$, または, $\mathrm{V}[X] = (\sqrt{3})^2\mathrm{V}[Z] = 3\cdot 1$.

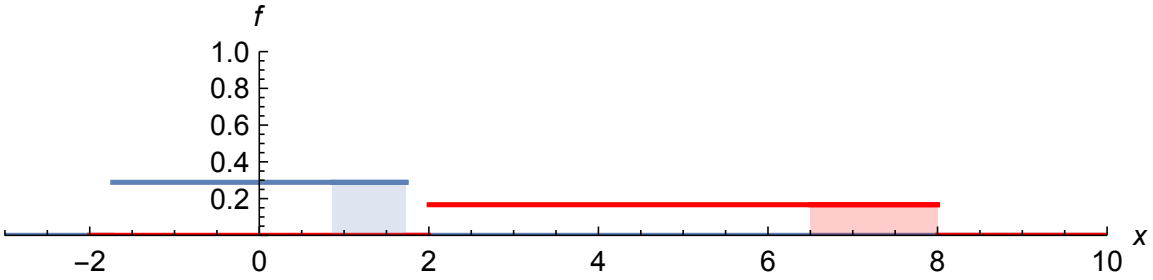
確率変数の変数変換 $X = aZ + b$ の意味

岩薩林 確率・統計 標準化 (p.81)

$Z \sim \mathrm{U}(c, d)$ (連続型一様分布) のとき, $X = aZ + b \sim \mathrm{U}(ac + b, ad + b)$.
一般に確率変数 X の確率密度関数のグラフは, Z のものを横に a 倍, 横に b 平行移動, (縦に $1/a$ 倍)

岩薩林 確率・統計 例題 4.4

$f_{\mathrm{X}}(x) = \frac{1}{a} \times f_{\mathrm{Z}}(\frac{x-b}{a})$
左 $Z \sim \mathrm{U}(-\sqrt{3}, \sqrt{3})$ 右 $X = aZ + b = \sqrt{3}Z + 5 \sim \mathrm{U}(2, 8)$ の確率密度関数



$$f_{\mathrm{X}}(x) = \frac{1}{a} \times f_{\mathrm{Z}}(\frac{x-b}{a}) = \frac{1}{a} \times \begin{cases} \frac{1}{2\sqrt{3}} & (-\sqrt{3} \le \frac{x-b}{a} \le +\sqrt{3}) \Leftrightarrow (b - a\sqrt{3} \le x < b + a\sqrt{3}) \\ 0 & (\text{他}) \end{cases}$$

対応する確率=塗った箇所の面積は同じ
$P(\frac{13}{2} < X \le 8) = P(\frac{\frac{13}{2}-5}{\sqrt{3}} < Z \le \frac{8-5}{\sqrt{3}})$.

確率変数の標準化

定義 (標準化された確率変数 岩薩林 確率・統計 例題 4.4(p.80))

確率変数 Z が, $\mathrm{E}[Z] = 0, \mathrm{V}[Z] = 1^2$ であるとき, Z は標準化された確率変数という.

任意の確率変数 X は, 1 次式 $Z = \frac{X-b}{a}$ で標準化された確率変数に変換できる.

定義 (確率変数の標準化 岩薩林 確率・統計 例題 4.4(p.80))

任意の確率変数 X に対して, $\mu = \mathrm{E}[X], \sigma^2 = \mathrm{V}[X], \sigma > 0$ とする.
確率変数 $Z = \frac{X-\mu}{\sigma}$ を「X を標準化した確率変数」という.

$\mathrm{E}[Z] = \mathrm{E}[\frac{X-\mu}{\sigma}] = 0$, $\mathrm{V}[Z] = \mathrm{V}[\frac{X-\mu}{\sigma}] = 1$
逆に, 標準化された Z に対して,
$\mathrm{E}[aZ + b] = b$, $\mathrm{V}[aZ + b] = a^2$.

標準一様分布

任意の連続型一様分布 $\mathrm{U}(c,d)$ を標準化すると, $\mathrm{U}(-\sqrt{3},+\sqrt{3})$ になる.

グラフの平行移動と拡大縮小

$y = f(x)$ のグラフを, y 軸を中心に横 (x 軸方向) に a 倍すると $y = f(\frac{x}{a})$.
さらに, 横 (x 軸方向) に b 平行移動すると $y = f(\frac{x-b}{a})$.
$y = x, y = x^2$ を例に考えてね.
$z = \frac{x-b}{a}, x = az + b$ とおいてる感じです.

確率変数の標準化とは

確率密度関数 $f_X(x)$ を横方向に拡大縮小して '幅' 1 にする（副産物として縦にも拡大縮小する），平行移動して平均値 '重心' の位置を 0 にすること

ここまで来たよ

4. 指数分布・一様分布と確率変数の標準化

5. 確率変数の標準化・標準正規分布・一般の正規分布
   - 標準正規分布
   - 一般の正規分布

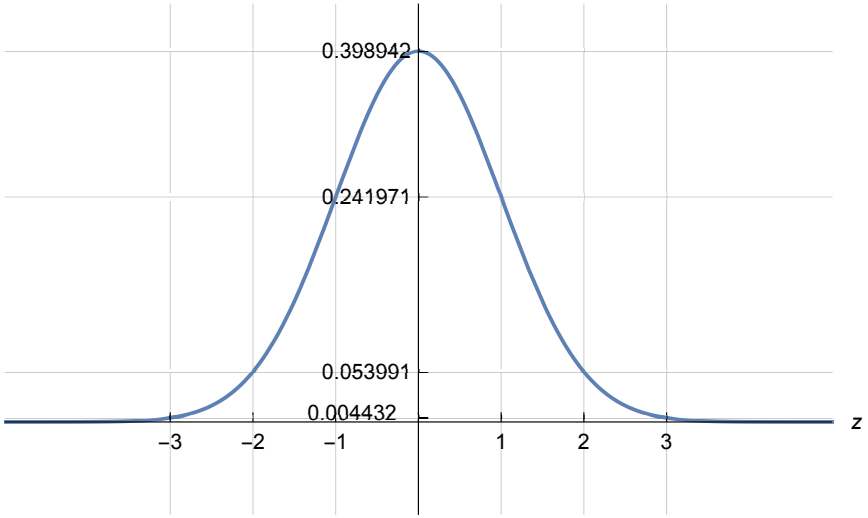
標準正規分布 standard normal distribution $\mathrm{N}(0,1^2)$ の性質

定義 (標準正規分布 $\mathrm{N}(0,1^2)$ 岩薩林 確率・統計 (4.17))

次の確率密度関数 を持つ確率変数 Z を，標準正規分布 $\mathrm{N}(0,1^2)$ にしたがうという.

$$f(z;0,1^2)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}.$$

Python で `scipy.stats.norm(loc=0,scale=1)`,
Excel で`=norm.s.dist(z,FALSE)`



標準正規分布 $Z \sim \mathrm{N}(0, 1^2)$ の母期待値

k 次のモーメント (k: 自然数)

$$\mathrm{E}[Z^{2k-1}] = 0, \quad \text{奇関数}$$

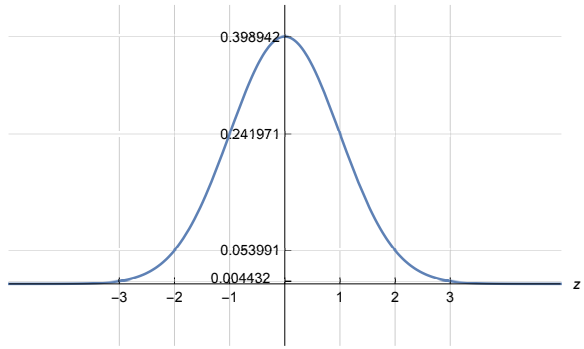
$$\mathrm{E}[Z^{2k}] = (2k-1)!!$$

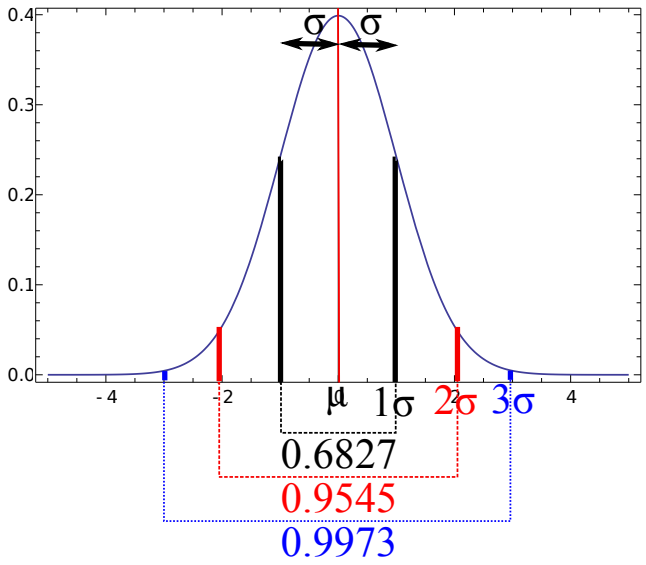
$$= 1 \cdot 3 \cdot 5 \cdots (2k-1) \quad \text{部分積分}$$

$$\text{全確率} \mathrm{E}[Z^0] = 1, \quad \text{岩薙林 確率・統計 (4.19)} \quad \text{微積分 II}$$

$$\text{母平均値} \mathrm{E}[Z] = 0, \quad \text{岩薙林 確率・統計 (4.20)}$$

$$\text{母分散} \mathrm{V}[Z] = 1 \quad \text{岩薙林 確率・統計 (4.21)}$$





たしかに, Z は標準化された分布. Google Colab Python による正規分布 hig3Moodle

標準正規分布の確率と $I(z)$ の数表

$Z \sim \mathrm{N}(0, 1^2)$ のとき,

$$P(c < Z \leq d) = \int_c^d f(z'; 0, 1^2)\,\mathrm{d}z' = \int_c^d \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z'^2}{2}}\,\mathrm{d}z'$$

岩薙林 確率・統計 (4.22)

不定積分は, 簡単な関数では書けないが, 累積分布関数を使って,

$$P(c < Z \leq d) = F(d) - F(c)$$

とし, $F(z)$ を `scipy.stats.norm().cdf(z)` や数表 岩薙林 確率・統計 付表 1(p.227) から求める. 分位数関数 $F^{-1}(q)$ も Python や数表 岩薙林 確率・統計 付表 1(p.227) に頼る.

注: つねに $F(-\infty) = 0, F(+\infty) = 1$. 標準正規分布は確率密度関数が偶関数なので $F(z) + F(-z) = 1$. 特に $F(0) = 1/2$.

岩薙林 確率・統計 付表 1 には $I(z) = F(z) - \frac{1}{2}, z > 0$ の表が載っている. 高校 数学 B

$z < 0$ については, $I(-z) = I(z)$ を経由して求める.

上側確率 $Q(z) = \frac{1}{2} - I(z) = 1 - F(z)$ の表を載せてる教科書も多い.

L05-Q1

Quiz(標準正規分布の確率)

Z は標準正規分布 $\mathrm{N}(0, 1^2)$ に従う連続型確率変数である.

1. 確率 $P(Z \leq 1.23)$ 累積分布関数 $F(z)$ で表そう．Python または表で小数として求めよう.
2. 確率 $P(-0.56 < Z \leq +1.23)$ を累積分布関数 $F(z)$ で表そう．Python または表で求めよう.
3. 確率 $P(Z > d) = 0.025$ となる d を累積分布関数 $F(z)$ の逆関数で表そう．Python または表で求めよう.

岩薩林 確率・統計 例題 4.9(p.91), 問題 8(p.92), 問題 10(p.96)

ここまで来たよ

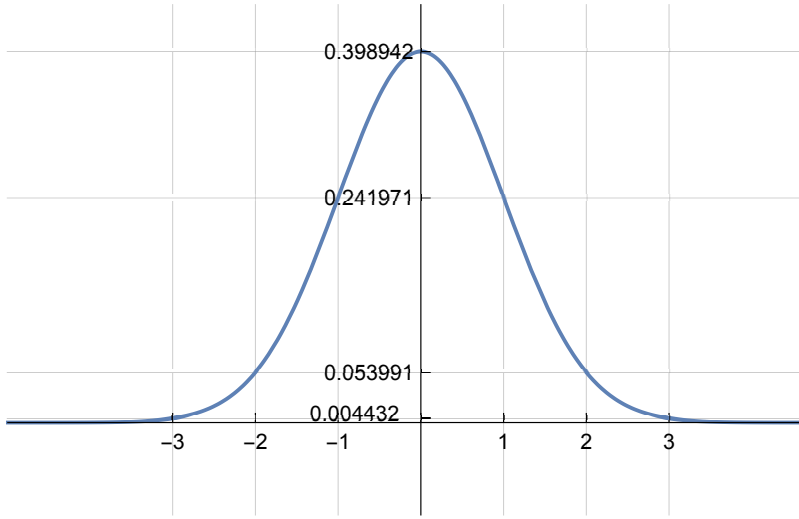
4. 指数分布・一様分布と確率変数の標準化

5. 確率変数の標準化・標準正規分布・一般の正規分布
   - 標準正規分布
   - 一般の正規分布

一般の正規分布 $\mathrm{N}(b, a^2)$

$Z \sim \mathrm{N}(0, 1^2)$ に対して, $X = aZ + b$ を考える.

$$\begin{aligned} \mathrm{E}[X^0] &= 1, \\ \mu = \mathrm{E}[X] &= \mathrm{E}[aZ + b] = b, \\ \sigma^2 = \mathrm{V}[X] &= \mathrm{V}[aZ + b] = a^2, \end{aligned}$$



$$f_{\mathrm{Z}}(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \rightsquigarrow f_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi a^2}} \mathrm{e}^{-\frac{(x-b)^2}{2a^2}}$$

定義 (一般の正規分布 $\mathrm{N}(\mu, \sigma^2)$)

次の確率密度関数を持つ確率変数 X を，母平均値 μ, 母分散 σ^2 の正規分布 $\mathrm{N}(\mu, \sigma^2)$ (normal distribution) にしたがうという.

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

L05-Q2

Quiz(正規分布の確率)

連続型確率変数 X は, 確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi \cdot 3^2}} \mathrm{e}^{-\frac{(x-4)^2}{2\cdot 3^2}}$$

にしたがう.

1. $\mathrm{E}[X]$ を求めよう.
2. $\mathrm{V}[X]$ を求めよう.
3. $f(x)$ のグラフを, 標準正規分布の確率密度関数と重ねて描こう.

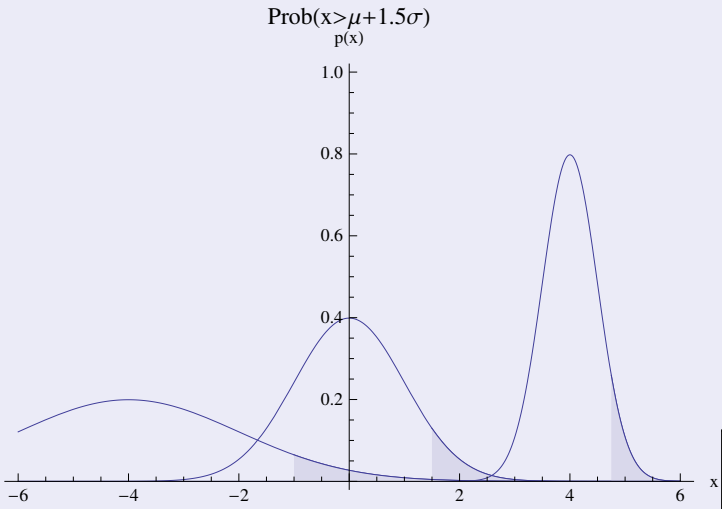
Python でも描いてみよう.

正規分布 $\mathrm{N}(\mu, \sigma^2)$ の確率 I

命題 (標準化前後の確率は同じ)

一般に, 確率は標準化 $Z = \frac{Z-b}{a}$ の前後で変わらない.

$$P(c < X \leq d) = P(\tfrac{c-\mu}{\sigma} < \tfrac{X-\mu}{\sigma} \leq \tfrac{d-\mu}{\sigma}) = P(\tfrac{c-\mu}{\sigma} < Z \leq \tfrac{d-\mu}{\sigma})$$



斜線部の面積はどれも同じ

標準化しても確率が同じことの別説明 (置換積分)

$X \sim \mathrm{N}(\mu, \sigma^2)$ のとき, 積分で $z = \frac{x-\mu}{\sigma}$, $\mathrm{d}z = \frac{1}{\sigma}\mathrm{d}x$ とすると,

$$
\begin{aligned}
P(c < X \leq d) &= \int_c^d \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x \\
&= \int_{\frac{c-\mu}{\sigma}}^{\frac{d-\mu}{\sigma}} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{z^2}{2}} \, \sigma \mathrm{d}z \\
&= P\left(\frac{c-\mu}{\sigma} < Z \leq \frac{d-\mu}{\sigma}\right)
\end{aligned}
$$

L05-Q3

Quiz(正規分布の確率)

確率変数 $X \sim \mathrm{N}(3, 2^2)$, $Z \sim \mathrm{N}(0, 1^2)$ とする.

1. X, Z の確率密度関数の式を書こう. グラフを重ねて描こう.
2. 母期待値 $\mathrm{E}[X^2]$ を求めよう.
3. 確率 $P(X > 5) = P(c < Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
4. 確率 $P(+1 < X \leq 7) = P(c \leq Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
5. 確率 $P(+3 < X \leq 9) = P(c \leq Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
6. 確率 $P(Z \leq d) = 3/4$ となる d を, Python や表を使って小数で求めよう. 確率 $P(X \leq d') = 3/4$ となるような d' を d で表そう.

岩薩林 確率・統計 §4.5 例題 4.10,4.11, 問題 10, 第 4 章練習問題 5

一般の正規分布・多次元の確率変数

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　L06(2023-05-15 Mon)

最終更新: Time-stamp: "2023-05-15 Mon 12:10 JST hig"

今日の目標

- 一般の正規分布の確率密度関数, グラフを書ける．母期待値，確率，分位数を求められる． 岩薩林 確率・統計 §4.5
- 多次元確率変数の同時分布から周辺分布, 母期待値, 母共分散を求められる． 岩薩林 確率・統計 §3.3

L05-Q1

Quiz 解答: 標準正規分布の確率

累積分布関数を $F(z)$ とする. `rvz=stats.norm(loc=0,scale=1)` とする.

1. $P(Z \leq 1.23) = \int_{-\infty}^{1.23} f(z; 0, 1^2)\,\mathrm{d}z = F(1.23) - F(-\infty) = F(1.23) - 0 =$ `rvz.cdf(1.23)` $= 0.5708$.
2. $P(-0.56 < Z \leq +1.23) = \int_{-0.56}^{1.23} f(z; 0, 1^2)\,\mathrm{d}z = F(1.23) - F(-0.56) =$ rvz.cdf(1.23) − rvz.cdf(-0.56) $= 0.6030$.
3. $d = F^{-1}(1 - 0.025) =$ `rvz.ppf(0.975)` $= 1.960$.

L05-Q2

Quiz 解答: 正規分布の確率

1. $\mathrm{E}[X] = 4$.
2. $\mathrm{V}[X] = 3^2$.

3. $x = 4$ を真ん中に幅 3 くらいの正規分布の確率密度関数のグラフ．標準正規分布を，横に 3 倍，縦に $1/3$ 倍，右に 4 平行移動．

ここまで来たよ

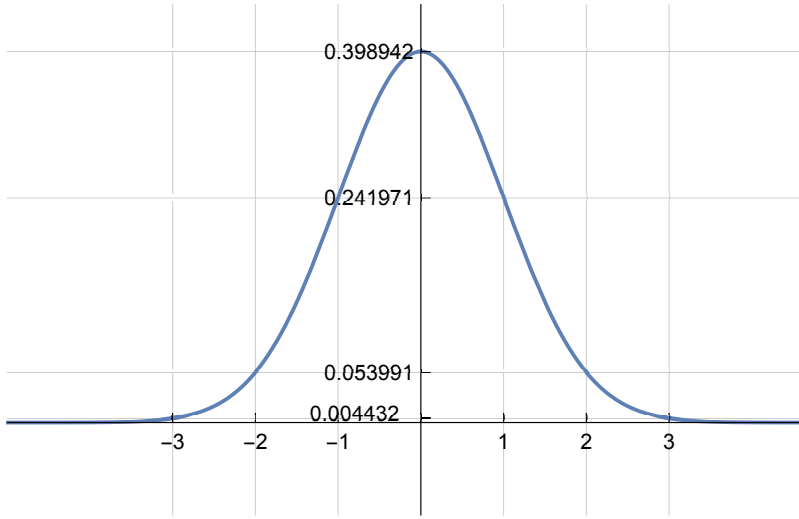
5 確率変数の標準化・標準正規分布・一般の正規分布

6 一般の正規分布・多次元の確率変数
- 一般の正規分布
- 2 次元の確率分布
- 母共分散
- 多次元の連続型確率分布

一般の正規分布 $\mathrm{N}(b, a^2)$

$Z \sim \mathrm{N}(0, 1^2)$ に対して, $X = aZ + b$ を考える.

$$\begin{aligned} \mathrm{E}[X^0] &= 1, \\ \mu = \mathrm{E}[X] &= \mathrm{E}[aZ + b] = b, \\ \sigma^2 = \mathrm{V}[X] &= \mathrm{V}[aZ + b] = a^2, \end{aligned}$$



$$f_{\mathrm{Z}}(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}} \rightsquigarrow f_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi a^2}} \mathrm{e}^{-\frac{(x-b)^2}{2a^2}}$$

定義 (一般の正規分布 $\mathrm{N}(\mu, \sigma^2)$)

次の確率密度関数を持つ確率変数 X を，母平均値 μ, 母分散 σ^2 の正規分布 $\mathrm{N}(\mu, \sigma^2)$ (normal distribution) にしたがうという.

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

L06-Q1

Quiz(正規分布の確率)

連続型確率変数 X は, 確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi \cdot 3^2}} \mathrm{e}^{-\frac{(x-4)^2}{2\cdot 3^2}}$$

にしたがう.

1. $\mathrm{E}[X]$ を求めよう.
2. $\mathrm{V}[X]$ を求めよう.
3. $f(x)$ のグラフを, 標準正規分布の確率密度関数と重ねて描こう.

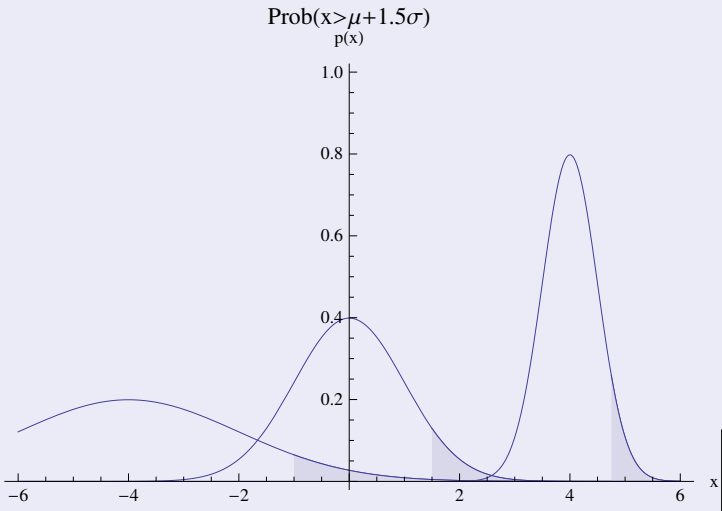
Python でも描いてみよう.

正規分布 $\mathrm{N}(\mu, \sigma^2)$ の確率 I

命題 (標準化前後の確率は同じ)

一般に, 確率は標準化 $Z = \frac{X-\mu}{\sigma}$ の前後で変わらない.

$$P(c < X \leq d) = P(\tfrac{c-\mu}{\sigma} < \tfrac{X-\mu}{\sigma} \leq \tfrac{d-\mu}{\sigma}) = P(\tfrac{c-\mu}{\sigma} < Z \leq \tfrac{d-\mu}{\sigma})$$



斜線部の面積はどれも同じ

標準化しても確率が同じことの別説明 (置換積分)

$X \sim \mathrm{N}(\mu, \sigma^2)$ のとき, 積分で $z = \frac{x-\mu}{\sigma}$, $\mathrm{d}z = \frac{1}{\sigma}\mathrm{d}x$ とすると,

$$
\begin{aligned}
P(c < X \leq d) &= \int_c^d \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x \\
&= \int_{\frac{c-\mu}{\sigma}}^{\frac{d-\mu}{\sigma}} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{z^2}{2}} \, \sigma \mathrm{d}z \\
&= P\left(\frac{c-\mu}{\sigma} < Z \leq \frac{d-\mu}{\sigma}\right)
\end{aligned}
$$

L06-Q2

Quiz(正規分布の確率)

確率変数 $X \sim \mathrm{N}(3, 2^2)$, $Z \sim \mathrm{N}(0, 1^2)$ とする.

1. X, Z の確率密度関数の式を書こう. グラフを重ねて描こう.
2. 母期待値 $\mathrm{E}[X^2]$ を求めよう.
3. 確率 $P(X > 5) = P(c < Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
4. 確率 $P(+1 < X \leq 7) = P(c \leq Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
5. 確率 $P(+3 < X \leq 9) = P(c \leq Z \leq d)$ となるように c,d を定めよう. また, 確率を Python や表を使って小数で求めよう.
6. 確率 $P(X \leq d) = 3/4$ となる d を, 確率 $P(Z \leq d') = 3/4$ となるような d' で表そう. また, Python や表を使って小数で求めよう.

岩薩林 確率・統計 §4.5 例題 4.10,4.11, 問題 10, 第 4 章練習問題 5

ここまで来たよ

5 確率変数の標準化・標準正規分布・一般の正規分布

6 一般の正規分布・多次元の確率変数
- 一般の正規分布
- 2 次元の確率分布
- 母共分散
- 多次元の連続型確率分布

2 つの離散型確率変数の同時確率分布

高校 数学 B | 岩薩林 確率・統計 p.56

例 6 枚のカードから無作為に 1 枚引く. ♡7 ♡8 ♡9 ♢8 ♠9 ♣9

2 つの離散型確率変数の同時分布

$X =$ 数, $Y = 0$(赤札), 1(黒札) とすると (x, y) を得る確率は 2 変数の確率関数で書ける. 同時分布, 結合分布, joint distribution という.

$$p(x,y) = \begin{cases} \frac{1}{3} & ((x,y) = (8,0)) \\ \frac{1}{6} & ((x,y) = (9,0)) \\ \frac{1}{3} & ((x,y) = (9,1)) \\ \frac{1}{6} & ((x,y) = (7,0)) \\ 0 & (\text{他}) \end{cases}$$

表で書いた方がまし. ここでは,「他」は省略.

岩薩林 確率・統計 では, p_{xy} とも.

$P(X = x, Y = y)$ とも. $P($条件$)$ 型の表記.

$y \backslash x$	7	8	9	計
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	
1	0	0	$\frac{1}{3}$	
計				

定義 (同時分布の母期待値 岩薩林 確率・統計 (3.16)p.60)

$$\mathrm{E}[g(X,Y)] \overset{\text{定義}}{=} \sum_{x=-\infty}^{+\infty} \sum_{y=-\infty}^{+\infty} g(x,y) \cdot p(x,y)$$

定義 (同時分布の確率)

条件 $a < X \leq b$ かつ $c < Y \leq d$ で定まる確率は,

$$P(a < X \leq b \text{ かつ } c < Y \leq d) = \sum_{a<x\leq b} \sum_{c<y\leq d} p(x,y) \qquad \text{(P2)}$$

L06-Q3

Quiz(多変数の確率変数の期待値)

2 変数 X, Y の離散型確率分布を考える. 同時分布 $p(x, y)$ が下の表で与えられる.

$y \backslash x$	1	2	3
0	0	$\frac{2}{12}$	$\frac{1}{12}$
2	$\frac{4}{12}$	0	$\frac{5}{12}$

1. 母期待値 $\mathrm{E}[2X^2 + e^Y]$ を求めよう.
2. 確率 $P(XY \geq 2)$ を求めよう.
3. 周辺分布 $p_{\mathrm{X}}(x)$, $p_{\mathrm{Y}}(y)$ を求めよう.

命題 (2 次元の確率分布の母期待値の性質 高校 数学 B 岩薩林 確率・統計 (3.17)p.61)

$$\mathrm{E}[g_1(X,Y)+g_2(X,Y)] = \mathrm{E}[g_1(X,Y)] + \mathrm{E}[g_2(X,Y)] \qquad \text{(E5)}$$

$$\text{特に} \mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] \qquad \text{(E6,(3.17))}$$

証明

$$\begin{aligned}\mathrm{E}[g_1(X,Y)+g_2(X,Y)] &= \sum_x \sum_y (g_1(x,y)+g_2(x,y))p(x,y) \\ &= \mathrm{E}[g_1(X,Y)] + \mathrm{E}[g_2(X,Y)].\end{aligned}$$

周辺分布

定義 (確率変数の周辺分布)

同時分布 $p(x,y)$ に対して,
X の周辺分布 $p_{\mathrm{X}}(x)$, Y の周辺分布 $p_{\mathrm{Y}}(y)$ は,

$$p_{\mathrm{X}}(x) = p(x,\bullet) = \sum_{y} p(x,y), \quad p_{\mathrm{Y}}(y) = p(\bullet,y) = \sum_{x} p(x,y)$$

要するに 一方を無視した分布. 小計.

岩薩林 確率・統計 例題 3.4(p.57)

$y\backslash x$	7	8	9	計
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	
1	0	0	$\frac{1}{3}$	
計				

命題 (X だけ, Y だけの関数の母期待値)

x だけ, y だけの関数の母期待値は,
下の左辺=同時分布で計算しても
下の右辺=周辺分布で計算しても
同じ結果.

$$\mathrm{E}[g(X)] = \sum_x \sum_y g(x) \cdot p(x,y) = \sum_x g(x) \sum_y p(x,y) = \sum_x g(x) \cdot p_{\mathrm{X}}(x)$$

$$\mathrm{E}[g(Y)] = \sum_y \sum_x g(y) \cdot p(x,y) = \sum_y g(y) \sum_x p(x,y) = \sum_y g(y) \cdot p_{\mathrm{Y}}(y)$$

L06-Q4

Quiz

さっきの問で

1. 周辺分布を求めよう.
2. 周辺分布から $\mathrm{E}[2X^2 + \mathrm{e}^Y]$ を求めよう.
3. 母分散 $\mathrm{V}[X]$ を求めよう.

ここまで来たよ

5 確率変数の標準化・標準正規分布・一般の正規分布

6 一般の正規分布・多次元の確率変数
- 一般の正規分布
- 2 次元の確率分布
- 母共分散
- 多次元の連続型確率分布

母共分散 高校 数学 B

定義 (母共分散 covariance 岩薩林 確率・統計 (3.19)p.61,(3.20)p.62)

X, Y が確率変数で, $\mu_X = \mathrm{E}[X], \mu_Y = \mathrm{E}[Y]$ とおいたとき,

$$\text{母共分散}\mathrm{Cov}[X,Y] \overset{\text{定義}}{=} \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] \tag{3.19}$$

$$= \text{岩薩林 確率・統計 (3.20)} \cdots = \mathrm{E}[XY] - \mathrm{E}[X] \times \mathrm{E}[Y]. \quad \text{(C1,(3.20))}$$

L06-Q5

母共分散

さっきの問で

1. 母期待値 $\mathrm{E}[XY]$ を求めよう.
2. 母共分散 $\mathrm{Cov}[X,Y]$ を求めよう.

2 次元の確率分布の母期待値のよくある間違い

$$\mathrm{E}[g(X,Y)] \neq g(\mathrm{E}[X], \mathrm{E}[Y])$$

だって, $\sin(\log_x y) \neq \log_{\sin x}(\sin y)$ じゃん.
特に $\mathrm{E}[X \times Y] \neq \mathrm{E}[X] \times \mathrm{E}[Y]$. $\times$ を, $+$ と区別しよう.

2 次元の確率分布の母分散のよくある間違い

$$\mathrm{V}[X+Y] \neq \mathrm{V}[X] + \mathrm{V}[Y]$$

母分散を, 母期待値と区別しよう.

L06-Q6

母分散

さっきの問で母分散 $\mathrm{V}[X+Y]$ を求めよう.

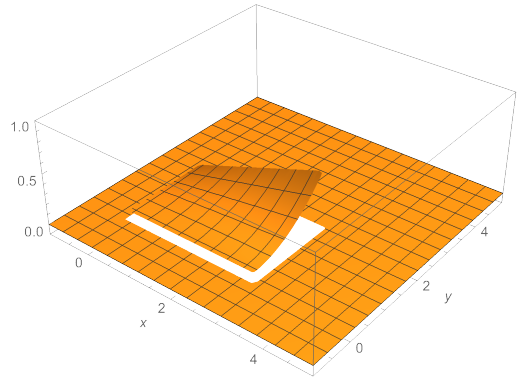
ここまで来たよ

5 確率変数の標準化・標準正規分布・一般の正規分布

6 一般の正規分布・多次元の確率変数
- 一般の正規分布
- 2 次元の確率分布
- 母共分散
- 多次元の連続型確率分布

多次元の連続型確率分布 岩薩林 確率・統計 §4.6

2 次元の同時確率密度関数 $f(x,y)$.
$f(x,y) \geq 0$, 大きいほど, その (x,y) が「出やすい」



定義 (連続型次元確率変数の母期待値 岩薩林 確率・統計 SS3.3,4.6)

離散型 $$\mathrm{E}[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot p(x,y) \tag{3.16}$$

連続型 $$\mathrm{E}[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) \cdot f(x,y)\, \mathrm{d}x\mathrm{d}y \tag{4.32}$$

確率 $P(a < X \leq b \text{かつ} c < Y \leq d) = \int_a^b (\int_c^d f(x,y)\mathrm{d}y)\mathrm{d}x$. 平面の領域上にある部分の体積.

条件付き確率とベイズの公式

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　L07(2023-05-22 Mon)

最終更新: Time-stamp: "2023-05-22 Mon 11:36 JST hig"

今日の目標

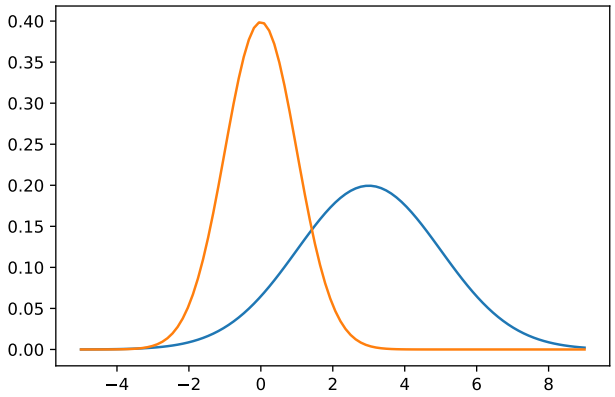
- 条件付き確率 岩薩林 確率・統計 §2.3, p.59, 同時確率分布, 周辺分布の間の計算ができる
- ベイズの定理 岩薩林 確率・統計 p.42 で, 2 つの条件付き確率の間の計算ができる

L06-Q1
L06-Q2

Quiz 解答: 正規分布の確率

X を標準化すると, $Z = \frac{X-3}{2}$.

1. $f_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi \cdot 2^2}} \mathrm{e}^{-\frac{(x-3)^2}{2\cdot 2^2}}$, $f_{\mathrm{Z}}(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}}$. X, Z の累積分布関数をそれぞれ $F_{\mathrm{X}}(x)$ =scipy.stats.norm(loc=2,scale=3).cdf(x), $F_{\mathrm{Z}}(z)$ =scipy.stats.norm(loc=0,scale=1).cdf(z) とかく.



2. $\mathrm{E}[X^2] = \mathrm{V}[X] + \mathrm{E}[X]^2 = 2^2 + 3^2$.

3. $P(X \geq 5) = P(5 \leq X < +\infty) = P(\frac{5-3}{2} \leq \frac{X-3}{2} < +\infty) = P(\frac{5-3}{2} \leq Z < +\infty) = 1 - F_{\mathrm{Z}}(1)$ =1-scipy.stats.norm(loc=0,scale=1).cdf(1) $(= 1 - F_{\mathrm{X}}(5)) = 0.15877.$
4. $P(1 \leq X \leq 7) = P(\frac{1-3}{2} \leq Z \leq \frac{7-3}{2}) = F_{\mathrm{Z}}(2) - F_{\mathrm{Z}}(-1) = 0.81859.$
5. $P(3 \leq X \leq 9) = P(\frac{3-3}{2} \leq Z \leq \frac{9-3}{2})$.
6. $d' = (d-3)/2$. $d' = (F_{\mathrm{Z}}^{-1}(3/4) =$ scipy.stats.norm(loc=0,scale=1).ppf(0.75).

L06-Q3

Quiz 解答: 多変数の確率変数の期待値

1. $p(x,y) =$

$y\backslash x$	1	2	3
0	0	$\frac{2}{12}$	$\frac{1}{12}$
2	$\frac{4}{12}$	0	$\frac{5}{12}$

$g(x,y) = 2x^2 + \mathrm{e}^y =$

$y\backslash x$	1	2	3
0	$2\cdot 1^2 + \mathrm{e}^0$	$2\cdot 2^2 + \mathrm{e}^0$	$2\cdot 3^2 + \mathrm{e}^0$
2	$2\cdot 1^2 + \mathrm{e}^2$	$2\cdot 2^2 + \mathrm{e}^2$	$2\cdot 3^2 + \mathrm{e}^2$

$\mathrm{E}[2X^2 + \mathrm{e}^Y] = (2\cdot 1^2 + \mathrm{e}^0)0 + (2\cdot 2^2 + \mathrm{e}^0)\frac{2}{12} + (2\cdot 3^2 + \mathrm{e}^0)\frac{1}{12}$
$+(2\cdot 1^2 + \mathrm{e}^2)\frac{4}{12} + (2\cdot 2^2 + \mathrm{e}^2)0 + (2\cdot 3^2 + \mathrm{e}^2)\frac{5}{12}$.

2. $\frac{4}{12} + 0 + \frac{5}{12} = \frac{9}{12}$.

3. $$p_{\mathrm{X}}(x) = \begin{cases} 4/12 & (x=1) \\ 2/12 & (x=2) \\ 6/12 & (x=3) \\ 0 & (\text{他}) \end{cases} \qquad p_{\mathrm{Y}}(y) = \begin{cases} 3/12 & (y=0) \\ 9/12 & (y=2) \\ 0 & (\text{他}) \end{cases}$$

4. (1 の別解) $\mathrm{E}[2X^2 + \mathrm{e}^Y] = 2\mathrm{E}[X^2] + \mathrm{E}[\mathrm{e}^Y] =$ 周辺分布で計算.

ここまで来たよ

6 一般の正規分布・多次元の確率変数

7 条件付き確率とベイズの公式
- 条件付き確率
- ベイズの定理
- ベイズ推定

同時確率分布と周辺分布 岩薙林 確率・統計 §3.3

X, Y 確率変数

同時確率分布

$p(x,y) = P(X = x \text{かつ} Y = y) =$

$y \backslash x$	150	160	165
45	3/8	0	1/12
50	1/8	1/3	1/12

コンマ前が X, 後が Y の値. 明示のため $p_{\mathrm{XY}}(x,y)$ と書くことも.

周辺確率分布 $p_{\mathrm{X}}(x) = P(X = x), p_{\mathrm{Y}}(y) = P(Y = y)$

意味 確率の小計, X の値を無視した Y の値だけの分布, その逆

確率変数が等しいとは「$X = 3Y + 15$」のとき,

$p(x,y) =$

$y \backslash x$	150	160	165
45	p	0	0
50	0	0	$1-p$

の形に限られる.

大注意 表の縦横, 変数名 X, Y には意味がなく, 問題などでは入れ替えたりする.

条件付き確率 岩薩林 確率・統計 §2.3

定義 (条件付き確率 岩薩林 確率・統計 (2.12))

条件 (事象) A のもとでの, 事象 B の条件付き確率

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

縦棒「|=given」の前後は不平等: P(確率を考える事象|条件の事象)
$P(B|A) \neq P(A|B), \neq P(A \cup B)$.

意味 A が起きる場合に限定した, B が起きる確率

$P(B|A)$ を P (of)B given A と発音する人がいる. given 条件=条件が与えられたとき, という言い方から.

縦棒の前だけ見る (縦棒の後を固定する) と, ただの確率. 全事象は 1.

離散型確率変数に対する条件付き確率

同時確率分布

$p(x,y) = P(X = x \text{かつ} Y = y) =$

$y \backslash x$	150	160	165
45	3/8	0	1/12
50	1/8	1/3	1/12

離散型確率変数に対する条件付き確率 岩薙林 確率・統計 (3.15)

条件 (事象)$Y = y$ のもとでの, 事象 $X = x$ の条件付き確率の記号
$P(X = x|Y = y) = p_{\mathrm{X|Y}}(x \text{ の値}|y \text{ の値})$

例 $p_{\mathrm{X|Y}}(150|45), p_{\mathrm{Y|X}}(45|150)$. 縦棒の前だけ見る (縦棒の後ろを固定しする) と, 前の変数についての, ただの確率分布. 意味 自分の言葉でどうぞ

命題 (条件付き確率を同時確率分布と周辺分布で表す式 岩薙林 確率・統計 (2.12))

$$P(X = x|Y = y) = p_{\mathrm{X|Y}}(x|y) = \frac{p(x,y)}{p_{\mathrm{Y}}(y)},$$

$$P(Y = y|X = x) = p_{\mathrm{Y|X}}(y|x) = \frac{p(x,y)}{p_{\mathrm{X}}(x)}.$$

L07-Q1

Quiz(条件付き分布)

次の 6 枚のカードから無作為に 1 枚のカードを引く.

♡7 ♡8 ♡9 ◇8 ♠9 ♣9

$X =$ 数, $Y = 0$(赤札), 1(黒札) とすると同時分布は次のようになる.

$y\backslash x$	7	8	9	計
0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	
1	0	0	$\frac{1}{3}$	
計				

1. 9 の札が出る条件のもとで赤札が出る, 条件付き確率を求めよう.
2. 赤札が出る条件のもとで 9 の札が出る, 条件付き確率を求めよう.

条件付き確率と周辺分布から，同時確率分布，周辺分布

命題 (同時確率分布, 周辺分布, 条件付き確率の関係)

乗法公式 (同時確率分布との関係)

岩薙林 確率・統計 (2.13)

$$\begin{aligned} p_{\mathrm{XY}}(x,y) =& p_{\mathrm{X}|\mathrm{Y}}(x|y) \times p_{\mathrm{Y}}(y) \\ =& p_{\mathrm{Y}|\mathrm{X}}(y|x) \times p_{\mathrm{X}}(x). \end{aligned}$$

全確率の法則 (周辺分布との関係)

岩薙林 確率・統計 (2.17),(2.20)

$$p_{\mathrm{X}}(x) = \sum_{y} p_{\mathrm{X}|\mathrm{Y}}(x|y) \times p_{\mathrm{Y}}(y),$$

$$p_{\mathrm{Y}}(y) = \sum_{x} p_{\mathrm{Y}|\mathrm{X}}(y|x) \times p_{\mathrm{X}}(x)$$

L07-Q2 岩薙林 確率・統計 例題 3.3,3.5,3 章練習問題 2

Quiz(同時分布・条件付き分布・周辺分布)

離散型確率変数 X は値 $2,3$, Y は値 $20,30$ をとる. 次の通りとする.

$$p_{\mathrm{Y}}(y) = \begin{cases} \frac{1}{3} & (y=20) \\ \frac{2}{3} & (y=30) \\ 0 & (他) \end{cases}$$

$$p_{\mathrm{X|Y}}(x|20) = \begin{cases} \frac{3}{4} & (x=2) \\ \frac{1}{4} & (x=3) \end{cases}, \quad p_{\mathrm{X|Y}}(x|30) = \begin{cases} \frac{1}{2} & (x=2) \\ \frac{1}{2} & (x=3) \end{cases}.$$

1. 同時分布を求めよう
2. X の周辺分布を求めよう.
3. 条件付き確率 $p_{\mathrm{Y|X}}(20|2), p_{\mathrm{Y|X}}(30|3)$ を求めよう.

最後の問は次のセクションに延期.

ここまで来たよ

6 一般の正規分布・多次元の確率変数

7 条件付き確率とベイズの公式
- 条件付き確率
- ベイズの定理
- ベイズ推定

例えばこんな問 I

L07-Q3

Quiz(ベイズの定理)

外見で区別できない, 甘い品種 1 と渋い品種 2 の柿がある.
甘い品種 1 は, 確率 0.95 で赤に, 確率 0.05 で黄色になる.
渋い品種 2 は, 確率 0.125 で赤に, 確率 0.875 で黄色になる.

1. かごの柿の $1/5$ が甘い柿である. いま, 無作為に 1 個の柿を取りだしたところ, 赤い柿だった. 取り出した赤い柿が甘い確率を求めよう.

ベイズの定理 I

定理 (ベイズの定理 岩薩林 確率・統計 (2.18),(2.19))

$$p_{\mathrm{X}|\mathrm{Y}}(x|y) \overset{\text{定義}}{=} \frac{p(x,y)}{p_{\mathrm{Y}}(y)} = \frac{p_{\mathrm{Y}|\mathrm{X}}(y|x)p_{\mathrm{X}}(x)}{\sum_{x'} p_{\mathrm{Y}|\mathrm{X}}(y|x')p_{\mathrm{X}}(x')},$$

$$p_{\mathrm{Y}|\mathrm{X}}(y|x) \overset{\text{定義}}{=} \frac{p(x,y)}{p_{\mathrm{X}}(x)} = \frac{p_{\mathrm{X}|\mathrm{Y}}(x|y)p_{\mathrm{Y}}(y)}{\sum_{y'} p_{\mathrm{X}|\mathrm{Y}}(x|y')p_{\mathrm{Y}}(y')}.$$

$p_{\mathrm{X}|\mathrm{Y}}(x|y)$ を $p_{\mathrm{Y}|\mathrm{X}}(y|x')$ (と $p_{\mathrm{X}}(x')$) で書き表す式,
$p_{\mathrm{Y}|\mathrm{X}}(y|x)$ を $p_{\mathrm{X}|\mathrm{Y}}(x|y')$ (と $p_{\mathrm{Y}}(y')$) で書き表す式.
定義の分子分母を, 乗法公式と全確率の公式で書き直しただけ (なので余分の記憶は不要).

L07-Q4

ベイズの定理

さっきの問で, 条件付き確率 $p_{\mathrm{Y}|\mathrm{X}}(20|2), p_{\mathrm{Y}|\mathrm{X}}(30|3)$ を求めよう.

ここまで来たよ

6 一般の正規分布・多次元の確率変数

7 条件付き確率とベイズの公式
- 条件付き確率
- ベイズの定理
- ベイズ推定

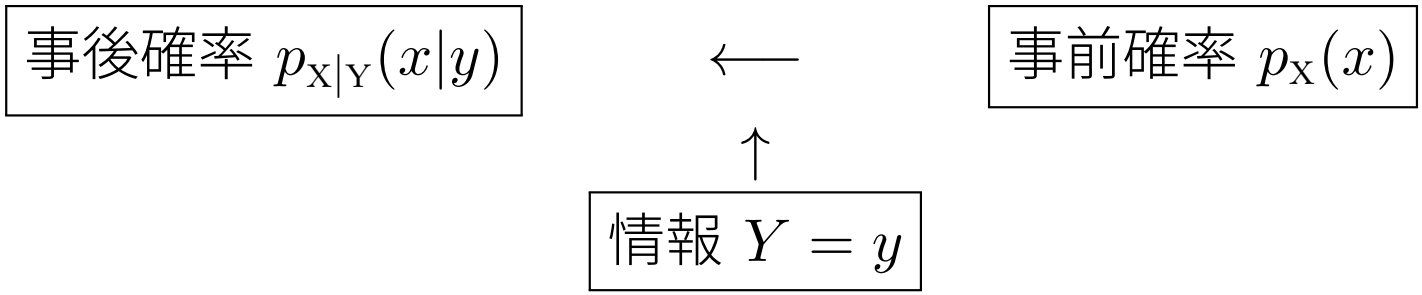
L07-Q5

Quiz(ベイズ推定)

袋に何個かの色つきボールが入っている. ボールを割ると, 当たり外れの記された紙がでてくる.
当たりのボールのうち赤いボールが $\frac{1}{10}$, 白いボールが $\frac{9}{10}$,
外れのボールのうち赤いボールが $\frac{7}{10}$, 白いボールが $\frac{3}{10}$ とわかっている
袋の中の当たりのボールは, $\frac{2}{10}$ とわかっている. ボールを 1 個取り出したときの当たりの確率 (事前確率) は $\frac{2}{10}$ である.
ボール 1 個を取り出したところ, 赤いボールだった. このとき, 当たりの確率 (事後確率) を求めよう.

岩薩林 確率・統計 例題 2.5, 例題 2.6,2 章練習問題 4,2 章練習問題 5

ベイズ推定の考え方



追加の情報が得られるたびに, 事後確率は正確になっていく.

主観確率事前確率に主観を許す考え方. 事後確率にも主観が含まれる.

L07-Q6

Quiz(ベイズ推定)

ある病気の人の割合は全体の 0.005 と思われている.
検査では, 病気の人の 0.99 は陽性となり (真陽性), 0.01 の人は陰性になる (偽陰性). また, 病気でない人の 0.02 は (誤って) 陽性となり (偽陽性), 0.98 の人は陰性になる (真陽性).

1.

1 回の検査で陽性となった場合, その人が病気である確率を求めよう.
2 回検査して 2 回とも陽性の場合は?

確率変数の独立, 独立同分布

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　L08(2023-05-29 Mon)

最終更新: Time-stamp: "2023-05-29 Mon 11:26 JST hig"

今日の目標

- 確率変数の独立性を判定し利用できる　岩薩林 確率・統計 §3.3
- 独立同分布の和の母平均値/母分散を計算できる

岩薩林 確率・統計 p.113, 定理 5.2

L07-Q1

Quiz 解答: 条件付き分布

1. $p_{\mathrm{Y}|\mathrm{X}}(0|9) = \dfrac{p(9,0)}{p_{\mathrm{X}}(9)} = \dfrac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}} = \dfrac{1}{3}.$
2. $p_{\mathrm{X}|\mathrm{Y}}(9|0) = \dfrac{p(9,0)}{p_{\mathrm{Y}}(0)} = \dfrac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}+\frac{1}{6}} = \dfrac{1}{4}.$

L07-Q2,Q4

Quiz 解答: 同時分布・条件付き分布・周辺分布

$y\backslash x$	2	3	
20	3/12	1/12	4/12
30	4/12	4/12	8/12
	7/12	5/12	1

L07-Q3

L07-Q5

Quiz 解答: ベイズ推定

Y を色, X を当たり外れとすると,

$$
\begin{aligned}
&P(X = \text{当}|Y = \text{赤}) \\
=&\frac{P(Y = \text{赤}|X = \text{当})P(X = \text{当})}{P(Y = \text{赤}|X = \text{当})P(X = \text{当}) + P(Y = \text{赤}|X = \text{外})P(X = \text{外})} \\
=&\frac{\frac{1}{10}\frac{2}{10}}{\frac{1}{10}\frac{2}{10}+\frac{7}{10}\frac{8}{10}} = \frac{2}{58}.
\end{aligned}
$$

$Y\backslash X$	当たり	外れ
赤	$\frac{1}{10}\cdot\frac{2}{10}$	$\frac{7}{10}\cdot\frac{8}{10}$
白	$\frac{9}{10}\cdot\frac{2}{10}$	$\frac{3}{10}\cdot\frac{8}{10}$
合計	$\frac{2}{10}$	$\frac{8}{10}$

L07-Q6

Quiz 解答: ベイズ推定

検査 $Y = \begin{cases} 80 & (\text{陽性}) \\ 20 & (\text{陰性}) \end{cases}$, 病状 $X = \begin{cases} 100 & (\text{病気}) \\ 0 & (\text{病気でない}) \end{cases}$ で表す.

1. $$p_{\mathrm{X}}(100) = 0.005, p_{\mathrm{Y}|\mathrm{X}}(80|100) = 0.99, p_{\mathrm{Y}|\mathrm{X}}(80|0) = 0.02$$

同時確率分布

$x \backslash y$	80	20	計
100	真陽性 0.99×0.005	偽陰性 0.01×0.005	0.005
0	偽陽性 0.02×0.995	真陰性 0.98×0.995	0.995

$$\begin{aligned} p_{\mathrm{X}|\mathrm{Y}}(100|80) &= \frac{p_{\mathrm{Y}|\mathrm{X}}(80|100) \times p_{\mathrm{X}}(100)}{p_{\mathrm{Y}|\mathrm{X}}(80|100) \times p_{\mathrm{X}}(100) + p_{\mathrm{Y}|\mathrm{X}}(80|0) \times p_{\mathrm{X}}(0)} \\ &= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.02 \times 0.995} = 0.199185. \end{aligned}$$

2. $0.99^2 \cdot 0.005 / (0.99^2 \cdot 0.005 + 0.02^2 \cdot 0.995) = 0.924884.$

ここまで来たよ

7 条件付き確率とベイズの公式

8 確率変数の独立, 独立同分布
- 確率変数の独立性
- 独立同分布

確率変数が「同じ」対「同分布」

2 つの確率変数 X, Y.

X, Y は同じ　定義:$X = Y$, 各試行で同じ値.

- X, Y が同じ, ならば, 同分布 (次)
- 例: コイン 1 枚 (表 (1) の確率 $0 < p < 1$) を 1 回投げた結果に, 2 つの変数名 X, Y をつけた
- もちろん周辺分布は等しい
  $\mathrm{E}[g(X)] = \mathrm{E}[g(Y)]$.

$y \backslash x$	0	1
0	$1-p$	0
1	0	p

X, Y は同分布　定義: 周辺分布が等しい $p_{\mathrm{X}}(a) = p_{\mathrm{Y}}(a)$.

- 「同じ」とはかぎらない
- $\mathrm{E}[g(X)] = \mathrm{E}[g(Y)]$.

$y \backslash x$	0	1
0	$1-\frac{3}{2}p$	$\frac{1}{2}p$
1	$\frac{1}{2}p$	$\frac{1}{2}p$

X, Y は独立同分布

例: 形の同じ青色コイン X(表 p) と赤色コイン Y(表 q). $p = q$

$y \backslash x$	0	1	
0	$(1-p)(1-q)$	$p(1-q)$	$1-q$
1	$(1-p)q$	pq	q
	$1-p$	p	1

例

$y \backslash x$	0	1	
0	4/9	2/9	2/3
1	2/9	1/9	1/3
	2/3	1/3	1

独立: この形 ($p \neq q$ も可) $\rightsquigarrow$ カモ

同分布: $p = q$

- X の確率が Y によらない. Y の確率が X によらない.
  - 条件付き分布 $p_{\mathrm{Y}|\mathrm{X}}(y|x) = p_{\mathrm{Y}}(y)$. x の条件によらない.
  - 条件付き分布 $p_{\mathrm{X}|\mathrm{Y}}(x|y) = p_{\mathrm{X}}(x)$. y の条件によらない.

確率変数の独立性 高校 数学 B

定義 (事象の独立性 岩薩林 確率・統計 (2.14)p.40)

事象 A, B が独立 (independent) とは次が成立すること.

$$P(A \cap B) = P(A) \times P(B)$$

定義 (確率変数の独立性 岩薩林 確率・統計 (3.14)p.58)

確率変数 X, Y が独立 (independent) とは次が成立すること.

$$p(x, y) = p_{\mathrm{X}}(x) \times p_{\mathrm{Y}}(y)$$

X, Y が独立とは

- すべての行 (周辺分布も) がベクトルとして平行. すべての列 (周辺分布も) がベクトルとして平行
- 条件付き確率が周辺分布に等しい
  $p_{\mathrm{X}|\mathrm{Y}}(x|\text{何でも}) = p_{\mathrm{X}}(x)$, $p_{\mathrm{Y}|\mathrm{X}}(y|\text{何でも}) = p_{\mathrm{Y}}(y)$.
- ベイズ推定は役立たない. 事前確率=事後確率.
- X, Y が互いに「無関係」であること
- 同時分布が, 周辺分布だけから (積で) 決まっちゃうこと

独立である例

例

$y\backslash x$	0	1
10	$4/9$	$2/9$
11	$2/9$	$1/9$

例

赤白のサイコロを各 1 個振って, $X =$ 赤の目, $Y =$ 白の目.

例

下のカードから引いて, $X =$ 色, $Y =$ 数.

♡8　♢8　♢9

♠8　♣8　♣9

独立でない例

サイコロで, 目が 1–4 なら $X = 0$, 目が 5,6 なら $X = 1$.

例 1 $X = Y$. 確率変数 X, Y は同じ.

$y \backslash x$	0	1
0	$2/3$	0
1	0	$1/3$

例 2 例 1 の X, 同じサイコロの目で，1–2 なら $Y = 1$, 3–6 なら $Y = 0$.

$y \backslash x$	0	1
0	$1/3$	$1/3$
1	$1/3$	0

命題 (独立と限らない 2 次元の確率分布の母期待値の性質)

高校 数学 B | 岩薩林 確率・統計 (3.17)p.61 | 確率統計 I(2023)L06

$$\mathrm{E}[g_1(X,Y)+g_2(X,Y)] = \mathrm{E}[g_1(X,Y)] + \mathrm{E}[g_2(X,Y)] \quad \text{(E5)}$$

$$\text{特に} \mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y] \quad \text{(E6,(3.17))}$$

$$\mathrm{E}[g_1(X) \times g_2(Y)] \neq \mathrm{E}[g_1(X)] \times \mathrm{E}[g_2(Y)] \quad \text{(8.1)}$$

$$\text{母共分散} \mathrm{Cov}[X,Y] \overset{\text{定義}}{=} \mathrm{E}[(X-\mu_{\mathrm{X}})(Y-\mu_{\mathrm{Y}})] \quad \text{(3.19)}$$

$$= \mathrm{E}[XY] - \mathrm{E}[X] \times \mathrm{E}[Y]. \quad \text{(C1,(3.20))}$$

L08-Q1

Quiz(2 次元の独立でない確率変数の母期待値)

確率変数 X, Y について, $\mathrm{E}[X] = 3, \mathrm{E}[X^2] = 16, \mathrm{E}[Y] = 10, \mathrm{E}[Y^2] = 102, \mathrm{E}[XY] = 25$ が成立する. 次の量を求めよう.

1. $\mathrm{E}[(X+2)(Y+3)]$
2. $\mathrm{Cov}[X, Y]$

命題 (X, Y が独立のとき成立するいい性質)

岩薩林 確率・統計 (3.18)p.61

$$\mathrm{E}[g_1(X) \times g_2(Y)] = \mathrm{E}[g_1(X)] \times \mathrm{E}[g_2(Y)] \quad \text{(IE1)}$$

$$\text{特に} \mathrm{E}[XY] = \mathrm{E}[X] \times \mathrm{E}[Y] \quad \text{(IE2,(3.18))}$$

$$0 = \mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X] \times \mathrm{E}[Y], \quad \text{(IC1,(3.21))}$$

$$\mathrm{V}[g_1(X) + g_2(Y)] = \mathrm{V}[g_1(X)] + \mathrm{V}[g_2(Y)] \quad \text{(IC2)}$$

岩薩林 確率・統計 (3.21)p.63

$$\text{特に} \mathrm{V}[X + Y] = \mathrm{V}[X] + \mathrm{V}[Y] \quad \text{(IC3,(3.21))}$$

岩薩林 確率・統計 (3.21)p.63

$$\mathrm{V}[aX + bY + c] = a^2 \mathrm{V}[X] + b^2 \mathrm{V}[Y] \quad \text{(IC4)}$$

Cov は「独立でない度」，ただし，
IC1, IC3,IC4 は, 独立でなくても $\mathrm{Cov}[X, Y] = 0$ だけで成り立つ.
$\mathrm{Cov}[X, Y] = 0$ は独立の **必要条件** (だが **十分条件** ではない)

一部の証明連続型で書くけど, $\int \to \sum$ で離散型でも同様.

$$
\begin{aligned}
\mathrm{E}[g_1(X)\cdot g_2(Y)] &= \iint g_1(x)g_2(y)\cdot f(x,y)\mathrm{d}y\,\mathrm{d}x \\
&\overset{\text{独立}}{=} \iint g_1(x)g_2(y)\cdot f_{\mathrm{X}}(x)\times f_{\mathrm{Y}}(y)\mathrm{d}y\,\mathrm{d}x \\
&= \int g_1(x)f_{\mathrm{X}}(x)\left(\int g_2(y)\cdot f_{\mathrm{Y}}(y)\mathrm{d}y\right)\mathrm{d}x \\
&= \mathrm{E}[g_1(X)]\times \mathrm{E}[g_2(Y)]
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{V}[X+Y] &\overset{\mathsf{V1}}{=} \mathrm{E}[(X+Y)^2]-\mathrm{E}[X+Y]^2 \\
&\overset{\mathsf{E6}}{=} \mathrm{E}[X^2]+2\mathrm{E}[XY]+\mathrm{E}[Y^2]-(\mathrm{E}[X]^2+2\mathrm{E}[X]\mathrm{E}[Y]+\mathrm{E}[Y]^2) \\
&\overset{\mathsf{V1},\mathsf{C1}}{=} \mathrm{V}[X]+2\mathrm{Cov}[X,Y]+\mathrm{V}[Y] \\
&\overset{\mathsf{IC1}}{=} \mathrm{V}[X]+0+\mathrm{V}[Y]
\end{aligned}
$$

L08-Q2

Quiz(独立な確率変数の母期待値)

独立な確率変数 X, Y を考える.
$\mathrm{E}[X] = 2, \mathrm{E}[Y] = 3, \mathrm{V}[X] = 5, \mathrm{V}[Y] = 11$ である.

1. 母期待値 $\mathrm{E}[XY]$ を求めよう.
2. 母共分散 $\mathrm{Cov}[X, Y]$ を求めよう.
3. $\mathrm{E}[(-2X + 3Y)(X + 5Y)]$ を求めよう.
4. $\mathrm{V}[-2X + 3Y]$ を求めよう.

岩薩林 確率・統計 問題 6(p.64)

岩薩林 確率・統計 第 3 章練習問題 2(p.73)

岩薩林 確率・統計 第 3 章練習問題 5(p.73)

ここまで来たよ

7 条件付き確率とベイズの公式

8 確率変数の独立, 独立同分布
- 確率変数の独立性
- 独立同分布

独立同分布

定義 (独立同分布 (i.i.d.) 岩薩林 確率・統計 定理 4.2(p.87) の仮定,p.113)

離散型/連続型確率変数 $X_1, X_2, \ldots, X_n$ が, たがいに独立で, 同分布 (同じ周辺分布 $p_{\mathrm{X}_i}(x_i) = p(x)$ を持つとする.
このとき $X_1, \ldots, X_n$ は独立同分布に従う (i.i.d.=independent and identically-distributed) という.

母期待値も共通になる. $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$.
例: 同じ形の n 色のサイコロ. X_1=青色, X_2=緑色, $\cdots$.

L08-Q3

Quiz(同分布の確率母期待値母分散)

X_1, X_2 は同分布にしたがう確率変数で，コインの表 (1) 裏 (0) を表す．確率関数は，

$$p_{\mathrm{X}_i}x_i = \begin{cases} 1-p & (x=0) \\ p & (x=1) \end{cases},$$

母平均値・母分散は

$$\mathrm{E}[X_i^k] = 0^k \cdot (1-p) + 1^k \cdot p = p, \mathrm{V}[X_i] = \mathrm{E}[X_i^2] - \mathrm{E}[X_i]^2 = p - p^2.$$

1. $X_1 = X_2$ であるとき (1 枚のコインに 2 個の名前をつけたとき),
   1. 確率 $P(X_1 = X_2 = 1)$ を求めよう.
   2. 母平均値 $\mathrm{E}[X_1 + X_2]$ を求めよう.
   3. 母分散 $\mathrm{V}[X_1 + X_2]$ を求めよう.
   4. 母共分散 $\mathrm{Cov}[X_1 + X_2]$ を求めよう.
2. X_1, X_2 が独立であるとき (形は同じだが色の違う別々のコインであるとき),
   1. 確率 $P(X_1 = X_2 = 1)$ を求めよう.
   2. 母平均値 $\mathrm{E}[X_1 + X_2]$ を求めよう.
   3. 母分散 $\mathrm{V}[X_1 + X_2]$ を求めよう.
   4. 母共分散 $\mathrm{Cov}[X_1 + X_2]$ を求めよう.

L08-Q4

Quiz(独立同分布の母期待値母分散)

確率変数 X_1, X_2, X_3 は独立同分布に従い, $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$ とする.
次の母期待値, 母分散を μ と σ^2 で表そう.

1. $\mathrm{E}[X_1 + X_1 + X_1]$
2. $\mathrm{E}[X_1 + X_2 + X_3]$
3. $\mathrm{E}[X_1 + X_1 + X_2]$
4. $\mathrm{V}[X_1 + X_1 + X_1]$
5. $\mathrm{V}[X_1 + X_2 + X_3]$
6. $\mathrm{V}[X_1 + X_1 + X_2]$
7. $\mathrm{E}[(X_1 + X_2 + X_3)^2]$

二項分布, 独立同分布の和

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　L09(2023-06-05 Mon)

最終更新: Time-stamp: "2023-06-05 Mon 07:43 JST hig"

今日の目標

- 二項分布の母期待値, 母分散, 確率を求められる
  岩薩林 確率・統計 §3.4
- 独立同分布にしたがう確率変数の和の母平均値，母分散を求められる 岩薩林 確率・統計 §4.4

L08-Q1

Quiz 解答:2 次元の独立でない確率変数の母期待値

1. $\mathrm{E}[(X+2)(Y+3)] = \mathrm{E}[XY+3X+2Y+6] =$
   $\mathrm{E}[XY]+3\mathrm{E}[X]+2\mathrm{E}[Y]+6\mathrm{E}[1] = 25+9+20+6 = 60.$
2. $\mathrm{Cov}[X,Y] = \mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y] = 25-3\cdot 10 = -5.$
   $\mathrm{Cov}[X,Y] = \mathrm{E}[(X-\mu_{\mathrm{X}})(Y-\mu_{\mathrm{Y}})]$ を展開しても同じ結果.

L08-Q2

Quiz 解答: 独立な確率変数の母期待値

1. X, Y は独立なので $\mathrm{E}[XY] \overset{\mathsf{IE1}}{=} \mathrm{E}[X]\mathrm{E}[Y] = 2\cdot 3.$
2. X, Y は独立なので $\mathrm{Cov}[X,Y] \overset{\mathsf{IC1}}{=} 0.$
3. $\mathrm{E}[(-2X+3Y)(X+5Y)] \overset{\mathsf{E5}}{=} \mathrm{E}[-2X^2]+\mathrm{E}[-7XY]+\mathrm{E}[15Y^2] \overset{\mathsf{V1,IE1}}{=}$
   $-2(\mathrm{V}[X]+\mathrm{E}[X]^2)-7\mathrm{E}[X]\mathrm{E}[Y]+15(\mathrm{V}[Y]+\mathrm{E}[Y]^2) = 240.$

4. X, Y は独立なので $\mathrm{V}[g_1(X)+g_2(Y)] \overset{\text{IC2}}{=} \mathrm{V}[g_1(X)]+\mathrm{V}[g_2(Y)]$ であることに注意して,
$\mathrm{V}[-2X+3Y] \overset{\text{IC2}}{=} \mathrm{V}[-2X]+\mathrm{V}[3Y] \overset{\text{V2}}{=} (-2)^2\mathrm{V}[X]+3^2\mathrm{V}[Y] = 119.$

L08-Q3

Quiz 解答: 同分布の確率母期待値母分散

1.

$x_2 \backslash x_1$	0	1
0	$1-p$	0
1	0	p

x_1+x_2	確率
0	$1-p$
1	0
2	p

1. $P(X_1 = X_2 = 1) = p.$
2. $\mathrm{E}[X_1+X_2] = \mathrm{E}[X_1]+\mathrm{E}[X_2] = 2p.$
3. $\mathrm{V}[X_1+X_2] = \mathrm{E}[(X_1+X_2)^2]-\mathrm{E}[X_1+X_2]^2 = 2^2p-(2p)^2 = 4p(1-p).$
4. $\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1X_2]-\mathrm{E}[X_1]\mathrm{E}[X_2] = p-p^2 = p(1-p) = \mathrm{V}[X_1].$

2

$x_2 \backslash x_1$	0	1
0	$(1-p)^2$	$p(1-p)$
1	$(1-p)p$	p^2

$x_1 + x_2$	確率
0	$(1-p)^2$
1	$2p(1-p)$
2	p^2

1. $P(X_1 = X_2 = 1) = p_{X_1} x_1 \times p_{X_2} x_2 = p^2$.
2. $\mathrm{E}[X_1 + X_2] = \mathrm{E}[X_1] + \mathrm{E}[X_2] = 2p$.
3. $\mathrm{V}[X_1 + X_2] = \mathrm{V}[X_1] + \mathrm{V}[X_2] = 2p(1-p)$.
4. $\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1 X_2] - \mathrm{E}[X_1]\mathrm{E}[X_2] = p^2 - p^2 = 0$.

ここまで来たよ

8 確率変数の独立, 独立同分布

9 二項分布, 独立同分布の和
- 二項分布
- 独立同分布にしたがう確率変数の和
- ベルヌーイ分布と二項分布・独立同分布の性質の応用
- チェビシェフの不等式

復習+ちょっと (x で書かれた確率関数) I

L09-Q1

Quiz(離散的な確率変数の母平均値・母分散・母標準偏差・確率)

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.

$$p(x) = \begin{cases} \frac{x}{55} & (0 \leq x \leq 10) \\ 0 & (\text{他}) \end{cases}$$

1. 確率 $P(X \leq 5)$ 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.

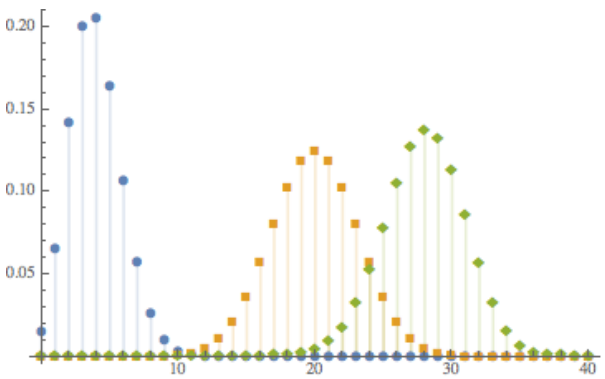
二項分布 高校 数学 B 岩薩林 確率・統計 §3.4

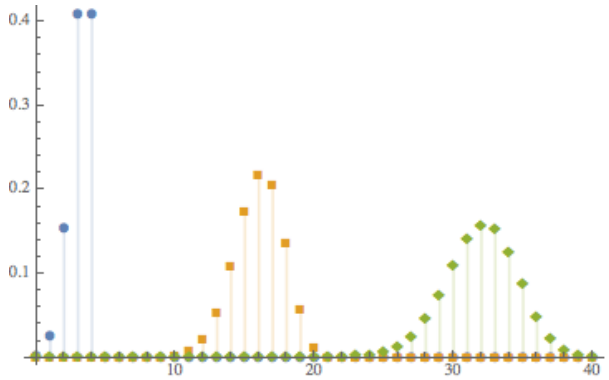
定義 (二項分布 岩薩林 確率・統計 (3.24)p.66)

離散型確率変数 X が次の確率分布を持つとき, X はパラメタ n, p の二項分布 $\mathrm{B}(n,p)$ にしたがうという.

$$p(x) = \begin{cases} \frac{n!}{x!(n-x)!}\, p^x (1-p)^{n-x} & (x = 0, 1, 2, 3, \ldots, n) \\ 0 & (\text{他}) \end{cases}$$

意味: 確率 p で表の出るコインを n 回投げたとき, x 回表が出る確率.





$${}_n\mathrm{C}_x = \binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$\mathrm{B}(40, 0.1), \mathrm{B}(40, 0.5), \mathrm{B}(40, 0.7),\ \mathrm{B}(4, 0.8), \mathrm{B}(20, 0.8), \mathrm{B}(40, 0.8)$

二項分布 $\mathrm{B}(n,p)$ の Python での扱い

Google Colab on Moodle

```
from scipy import stats

# 二項分布 = binomial distribution
# scipy.stats.binom(n=10,p=0.3) 確率分布名=binom(n,p)
rvx=stats.binom(n=10,p=0.3)

# 確率(質量)関数 probability mass function 離散型に使う用語
rvx.pmf(x) # =p(x)

# 累積分布関数
rvx.cdf(x)

# 母平均値
rvx.mean()
```

命題 (二項分布の母平均値と母分散 岩薩林 確率・統計 定理 3.1(p.70))

$$\mathrm{E}[X] = np, \mathrm{V}[X] = np(1-p)$$

(証明延期)

$\mathrm{E}[1] =$ 岩薩林 確率・統計 式 (3.26)–(3.27)

定理 (二項定理 高校 数学 A 岩薩林 確率・統計 (3.25)p.68)

$$(a+b)^n = \sum_{x=0}^{n} {}_n\mathrm{C}_x a^x b^{n-x}$$

L09-Q2

Quiz(二項分布)

確率 $p = \frac{2}{3}$ で表のでるコインを 100 回投げる.

1. 表がでる回数 X は確率変数である. X のしたがう確率分布を答えよう.
2. 表が 40 回でる確率を求めよう. 階乗 $n!$ とべき乗 a^b と分数 $\frac{a}{b}$ は簡単化・約分しなくてよい. P,C などの記号は使わないで答えること.
3. 表がでる回数の母平均値, 母分散を求めよう.

L09-Q3

Quiz(二項分布または独立同分布)

あるスーパーのおでんセット 1 パックには, 確率 $\frac{7}{10}$ で卵が 2 個, 確率 $\frac{3}{10}$ で卵が 0 個 (!) 入っている. おでんセットを 10 パックを買ったときに得られる卵の合計の個数を確率変数 Y とする. 各パックの卵の個数は独立とする.

1. 10 パック中 X パックが卵 2 個入りとする. Y を X で表そう. X のしたがう確率分布を答えよう.
2. 母平均値 $\mathrm{E}[Y]$ を求めよう.
3. 母分散 $\mathrm{V}[Y]$ を求めよう.
4. 確率 $P(Y = 16)$ を求めよう.

ここまで来たよ

8 確率変数の独立, 独立同分布

9 二項分布, 独立同分布の和
- 二項分布
- 独立同分布にしたがう確率変数の和
- ベルヌーイ分布と二項分布・独立同分布の性質の応用
- チェビシェフの不等式

独立同分布

$X = X_1, Y = X_2$ は独立同分布 定義: 独立かつ同分布

例: 形の同じ青色コイン X_1 と赤色コイン X_2.
表がでる (1) の確率 p

$x_2 \backslash x_1$	0	1
0	$(1-p)^2$	$p(1-p)$
1	$p(1-p)$	p^2

定義 (独立同分布 (i.i.d.) 岩薩林 確率・統計 定理 4.2(p.87) の仮定,p.113 **)**

離散型/連続型確率変数 $X_1, X_2, \ldots, X_n$ が, たがいに独立で, すべて同分布に従う (同じ周辺分布 $p_{X_i}(x_i) = p(x)$ を持つ) とする.
このとき $X_1, \ldots, X_n$ は独立同分布に従う (i.i.d.=independent and identically-distributed) という.

母期待値も共通になる. $\mathrm{E}[X_i] = \mu, \mathrm{V}[X_i] = \sigma^2$.
例: 同じ形の n 色のサイコロ. X_1=青色, X_2=緑色, $\cdots$.

L09-Q4

Quiz(同分布の確率母期待値母分散)

X_1, X_2 は同分布にしたがう確率変数で，コインの表 (1) 裏 (0) を表す．確率関数は，

$$p_{\mathrm{X}_i}(x_i) = \begin{cases} 1-p & (x=0) \\ p & (x=1) \end{cases},$$

母平均値・母分散は

$$\mathrm{E}[X_i^k] = 0^k \cdot (1-p) + 1^k \cdot p = p, \mathrm{V}[X_i] = \mathrm{E}[X_i^2] - \mathrm{E}[X_i]^2 = p - p^2.$$

1. $X_1 = X_2$ であるとき (1 枚のコインに 2 個の名前をつけたとき),
   1. 確率 $P(X_1 = X_2 = 1)$ を求めよう.
   2. 母平均値 $\mathrm{E}[X_1 + X_2]$ を求めよう.
   3. 母分散 $\mathrm{V}[X_1 + X_2]$ を求めよう.
   4. 母共分散 $\mathrm{Cov}[X_1 + X_2]$ を求めよう.
2. X_1, X_2 が独立であるとき (形は同じだが色の違う別々のコインであるとき),
   1. 確率 $P(X_1 = X_2 = 1)$ を求めよう.
   2. 母平均値 $\mathrm{E}[X_1 + X_2]$ を求めよう.
   3. 母分散 $\mathrm{V}[X_1 + X_2]$ を求めよう.
   4. 母共分散 $\mathrm{Cov}[X_1 + X_2]$ を求めよう.

独立同分布にしたがう確率変数の和の性質

岩薙林 確率・統計 例題 4.6

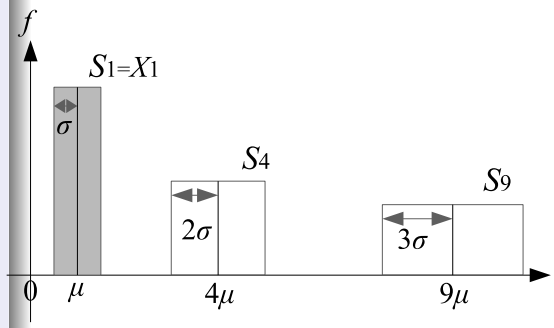
命題 (i.i.d にしたがう確率変数の和)

$X_1, \ldots, X_n$: i.i.d. 母平均値 $\mathrm{E}[X_i] = \mu$, 母分散 $\mathrm{V}[X_i] = \sigma^2$.
和の確率変数 $S_n = X_1 + \cdots + X_n$.

$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] \overset{\text{同分布}}{=} n \times \mu.$$

$$\mathrm{V}[S_n] \overset{\text{独立}}{=} \sum_{i=1}^{n} \mathrm{V}[X_i] \overset{\text{同分布}}{=} n \times \sigma^2$$

S_n の確率 (密度) 関数はこんな感じ?



L09-Q5

Quiz(独立同分布にしたがう確率変数の和)

ダーツで, 1 回投げるごとに点数が得られ, n 回投げた合計点で競うルールでプレイしている.
あるプレイヤーの i 回目の点数を確率変数 X_i とすると, X_i は独立同分布にしたがい, $\mathrm{E}[X_i] = 80, \mathrm{V}[X_i] = 200$ だという.
このプレイヤーが n 回投げた合計点を確率変数 $S_n = \sum_{i=1}^{n} X_i$ とする.

1. $\mathrm{E}[S_{10}]$, $\mathrm{V}[S_{10}]$ を求めよう.
2. $\mathrm{E}[\frac{1}{10}S_{10}]$, $\mathrm{V}[\frac{1}{10}S_{10}]$ を求めよう.
3. $\mathrm{E}[\frac{1}{100}S_{100}]$, $\mathrm{V}[\frac{1}{100}S_{100}]$ を求めよう.

ここまで来たよ

8 確率変数の独立, 独立同分布

9 二項分布, 独立同分布の和
- 二項分布
- 独立同分布にしたがう確率変数の和
- ベルヌーイ分布と二項分布・独立同分布の性質の応用
- チェビシェフの不等式

ベルヌーイ分布

定義 (ベルヌーイ分布 岩薩林 確率・統計 名前は出てこない)

$n = 1$ の二項分布 $\mathrm{B}(1, p)$ をパラメタ p のベルヌーイ分布という.

$$P(X = x) = p(x) = p^x \cdot (1-p)^{1-x} = \begin{cases} 1-p & (x = 0) \\ p & (x = 1) \\ 0 & (\text{他}) \end{cases}$$

意味: ベルヌーイ試行=(不公平な) コイン投げ. 表がでる確率 p. 表 $x = 1$.

命題 (ベルヌーイ分布のモーメントと母平均値と母分散)

$$\mathrm{E}[X^k] = \boxed{p} \ (k = 1, 2, \ldots), \quad \mathrm{V}[X] = \boxed{p(1-p)}.$$

命題 (ベルヌーイ分布と二項分布のもうひとつの関係)

岩薩林 確率・統計 例題 4.6 の最初の 1 文

$X_1, X_2, \ldots, X_n$ が独立同分布 $\mathrm{B}(1, p)$ にしたがうとき,
$S_n = X_1 + \cdots + X_n$ は $S_n \sim \mathrm{B}(n, p)$.

なぜなら

X_i は1枚のコインのうち表が出た枚数. S_n は n 枚のコインのうち表が出た枚数.

二項分布の母平均値・母分散の導出

$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathrm{B}(1, p)$ とすると, $\mathrm{E}[X_i] = p, \mathrm{V}[X_i] = p(1-p)$.

このとき, 独立同分布の和 確率統計 I(2022)L09 $S_n = \sum_{i=1}^{n} X_i \sim \mathrm{B}(n, p)$ に対して,

$$\mathrm{E}[S_n] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n] \overset{\text{同分布}}{=} n \times p$$

$$\mathrm{V}[S_n] \overset{\text{独立}}{=} \mathrm{V}[X_1] + \cdots + \mathrm{V}[X_n] \overset{\text{同分布}}{=} n \times p(1-p).$$

L09-Q6
(Q3 と同じ問を別解法で)

ここまで来たよ

8 確率変数の独立, 独立同分布

9 二項分布, 独立同分布の和
- 二項分布
- 独立同分布にしたがう確率変数の和
- ベルヌーイ分布と二項分布・独立同分布の性質の応用
- チェビシェフの不等式

チェビシェフの不等式

チェビシェフの不等式 Chebyshev's inequality

岩薩林 確率・統計 §4.3(4.15)

X を離散型または連続型確率変数とする. $\mu = \mathrm{E}[X]$: 母平均値,
$\sigma^2 = \mathrm{V}[X]$: 母分散
$a > 0$: 任意の正の実数.
このとき次が成立する.

$$P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}$$

どんな X にも使えて便利な不等式. これが母平均値・母標準偏差のひとつの意味づけ.

$a > 1$ と思う. 母平均値 μ からの距離が母標準偏差 $\times a$ 以上離れた値が出る確率は, $1/a^2$ 以下. 母平均値から離れるほど, 確率は小さい. 母標準偏差 σ は離れ方の基準 (分布の幅).

チェビシェフの不等式の証明 (離散型)

$\mathrm{E}[X] = \mu, \mathrm{V}[X] = \sigma^2$ とする.

$$g(x) = \begin{cases} 1 & (|x-\mu| \geq a\sigma) \\ 0 & (|x-\mu| < a\sigma) \end{cases}$$

とおく. 意味のわからない $\mathrm{E}[g(X)\cdot(X-\mu)^2]$ を定義に戻って書くと,

$$\begin{aligned}
(a\sigma)^2 \times P(|X-\mu| \geq a\sigma) &= (a\sigma)^2 \sum_{\mu - a\sigma \leq x \leq \mu + a\sigma} p(x) \\
&= \sum_x g(x)\cdot(a\sigma)^2\, p(x) \\
&\leq \sum_x g(x)\cdot(x-\mu)^2\, p(x) \\
&\leq \sum_x (x-\mu)^2\, p(x) = \mathrm{V}[X] = \sigma^2
\end{aligned}$$

連続型での証明 岩薩林 確率・統計 §4.3

中心極限定理・独立同分布の和の正規近似

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 L10(2023-06-12 Mon)

最終更新: Time-stamp: "2024-06-17 Mon 12:14 JST hig"

今日の目標

- チェビシェフの不等式 (母平均値・母分散の正確な意味) が説明できる 岩薩林 確率・統計 §4.3
- 大数の法則・中心極限定理が説明できる 岩薩林 確率・統計 §4.4
- 独立同分布の和の確率を正規近似できる

L09-Q1

L09-Q2

Quiz 解答: 二項分布

1. $X \sim \mathrm{B}(100, \frac{2}{3})$
2. $P(X = 40)$ を求めればいいから,
   ${}_{100}\mathrm{C}_{40} p^{40}(1-p)^{100-40} = \frac{100!}{40!60!}(\frac{2}{3})^{40}(1-\frac{2}{3})^{60}$.
3. $\mathrm{E}[X] = n \times p = \frac{200}{3}$. $\mathrm{V}[X] = n \times p(1-p) = \frac{200}{9}$.

L09-Q3

Quiz 解答: 二項分布または独立同分布

1. $X \sim \mathrm{B}(10, 0.3)$ とすると, $Y = 2X$.
2. $\mathrm{E}[Y] = \mathrm{E}[2X] = 2 \times 10 \cdot 0.7 = 14$ 個.
3. $\mathrm{V}[Y] = 2^2 \mathrm{V}[X] = 2^2 \times 10 \cdot 0.7 \cdot 0.3 = 8.4$ 個 ^{2}.
4. $P(Y = 16) = P(X = 8) = \frac{10}{2!8!} 0.7^8 0.3^2$

別解. 10 パックそれぞれに入っている卵の個数を確率変数 Y_i とすると,
$Y = \sum_{i=1}^{10} Y_i$ で Y_i は独立同分布にしたがうので,
$\mathrm{E}[Y] = 10\mathrm{E}[Y_i], \mathrm{V}[Y] = 10\mathrm{V}[Y_i]$.
$Y_i = 2X_i$ とおくと, X_i はパラメタ $p = 0.7$ のベルヌーイ分布にしたがう. すなわち, $X_i \sim \mathrm{B}(1, 0.7)$.
$\mathrm{E}[X_i] = 0.7$, $\mathrm{V}[Y_i] = 0.7 \cdot (1 - 0.3)$ なので, $\mathrm{E}[Y_i] = 2 \cdot 0.7$,
$\mathrm{V}[Y_i] = 2^2 \cdot 0.7 \cdot (1 - 0.3)$ である.
よって, $\mathrm{E}[Y] = 10\mathrm{E}[Y_i] = 14, \mathrm{V}[Y] = 10\mathrm{V}[Y_i] = 8.4$.

L09-Q5

Quiz 解答: 独立同分布にしたがう確率変数の和

1. $\mathrm{E}[S_{10}] = 800$, $\mathrm{V}[S_{10}] = 2000$.
2. $\mathrm{E}[\frac{1}{10}S_{10}] = 80$, $\mathrm{V}[\frac{1}{10}S_{10}] = 20$.
3. $\mathrm{E}[\frac{1}{100}S_{100}] = 80$, $\mathrm{V}[\frac{1}{100}S_{100}] = 2$.

L09-Q6 Q3 の別解，teamL09 裏 参照

ここまで来たよ

9 二項分布, 独立同分布の和

10 中心極限定理・独立同分布の和の正規近似
- チェビシェフの不等式
- 大数の法則
- 中心極限定理
- 正規近似

チェビシェフの不等式 Chebyshev's inequality 岩薩林 確率・統計 §4.3

定理 (チェビシェフの不等式 岩薩林 確率・統計 (4.15))

X を離散型または連続型確率変数とする.
母平均値を $\mu = \mathrm{E}[X]$, 母分散 $\sigma^2 = \mathrm{V}[X]$, $a > 0$: 任意の正の実数とする.
このとき次が成立する.

$$P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}$$

中心とか幅とか言ってた母平均値・母標準偏差の厳密な意味づけ.

$a > 1$ と思う. 母平均値 μ からの距離が母標準偏差 $\times a$ 以上離れた値が出る確率は, $1/a^2$ 以下. 母平均値から離れるほど, 確率は小さい. 母標準偏差 σ は離れ方の基準 (分布の幅).

チェビシェフの不等式の証明 (離散型)

$$\text{「離れた } x\text{」特徴関数} g(x) = \begin{cases} 1 & (|x-\mu| \geq a\sigma) \\ 0 & (|x-\mu| < a\sigma) \end{cases}$$

とおく．意味を考えず $\mathrm{E}[g(X)\cdot(X-\mu)^2]$ を定義に戻って書くと，

$$\begin{aligned}
(a\sigma)^2 \times P(|X-\mu| \geq a\sigma) &= (a\sigma)^2 \sum_{x<\mu-a\sigma \text{ or } \mu+a\sigma<x} p(x) \\
&= \sum_{x=-\infty}^{+\infty} g(x)\cdot(a\sigma)^2\, p(x) \\
&\leq \sum_{x=-\infty}^{+\infty} g(x)\cdot(x-\mu)^2\, p(x) \\
&\leq \sum_{x=-\infty}^{+\infty} 1\cdot(x-\mu)^2\, p(x) = \mathrm{V}[X] = \sigma^2
\end{aligned}$$

連続型での証明

ここまで来たよ

9 二項分布, 独立同分布の和

10 中心極限定理・独立同分布の和の正規近似
- チェビシェフの不等式
- 大数の法則
- 中心極限定理
- 正規近似

独立同分布にしたがう確率変数の和

岩薩林 確率・統計 例題 4.6

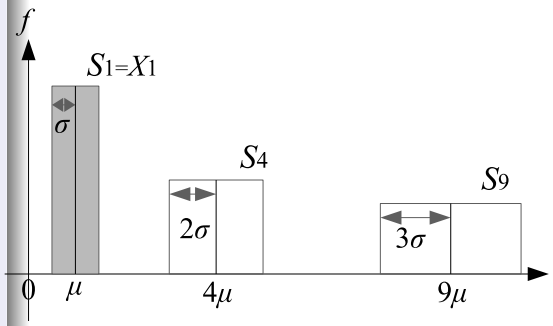
i.i.d にしたがう確率変数の和

$X_1, \ldots, X_n$: i.i.d. 母平均値 $\mathrm{E}[X_i] = \mu$, 母分散 $\mathrm{V}[X_i] = \sigma^2$.
和の確率変数 $S_n = X_1 + \cdots + X_n$.

$$\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] \overset{\text{同分布}}{=} n \times \mu.$$

$$\mathrm{V}[S_n] \overset{\text{独立}}{=} \sum_{i=1}^{n} \mathrm{V}[X_i] \overset{\text{同分布}}{=} n \times \sigma^2$$

S_n の確率密度関数はこんな感じ?

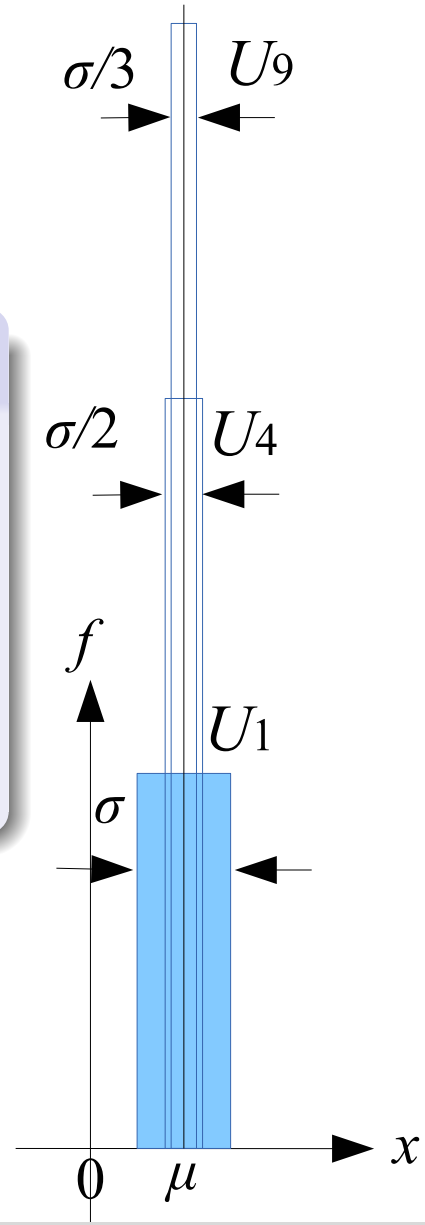


i.i.d にしたがう確率変数の和の $1/n$

$X_1, \ldots, X_n$: i.i.d.
新しい確率変数: $U_n = \frac{1}{n} S_n = \frac{1}{n}(X_1 + \cdots + X_n)$

$$\mathrm{E}[U_n] = \mathrm{E}\left[\frac{1}{n} S_n\right] = \frac{1}{n} \times n \times \mu.$$
$$\mathrm{V}[U_n] = \mathrm{V}\left[\frac{1}{n} S_n\right] = \left(\frac{1}{n}\right)^2 \times n \times \sigma^2.$$

U_n の確率密度関数はこんな感じ?



大数の (弱) 法則 岩薩林 確率・統計 §4.4

$X_1, \ldots, X_n$ が独立同分布にしたがい, $\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$ とする.

確率変数 $U_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ を考える (n 回の賞金の相加平均) と $\mathrm{E}[U_n] = \mu$, $\mathrm{E}[U_n] = \sigma^2/n$.

定理 (大数の (弱) 法則アバウト版 岩薩林 確率・統計 定理 4.1(p.84))

n が十分大きいとき確率変数 U_n は 'μ に近い値ばかり.' (U_n が μ から外れる確率はゼロに近づく)

弱法則の正確な表現

定理 (大数の (弱) 法則 岩薩林 確率・統計 定理 4.1(p.84))

$$\forall \epsilon > 0 \quad \lim_{n \to +\infty} P(|U_n - \mu| \geq \epsilon) = 0.$$

つまり n 大で U_n は $\mathrm{E}[U_n] = \mu$ に「必ず近い」(確率収束).

大数の弱法則の証明

U_n に対するチェビシェフの不等式を書くと,

$$P(|U_n - \mu| \geq a \times \frac{\sigma}{\sqrt{n}}) \leq \frac{1}{a^2}$$

不等式の右辺を ϵ にしたい下心から $a = \frac{\epsilon}{\frac{\sigma}{\sqrt{n}}}$ とすると, $n \to +\infty$ で

$$P(|U_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n} \to 0$$

これが母平均値・母期待値の直観的意味. 要するに,

何回も宝くじを買って賞金の相加平均をとると, 必ず E[賞金] に近い

確率にまつわるいろいろな「収束」

大数の強法則

U_n の「近づく」(!) 先が μ でない確率はゼロである

$$P(|\lim_{n\to+\infty} U_n - \mu| > 0) = 0$$

確率にまつわる「収束」にはいろいろある…

- 概収束
- 確率収束
- 平均収束
- 法則収束
- ⋮

ここまで来たよ

9 二項分布, 独立同分布の和

10 中心極限定理・独立同分布の和の正規近似
- チェビシェフの不等式
- 大数の法則
- 中心極限定理
- 正規近似

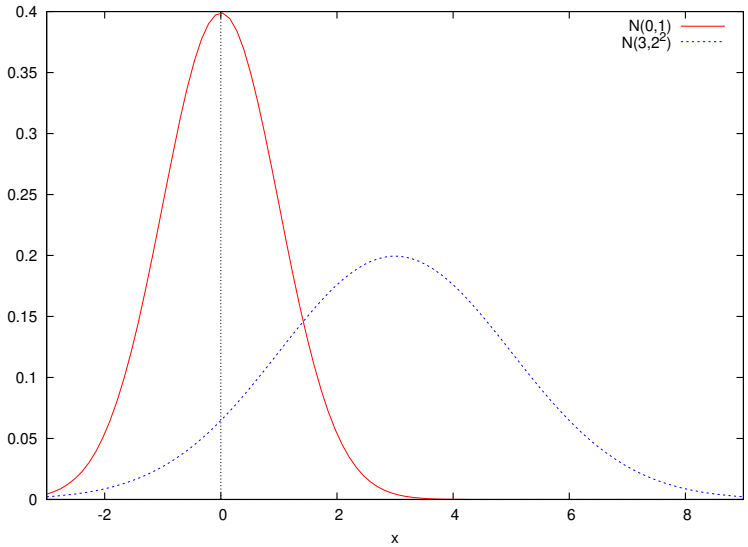
復習: 正規分布 岩薩林 確率・統計 §4.5

定理 ((一般の) 正規分布 $\mathrm{N}(\mu, \sigma^2)$ 岩薩林 確率・統計 (4.23))

母平均値 μ, 母分散 σ^2 の正規分布 $\mathrm{N}(\mu, \sigma^2)$ の確率密度関数は,

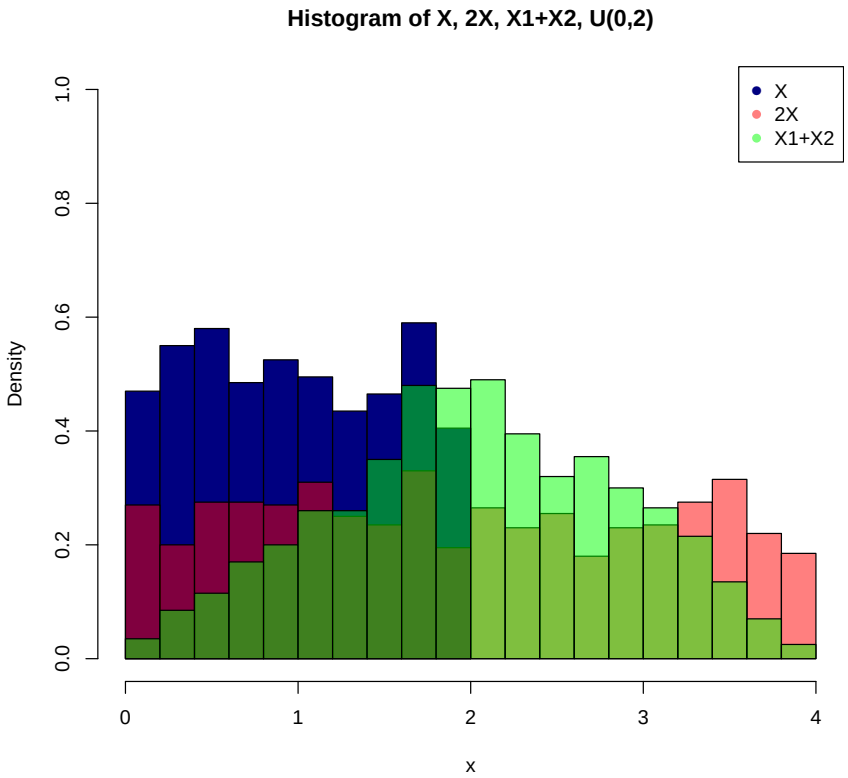
$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

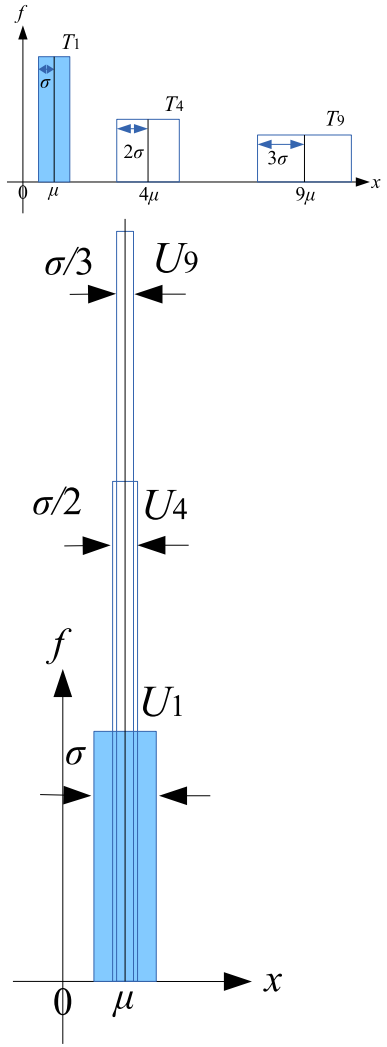
$f(x)$ =scipy.stats.norm(loc=μ,scale=σ).pdf(x)



対称軸は $x = \mu, f(\mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma}$,
$\sqrt{2}\sigma$ 離れると $1/e$.
$Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0, 1^2)$ 標準正規分布

一様分布に従う確率変数の定数倍と和





和の確率密度関数は，実際はどんな形なんだろう？

$X_1, X_2 \sim \mathrm{U}(0,2)$, i.i.d.

$2 \times X_1 \sim$

$X_1 + X_2 \sim$

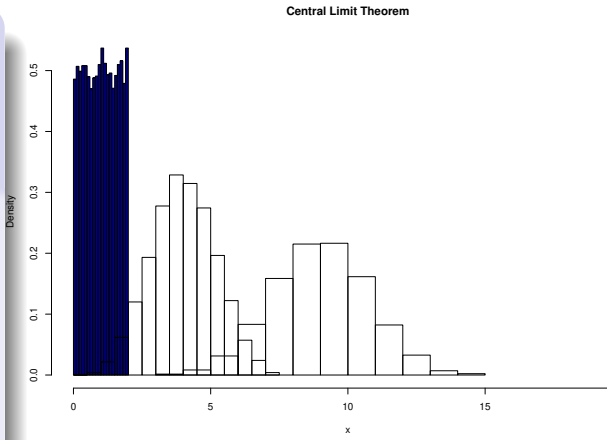
中心極限定理 岩薩林 確率・統計 §4.4

定理 (中心極限定理 (いいかげんバージョン) 岩薩林 確率・統計 定理 4.2(p.87))

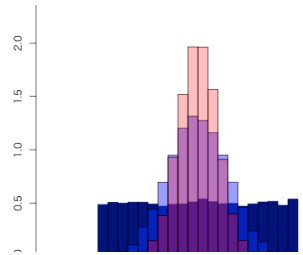
$X_1, \ldots, X_n$: 独立同分布にしたがう．$\mathrm{E}[X_i] = \mu$, $\mathrm{V}[X_i] = \sigma^2$. $n \to +\infty$ で,

- $S_n = X_1 + \cdots + X_n$, の確率分布は, 正規分布 $\mathrm{N}(n\mu, n\sigma^2)$ に似る
- $U_n = \frac{1}{n}(X_1 + \cdots + X_n)$ の確率分布は, 正規分布 $\mathrm{N}(\mu, \sigma^2/n)$ に似る
- 標準化した $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ の確率分布は, 標準正規分布 $\mathrm{N}(0, 1^2)$ に似る

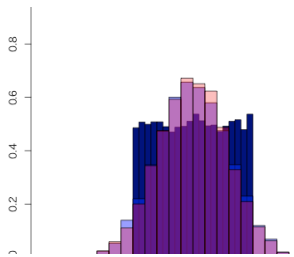
S



U



Z



ここまで来たよ

9 二項分布, 独立同分布の和

10 中心極限定理・独立同分布の和の正規近似
- チェビシェフの不等式
- 大数の法則
- 中心極限定理
- 正規近似

独立同分布にしたがう確率変数の和の正規近似 I

L10-Q1

Quiz(独立同分布と中心極限定理)

確率変数 X_i $(i = 1, \ldots, 400)$ は独立同分布にしたがい, $\mathrm{E}[X_i] = \frac{1}{10}$, $\mathrm{V}[X_i] = \frac{9}{100}$ である.
$S = X_1 + \cdots + X_{400}$ とする.
S の分布を正規分布で近似し, $P(S > 31)$ の確率を, 標準正規分布の累積分布関数 $\Phi(z)$ を用いて表し, さらに scipy.stats.norm.cdf() (または正規分布表) を用いて小数値として近似的に求めよう.

独立同分布にしたがう確率変数の和の正規近似 I

L10-Q2

Quiz(独立同分布と中心極限定理)

確率変数 X_i $(i = 1, \ldots, 400)$ は独立同分布にしたがい, $X_i \sim \mathrm{U}(3, 5)$ である. $S = X_1 + \cdots + X_{400}$ とする.
S の分布を正規分布で近似し, $P(S \leq 1500)$ の確率を, 標準正規分布の累積分布関数 $\Phi(z)$ を用いて表し, さらに scipy.stats.norm.cdf() (または正規分布表) を用いて小数値として近似的に求めよう.

二項分布の正規近似高校 数学 B I

L10-Q3

Quiz(ベルヌーイ分布の独立同分布の和と中心極限定理)

表が $\frac{4}{5}$, 裏が $\frac{1}{5}$ の確率ででる超いびつなコインを, 100 回投げる. 表が 73 回より多く 79 回以下でる確率を求めたい.
標準正規分布の累積分布関数 $\Phi(z)$ を用いて表し, さらに scipy.stats.norm.cdf() (または正規分布表) を用いて小数値として近似的に求めよう.

岩薩林 確率・統計 第 5 章練習問題 1(3)

二項分布の正規近似高校 数学 B

二項分布 $\mathrm{B}(n,p)$ は, 正規分布 $\mathrm{N}(np, np(1-p))$ で近似できる.

母集団と標本・母平均値/母分散/母比率の点推定

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 L11(2023-06-19 Mon)

最終更新: Time-stamp: "2023-06-19 Mon 12:06 JST hig"

今日の目標

- 母集団, 標本抽出, 推定を説明できる 岩薙林 確率・統計 §5.1,5.2
- 母平均値, 母期待値, 母分散, 母標準偏差，母比率を点推定できる 岩薙林 確率・統計 §6.1,§7.1,§7.2,§7.3
- 標本平均値の分布を説明できる 岩薙林 確率・統計 §5.3

L10-Q1

Quiz 解答: 独立同分布と中心極限定理

$n = 400$ が大きいと考えると, 中心極限定理より, S は近似的に正規分布 $\mathrm{N}(n\mu, n\sigma^2)$ すなわち $\mathrm{N}(40, 6^2)$ $Z = \frac{S-40}{6}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう. よって, 求める確率は,

$P(S > 31) = P(Z > -\frac{9}{6}) == P(-\frac{9}{6} < Z < +\infty) = F_{\mathrm{Z}}(\infty) - F_{\mathrm{Z}}(-\frac{9}{6}) = 1 - F_{\mathrm{Z}}(-\frac{9}{6}) = 0.9332.$

L10-Q2

Quiz 解答: 独立同分布と中心極限定理

$\mu = \mathrm{E}[X_i] = \frac{3+5}{2}, \sigma^2 = \mathrm{V}[X_i] = \frac{(5-3)^2}{12}.$

$n = 400$ が大きいと考えると, 中心極限定理より, S は近似的に正規分布 $\mathrm{N}(n\mu, n\sigma^2)$ すなわち $\mathrm{N}(1600, \frac{400}{3})$ にしたがう. よって, $Z = \frac{S-1600}{20/\sqrt{3}}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう. よって, 求める確率は,

$P(S \leq -5\sqrt{3}) = F_{\mathrm{Z}}(-5\sqrt{3})$

二項分布の正規近似高校 数学 B I

L10-Q3

Quiz(ベルヌーイ分布の独立同分布の和と中心極限定理)

表が $\frac{4}{5}$, 裏が $\frac{1}{5}$ の確率ででる超いびつなコインを, 100 回投げる. 表が 73 回より多く 79 回以下でる確率を求めたい.
標準正規分布の累積分布関数 $F_{\mathrm{Z}}(z)$ を用いて表し, さらに scipy.stats.norm.cdf() (または正規分布表) を用いて小数値として近似的に求めよう.

二項分布にしたがう $Y \sim \mathrm{B}(n,p)$ は，独立同分布 (ベルヌーイ分布) にしたがう $X_i \sim \mathrm{B}(1,p)$ $(i=1,\ldots,n)$ の和とみなせるので，やはり正規近似できる.

岩薩林 確率・統計 第 5 章練習問題 1(3)

命題 (二項分布の正規近似高校 数学 B)

二項分布 $\mathrm{B}(n,p)$ は, $\mathrm{N}(np, p(1-p))$ で近似できる.

ここまで来たよ

10 中心極限定理・独立同分布の和の正規近似

11 母集団と標本・母平均値/母分散/母比率の点推定
- 母集団と標本
- 母平均値・母分散の (点) 推定
- 母比率の (点) 推定

母集団と標本 (1) 無限母集団 特に 離散/連続型確率変数

岩薩林 確率・統計 §5.1,5.2

謎な仕組みのくじ (ひいたら戻す) の賞金額 (確率変数とも言う) の母平均値を, 結果から求めたい!

```
1  rvz=stats.norm(loc=0,scale=1) #N(0,1^2) 実は賞金は正規分布
2  rvz.rvs(size=5) # rvs=random variable sample
```

「確率変数 $Z \sim \mathrm{N}(0, 1^2)$ にしたがう試行を 5 回して，5 個の数値を得る」

- 母集団サイズ=$+\infty$ の母集団から，
- 標本サイズ=5 の標本 (sample) を
- 標本の個数=1 個を標本抽出する

無限母集団
数値は毎回生成されて何度でも引けるけど，抽象的に，無限個の数値のはいった袋があり，そこから数値に応じた確率で取り出してると考える．全部を取り出すことはできない．
推定

- 値 (賞金) の母平均値 $\mu = \mathrm{E}[X]$ を求めたい.
- 袋 (くじの仕組み=正規分布) を知れば定義の式使うだけ.
- しかし, いま袋の中 ($f(x)$) を見ることはできない.
- $+\infty$ 回くじを買わず, 5 個で何とかすませたい.

母集団と標本 (2) 有限母集団

岩薙林 確率・統計 §§5.1,5.2

某アイドルグループ全員 ($\to$ 有限母集団) の身長 x_i の平均値
$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ を求めたい!
無作為にメンバーが 1 名出てきて質問できる握手券を 5 枚買って質問する
無作為にだれかの身長が返される試行

- 母集団サイズ= 46 の母集団から,
- 標本サイズ= 5 の標本 (sample) を
- 標本の個数= 1 個を標本抽出する

推定

- 身長の母平均値 $\mu = \mathrm{E}[X]$ を求めたい.
- 全員の身長を知れば定義の式使うだけ.
- しかし, いま, 全員の身長 (母集団) を知ることはできない.
- 46 枚の握手券を買わず (実際は重複するかも), 5 枚で何とかすませたい.

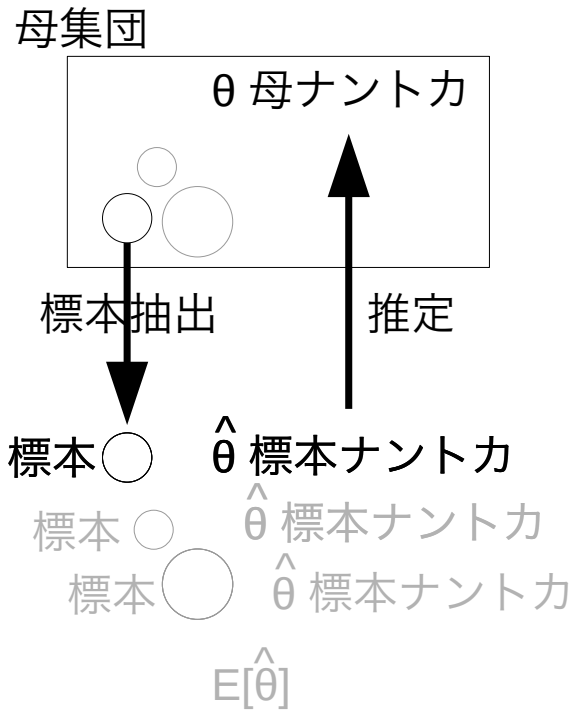
母集団・標本抽出・推定

岩薩林 確率・統計 例 11(p.115)

- 母集団 population =考えたい集団. どんな分布, 母平均値, 母分散, などわかっていないことがあるが, 全体を調べるわけにはいかない集団.
- 標本=sample (名詞) =母集団から '無作為に' とってきた一部分
- 標本抽出する sample(動詞)=母集団から '無作為に' とってくる ⇝ sampling (動名詞)
- 推定 する estimate(動詞) =標本を調べて母集団について正しそうな事実を見つける ⇝ estimation (名詞)
- 確率変数 X, $\overline{X}$ 分布をもつ変数
- 実現値, 観測値 x, $\bar{x}$ 標本を 1 つとって確定した値
- 推定量 $\hat{\theta}(x)$ 母集団の量 θ を推定する量

推定には誤差ある. 標本の選び方ごとに答は違うし.

岩薩林 確率・統計 図 p.109,115,137,167



科目参加者全体から抽出した標本: 身長, 滋賀県内高校

3 変量データ

- 身長 X = 身長 (参加者)(cm) 量的データ 連続
- $Y = \begin{cases} 1 & (\text{滋賀県高校 Yes}) \\ 0 & (\text{No}) \end{cases}$. 質的データ 名義尺度
- W A,B,C,D スクショする？ 質的データ (順序尺度)

母集団=回答者全体

1. 母集団サイズ 99 (クラス全体なら 131 だった)

今回の標本

1. 1 チームに割り当てる標本の個数 1 個
2. 標本サイズ 10

ここまで来たよ

10 中心極限定理・独立同分布の和の正規近似

11 母集団と標本・母平均値/母分散/母比率の点推定
- 母集団と標本
- 母平均値・母分散の (点) 推定
- 母比率の (点) 推定

母平均値の (点) 推定

組 $(X_1, X_2, \ldots, X_n)$ はサイズ n の標本. 各 X_i は母平均値 $\mu = \mathrm{E}[X_i]$, 母分散 $\sigma^2 = \mathrm{V}[X_i]$ の独立同分布にしたがう確率変数.

定義 (標本平均値 岩薩林 確率・統計 (5.4)p.114)

$$\text{標本平均値 } \overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) = \text{先週の} U_n$$

母平均値 $\mu = \mathrm{E}[X]$ の 'よい' 推定量になっている.

Pandas/Python では .mean(), Excel では関数 average()

定義 (標本期待値)

$$g(X)\text{の標本期待値 } \overline{g(X)} = \frac{1}{n}(g(X_1) + \cdots + g(X_n))$$

$\mathrm{E}[g(X)]$ の 'よい' 推定量になっている.

母期待値 $\mathrm{E}[g(X)]$ はひとつに定まっているが, 標本平均値 $\overline{X}$, $\overline{g(X)}$ は確率変数で, 試行=標本抽出のたびにかわる ($\overline{X}$ は確率分布をもつ)

よい (点) 推定量がもつ性質

- 不偏性 (unbiased ナントカ)
  推定量の母平均値は, 推定したい母ナントカに等しい 岩薩林 確率・統計 p.141
- 一致性 (consistency)
  推定量と母ナントカに一定の差がある確率は, 標本サイズを大きくすると zero になる 岩薩林 確率・統計 p.143
- 最尤性 (maximum likelihood)　確率統計 II

標本平均値 $\overline{X}$ の不偏性 岩薩林 確率・統計 p.113

母平均値 [母ナントカの推定量] =母ナントカ

$$\mathrm{E}[\overline{X}] = \frac{1}{n}(\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]) = \mu$$

標本平均値 $\overline{X}$ の一致性 大数の法則から 岩薩林 確率・統計 p.143

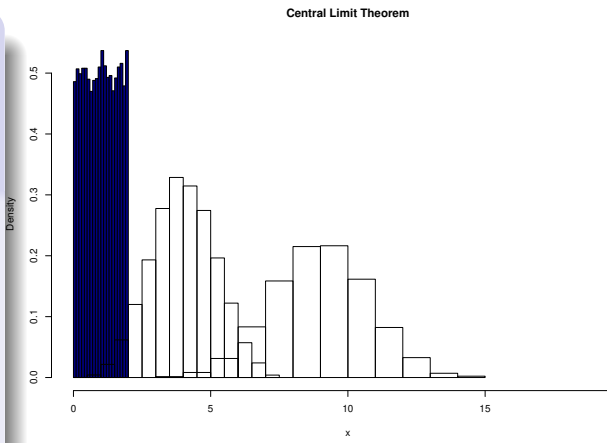
中心極限定理

定理 (中心極限定理 (いいかげんバージョン) 岩薩林 確率・統計 定理 4.2(p.87))

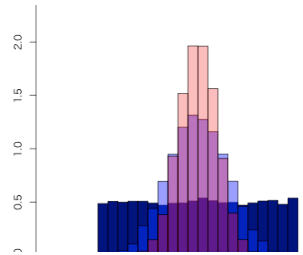
$X_1, \ldots, X_n$ が母平均値 μ, 母分散 σ^2 の独立同分布に従うとき, $n \to +\infty$ で

- $S_n = X_1 + \cdots + X_n$, の確率分布は, 正規分布 $\mathrm{N}(n\mu, n\sigma^2)$ に似る
- $U_n = \frac{1}{n}(X_1 + \cdots + X_n)$ の確率分布は, 正規分布 $\mathrm{N}(\mu, \sigma^2/n)$ に似る
- 標準化した $Z_n = \frac{S_n - n\mu}{\sqrt{n}\sigma}$ の確率分布は, 標準正規分布 $\mathrm{N}(0, 1^2)$ に似る

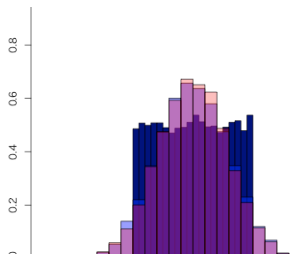
S



U



Z



L11-Q1

Quiz(確率変数としての標本平均値の分布)

正規分布 $\mathrm{N}(10, 6^2)$ にしたがう母集団 (正規母集団) から, サイズ $n = 4$ の標本 $X_1, \ldots, X_n$ を抽出し, 標本平均値 $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ を計算する.

1. $\overline{X}$ の母平均値を求めよう.
2. $\overline{X}$ の母分散を求めよう.
3. $\overline{X}$ を標準化する変換 $Z = \frac{\overline{X} - ?}{?}$ を書こう.
4. Z は標準正規分布にしたがう. 確率 $P(7 < \overline{X} \leq 13)$ を, $P(Z$ の不等式$)$ に直して, 標準正規分布の累積分布関数 $F(z)$ で表そう.
5. 確率 $P(-c < Z \leq +c) = 1 - \alpha = 1 - \frac{1}{5}$ となるような c を, 標準正規分布の累積分布関数の逆関数 $F^{-1}(z)$ で表そう.

L11-Q2

Quiz(母平均値, 母分散, 母比率の点推定)

フライドチキン屋さんのフライドチキンの大量の在庫 (=母集団) から, 無作為に 6 本のチキンを取り出したところ, 重さは次のようだった.
117g, 109g, 109g, 119g, 100g, 112g.

1. 重さの母平均値を点推定しよう.
2. 重さの二乗の母期待値を点推定しよう.
3. 重さの母分散を点推定しよう.
4. 110g 以上のものの母比率を点推定しよう.

桁落ちに注意

数値計算法

記述上の注意

- 母平均値 $= \mu = \mathrm{E}[X] \neq$ 標本平均値 $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$.
- 母分散 $= \sigma^2 = \mathrm{V}[X] \neq$ 不偏標本分散 $S^2 = \frac{1}{n-1}(\cdots)$.
- 母比率 $= p \neq$ 標本比率 $\hat{p} = \frac{k}{n}$.
- ここしばらくの問題で,「母ナントカを…と × 求めた○推定する」

上でタイプの間違いは厳しく弾圧します. × き

Moodle の練習問題スケジュール 今回から, 練習問題の締切を月 09:15 に変更. それ以降が再練習問題.

統計検定自分で申し込んで好きなときに CBT 受験できる

`https://www.toukei-kentei.jp`

- 統計検定 3 級=データ分析まで+この科目の一部
- 統計検定 2 級=この科目まで $+\alpha$. データサイエンスプログラム TM に勉強会記録あり.

母分散の (点) 推定

定義 (不偏標本分散 岩薩林 確率・統計 (5.11) の V(p.122))

$$\text{不偏標本分散 } S^2 = \frac{1}{n-1}[(X_1-\overline{X})^2+\cdots+(X_n-\overline{X})^2]$$
$$= \frac{n}{n-1}\left[\frac{1}{n}\sum_i X_i^2-(\overline{X})^2\right]$$

不偏標本分散は母分散 σ^2 の 'よい' 推定値になっている.

不偏標本分散は 岩薩林 確率・統計 p.122 の不偏分散のこと. 岩薩林 確率・統計 p.113 の標本分散は $n-1$ でなく n で割ったもの.

ここで, $\overline{X}$ は母平均値でなく, 上の標本平均値.

Pandas/Python では .var(ddof=1), Excel では関数 var.s()

$n-1$ の理由 こうするとちょうど不偏: $\mathrm{E}[S^2]=\sigma^2$.

直観的理由 $\overline{X}$ は X_i の重心だから, μ より近くにある. $(X_i-\overline{X})^2$ は $(X_i-\mu)^2$ より小さくなりがち ($\frac{n-1}{n}$ 倍) なので修正.

不偏標本分散の不偏性の確認

$n = 2$. $\mathrm{V}[X_i] = \sigma^2$. $\overline{X} = \frac{1}{2}(X_1 + X_2)$.

$$
\begin{aligned}
\mathrm{E}[S^2] =& \mathrm{E}[\tfrac{1}{2-1}((X_1 - \overline{X})^2 + (X_2 - \overline{X})^2)] \\
=& \mathrm{E}[(X_1 - \tfrac{1}{2}(X_1 + X_2))^2 + (X_2 - \tfrac{1}{2}(X_1 + X_2))^2] \\
=& 2 \cdot \frac{1}{4}\mathrm{E}[(X_1 - X_2)^2] \\
=& 2 \cdot \frac{1}{4}\mathrm{E}[X_1^2 - 2X_1X_2 + X_2^2] \\
=& 2 \cdot \frac{1}{4}((\sigma^2 + \mu^2) - 2\mu\mu + (\sigma^2 + \mu^2)) \\
=& 2 \cdot \frac{1}{4}(2\sigma^2) = \sigma^2.
\end{aligned}
$$

ここまで来たよ

10 中心極限定理・独立同分布の和の正規近似

11 母集団と標本・母平均値/母分散/母比率の点推定
- 母集団と標本
- 母平均値・母分散の (点) 推定
- 母比率の (点) 推定

比率=ratio 岩薩林 確率・統計 p.107

確率変数 $Y \sim \mathrm{B}(1,p)$ ベルヌーイ分布, を考える.
こういう Y は, いろんな母集団を,「X は…である」という条件の成立不成立で 2 つに類別して作れる. 名義データ カテゴリ変数

- $X \sim$ ある分布, $Y = \begin{cases} 1 & (X \text{ の条件成立}) \\ 0 & (X \text{ の不成立}) \end{cases}$. 例 $X > 10$ なら $Y = 1$.
- 母集団=日本国民, その国民血液型が A であるなら $Y = 1$.

定義 (母比率 岩薩林 確率・統計 p.107)

$\mathrm{B}(1,p)$ の p. または母集団で X の条件「…」から $\mathrm{B}(1,p)$ を作ったとき, '母集団の「…である」ものの母比率', ともいう.

有限母集団なら,

$$\text{母集団の「…である」母比率 } p = \frac{\text{「…である」メンバー } x \text{ の個数}}{\text{すべてのメンバーの個数}} = \mathrm{E}[Y]$$

やりたいこと: 母比率の推定

ベルヌーイ分布の p (母比率) を標本から推定したい!

- クラスの中で, 血液型 A 型の人の比率は? n 人に質問しただけで推定したい.
- 候補者 A の得票率は何% ? n 人に質問しただけで推定したい.
- 工場から出荷する製品のうち, 何% が不良品? n 個だけ抜き出して調査したい.
- このコインの表が出る確率は? n 回投げるだけで推定したい.

母比率の (点) 推定 岩薩林 確率・統計 p.115

定義 (標本比率 岩薩林 確率・統計 p.115)

標本のデータ n 個中 k 個が「…である」とき,

$$\text{標本比率 } \hat{p} = \frac{k}{n}$$

標本比率は「…」の母比率 p のよい推定値になっている.

$$\text{母平均値}\mathrm{E}[Y]\text{の推定値} = \text{標本平均値}\overline{Y}$$

$$= \frac{1}{n}[\underbrace{1 + \cdots + 1}_{k} + \underbrace{0 + \cdots + 0}_{n-k}] = \frac{k}{n} = \hat{p}.$$

岩薩林 確率・統計 問題 6(p.116)

母比率の区間推定

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 L12(2023-06-26 Mon)

最終更新: Time-stamp: "2024-06-28 Fri 17:08 JST hig"

今日の目標

- 母比率を区間推定できる 岩薙林 確率・統計 §7.3
- カイ二乗分布を説明できる 岩薙林 確率・統計 p.123

L10-Q3

Quiz 解答: ベルヌーイ分布の独立同分布の和と中心極限定理

100 回中表の出る回数は，$Y = X_1 + \cdots + X_{100}$, $X_i \sim \mathrm{B}(1, \frac{4}{5})$, 独立同分布, $\mathrm{E}[X_i] = \frac{4}{5}, \mathrm{V}[X_i] = \frac{4}{5}\frac{1}{5}$. よって, $\mathrm{E}[Y] = 80$, $\mathrm{V}[Y] = 4^2$ である (これは, $Y \sim \mathrm{B}(100, \frac{4}{5})$ からも求められる)

$n = 100$ が大きいと考えると, 中心極限定理より, Y は近似的に正規分布 $\mathrm{N}(80, 4^2)$ にしたがう.

標準化された $Z = \frac{Y-80}{4}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう.

よって, 求める確率は,

$P(73 < X \leq 79) = P(-\frac{7}{4} < Z \leq -\frac{1}{7}) = \Phi(-\frac{1}{4}) - \Phi(\frac{7}{4}) = 0.4599 - 0.0987 = 0.3612.$

L11-Q1

L11-Q2

Quiz 解答: 母平均値, 母分散, 母比率の点推定

在庫のフライドチキンの重さを X とすると,

1. 標本平均値 (の実現値) は $\overline{X} = \frac{1}{6}(117 + \cdots + 112) = 111\mathrm{g}$ なので, 母平均値は $111\mathrm{g}$ と推定できる.
2. 標本期待値 (の実現値) は $\overline{X^2} = \frac{1}{6}(117^2 \cdots + 112^2) = 37078/3\mathrm{g}^2$ なので, 母期待値は $37078/3\mathrm{g}^2$ と推定できる.
3. 不偏標本分散 (の実現値) は, $S^2 = \frac{1}{6-1}[(117-111)^2 + \cdots + (112-111)^2] = 46\mathrm{g}^2$ なので, 母分散は $46\mathrm{g}^2$ と推定できる.
4. 標本比率 (の実現値) は, $p = \frac{1}{6}[1+0+0+1+0+1] = 0.5$ なので, 母比率は 0.5 と推定できる.

ここまで来たよ

11 母集団と標本・母平均値/母分散/母比率の点推定

12 母比率の区間推定
- 母平均値の区間推定 (正規母集団, 母分散既知)
- 母比率の区間推定
- 不偏標本分散のちらばりとカイ二乗分布
- 母分散の区間推定

点推定 対 区間推定

点推定 岩薩林 確率・統計 §6.1
真の母平均値はわからないが, 標本平均値を使って,

「母平均値を A 円と推定する」

それどのくらい正確なの? 正確さは実は 母分散や標本サイズによる

区間推定 岩薩林 確率・統計 §6.1

「母平均値が, B 円以上 C 円以下である '確率' は $1-\alpha=0.95$」

推定の精度・正確さまで表現
ここで '確率' というのは不誠実. 正しい言葉遣いは, 信頼係数=信頼度で

「母平均値の信頼係数 $1-\alpha=0.95$ の信頼区間は B 円以上 C 円以下」

動く (確率変数である) のは母平均値 μ でなく, B, C のほう.

標本平均値のしたがう分布 (正規母集団, 母分散既知)

岩薩林 確率・統計 p.144

$\mathrm{N}(\mu, \sigma^2)$ にしたがう母集団 (正規母集団) の, サイズ n の標本を何回も取り出して, 毎回, 標本平均値 $\overline{X}$ を計算する.

その意味で, 標本平均値は確率変数

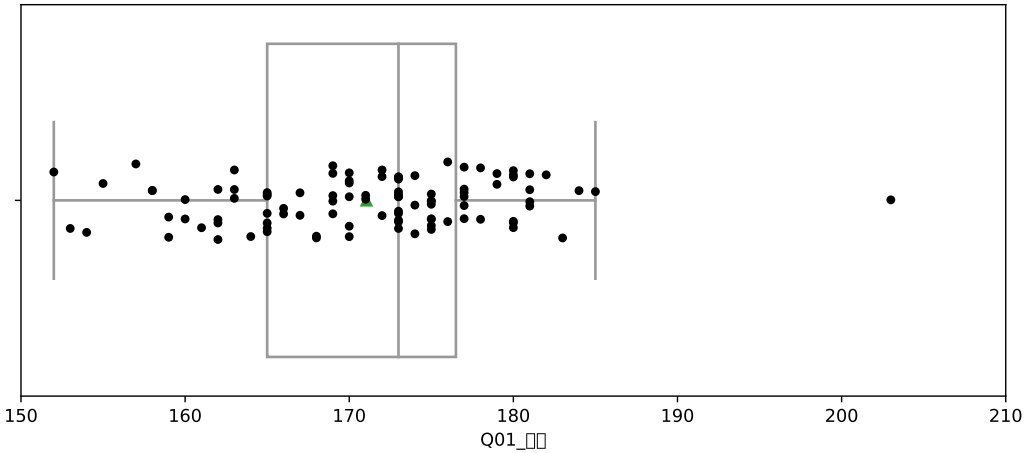
先週チーム課題で計算したのは, 1 標本で得られた標本平均値 (1 試行で得られた確率変数の値).

実は, $U_n = \overline{X} \sim \mathrm{N}(\mu, \sigma^2/n)$. $Z = \dfrac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} \sim \mathrm{N}(0, 1^2)$.

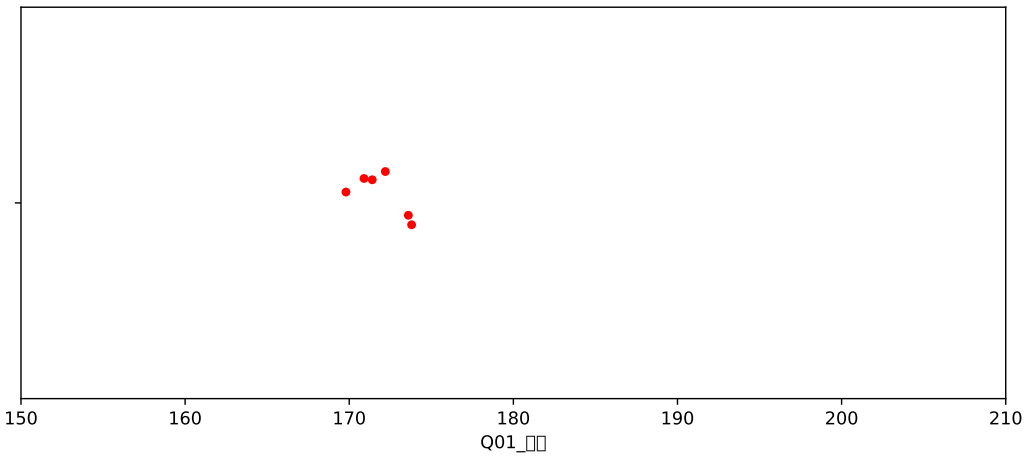
正規分布の再生性 確率統計 II(2023)L?? 岩薩林 確率・統計 p.99 から. 使わなくても, $n \to +\infty$ で正しいことは中心極限定理からわかる. 正規母集団でないときも, 標本サイズ n が大きい (30 くらい) なら, 近似的に成立することが多い.

身長の標本平均値の分布と母平均値

母集団



サイズ 10 の標本 6 個の標本平均値

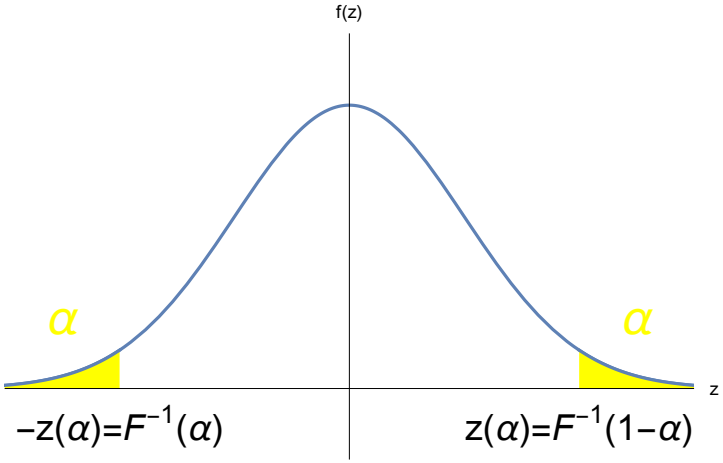


L12-Q1

Quiz(正規分布の上側確率)

$Z \sim \mathrm{N}(0, 1^2)$ とする. Z の累積分布関数を $\Phi(z)$ とする.

1. $P(u < Z)$ を u と Φ で表そう.
2. $P(u < Z) = \alpha$ となる u を α と Φ で表そう. この u のことをよく $z(\alpha)$ と書く.
3. Z の確率密度関数が偶関数であることから, $P(u < Z) = P(Z < -u)$ であることを説明しよう.
4. $P(Z < -u) + P(u < Z) = \alpha$ となる u を α と $z()$ で表そう.
5. $\alpha = 0.05$ のとき, 上のような u を求めよう.



定義 (標準正規分布の $z(\alpha)$)

$Z \sim \mathrm{N}(0, 1^2)$ のとき, 上側確率 $P(Z \geq z(\alpha)) = \alpha$ となる境い目を $z(\alpha)$ と定める. 岩薩林 確率・統計 付表 1 下

$z(\alpha) = F_{\mathrm{Z}}^{-1}(1-\alpha)$. 偶関数だから $z(1-\frac{\alpha}{2}) = -z(\frac{\alpha}{2})$.

標本平均値 $\overline{X}$ が母平均値 μ から大きく外れない確率は大きい (ここでは $1-\alpha = 1-0.05$) という式を書くと…

$$P\left(F_{\mathrm{Z}}^{-1}(\tfrac{\alpha}{2}) < \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}} < F_{\mathrm{Z}}^{-1}(1-\tfrac{\alpha}{2})\right) = 1-\alpha.$$

$$P\left(\mu + F_{\mathrm{Z}}^{-1}(\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n} < \overline{X} < \mu + F_{\mathrm{Z}}^{-1}(1-\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n}\right) = 1-\alpha.$$

μ について不等式を解くと,

$$P\left(\overline{X} - F_{\mathrm{Z}}^{-1}(1-\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n} < \mu < \overline{X} - F_{\mathrm{Z}}^{-1}(\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n}\right) = 1-\alpha.$$

$$P\left(\overline{X} - z(\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n} < \mu < \overline{X} - (-z(\tfrac{\alpha}{2})) \times \sqrt{\sigma^2/n}\right) = 1-\alpha.$$

業界の習慣で, しばしば $\alpha = 0.05, 0.01$

$$\alpha = 0.05 \rightsquigarrow P\left(\overline{X} - 1.96 \times \sqrt{\sigma^2/n} < \mu < \overline{X} + 1.96 \times \sqrt{\sigma^2/n}\right) = 1-0.05.$$

命題 (母平均値 (正規母集団, 母分散既知) の信頼区間

岩薩林 確率・統計 定理 6.1(p.18))

$\mathrm{N}(\mu, \sigma^2)$ にしたがう母集団の, σ^2 がわかっているとき, サイズ n の標本から区間推定すると, 母平均値 μ の 信頼係数 $1-\alpha$ の信頼区間 ($(1-\alpha)$ 信頼区間) は, $\overline{X}$ を標本平均値として,

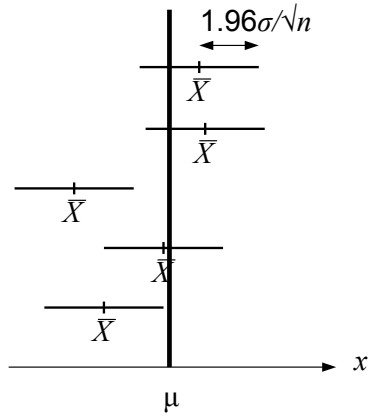
$$\overline{X} - z(\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n} < \mu < \overline{X} + z(\tfrac{\alpha}{2}) \times \sqrt{\sigma^2/n}.$$

何回も標本抽出して何個も信頼区間を求めたとき, 信頼区間が μ を含む確率は, 信頼係数 $1-\alpha$. 推定が外れる確率 α.

切りがいい α の $z(\alpha)$ は岩薩林 確率・統計 付表 1 下 (p.227)

$z(\frac{0.05}{2}) = 1.96, z(\frac{0.01}{2}) = 2.58.$

ゆるゆるな高校 数学 B では, $z(\frac{0.05}{2}) = 1.96$ の場合のみ.
$a < \mu < b$ でなく, 閉区間の記号 $[a, b]$ で.
真の母分散 σ^2 の代わりに, (不偏 $\frac{1}{n-1}$ でない) 標本分散 $S^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$ を使っていい.



L12-Q2

Quiz(母平均値の区間推定 (母分散既知))

確率変数 X は, 母分散が $\mathrm{V}[X] = \sigma^2 = 3^2$, 母平均値 $\mu = \mathrm{E}[X]$ が不明な正規分布 $\mathrm{N}(\mu, 3^2)$ にしたがう. X のサイズ 4 の標本を抽出したところ次のようになった (本当は整数値になる確率は 0 だけど, 計算を楽にするためのファンタジー).
$51, 52, 47, 50.$

1. 母平均値 μ を, 信頼係数 $1 - \alpha = 0.95$ で区間推定しよう.
2. 母平均値 μ を, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.

岩薩林 確率・統計 例題 6.4(p. 145)	岩薩林 確率・統計 問題 3(p.146)	岩薩林 確率・統計 第 6 章練習問題 1

推定が正確であるとは 信頼区間が 自分の言葉で であること.

Quiz(区間推定の性質)

標本からの母平均値の区間推定について, 正しいのはどれ?

1. 母分散が大きいほど, 信頼区間は大きくなる
2. 標本サイズが大きいほど, 信頼区間は大きくなる
3. 母平均値が大きいほど, 信頼区間は小さくなる
4. 信頼係数が大きいほど, 信頼区間は小さくなる

標本平均値が大きい $\Rightarrow$ 信頼区間は 平行移動する

母分散が大きい $\Rightarrow$ 信頼区間は 大きい

標本サイズ n が小さい 信頼区間は 小さい

ここまで来たよ

11 母集団と標本・母平均値/母分散/母比率の点推定

12 母比率の区間推定
- 母平均値の区間推定 (正規母集団, 母分散既知)
- 母比率の区間推定
- 不偏標本分散のちらばりとカイ二乗分布
- 母分散の区間推定

母比率の信頼区間

母比率は母平均値の一種なので, さっきの区間推定の式で, $\sigma^2 = p(1-p)$ とおく.
信頼係数 $1-\alpha$.

$$P\left(p - z(\tfrac{\alpha}{2}) \times \sqrt{\tfrac{p(1-p)}{n}} < \hat{p} < p + z(\tfrac{\alpha}{2}) \times \sqrt{\tfrac{p(1-p)}{n}}\right) = 1-\alpha.$$

平方根の中の $\sigma^2 = p(1-p)$ は $\hat{p}(1-\hat{p})$ と近似して p について解くと次を得る.

母比率の信頼区間 (母分散未知) 岩薩林 確率・統計 §7.3

X のサイズ n の標本で, 標本比率 $\hat{p} = k/n$ のとき, 母比率の 信頼係数 $1-\alpha$ の信頼区間 ($(1-\alpha)$ 信頼区間) は,

$$\hat{p} - z(\tfrac{\alpha}{2}) \times \sqrt{\tfrac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z(\tfrac{\alpha}{2}) \times \sqrt{\tfrac{\hat{p}(1-\hat{p})}{n}}.$$

$z(\frac{0.05}{2}) = 1.96, z(\frac{0.01}{2}) = 2.58.$

L12-Q3

Quiz(母比率の区間推定)

選挙で出口調査をしたところ, 50 人中 35 人が A 候補に投票したと答えた. 母集団を投票した人 2500 人とする. そのうち A 候補に投票した人の母比率 (得票率) を考える.

1. A 候補の得票率を, (点) 推定しよう
2. A 候補の得票率を, 信頼係数 $1-\alpha=0.95$ で区間推定しよう.
3. A 候補の得票率を, 信頼係数 $1-\alpha=0.99$ で区間推定しよう.

岩薩林 確率・統計 例題 7.6(p.170)	岩薩林 確率・統計 問題 7(p.171)	岩薩林 確率・統計 第 7 章練習問題 2(2)

注: 下限, 上限が 0,1 を越えるときは, 0,1 に直してしまっていい.

ここまで来たよ

11 母集団と標本・母平均値/母分散/母比率の点推定

12 母比率の区間推定
- 母平均値の区間推定 (正規母集団, 母分散既知)
- 母比率の区間推定
- 不偏標本分散のちらばりとカイ二乗分布
- 母分散の区間推定

ちらばり (不偏標本分散) のちらばりを考えたい

標本データのちらばりって? $\sqrt{\text{母分散}} \overset{\text{点推定}}{\leftarrow} \sqrt{\text{不偏標本分散}}$

- 母分散の点推定の精度って?

	の点推定	の区間推定
母平均値 μ	標本平均値 $\overline{X} = \frac{1}{n}[X_1 + \cdots]$	$\overline{X} - \boxed{\text{正}}\sqrt{\quad} < \mu < \overline{X} - \boxed{\text{負}}\sqrt{$
母分散 σ^2	不偏標本分散 $S^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots]$	$S^2 \times \boxed{\text{小}} < \sigma^2 < S^2 \times \boxed{\text{大}}$

母集団が正規分布にしたがうとき

- 標本平均値の分布 (正規分布) をうまく平行移動, 拡大縮小すると標準正規分布 $\mathrm{N}(0, 1^2)$
- 不偏標本分散の分布をうまく拡大縮小するとカイ二乗分布 χ^2_k

確率統計 II(2023)L??

カイ二乗分布 岩薙林 確率・統計 p.123

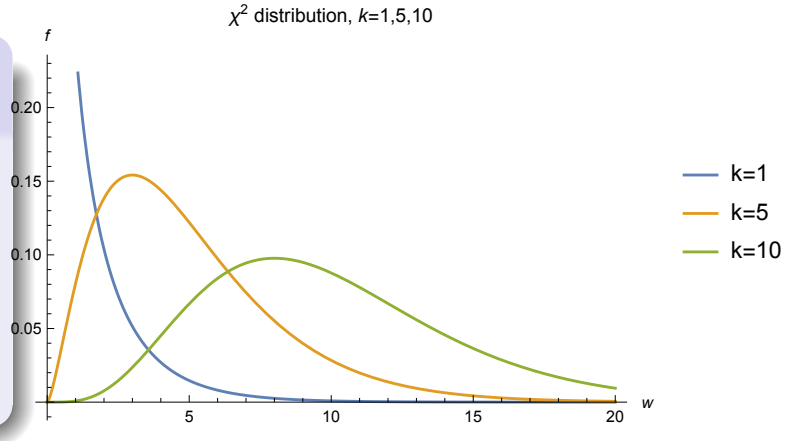
カイ二乗分布

$Z_1, \ldots, Z_k \sim \mathrm{N}(0, 1^2)$, iid のとき, 確率変数 $W = Z_1^2 + \cdots + Z_k^2$ は, 自由度 k のカイ二乗分布 χ_k^2 にしたがう.

言語	小	大	読み
英語	x	X	エクス
ギリシャ語	χ	X	カイ

χ_k^2 の確率密度関数

$$f_k(y) = \begin{cases} C_k \times y^{\frac{k}{2}-1} \mathrm{e}^{-\frac{1}{2}y} & (y \geq 0) \\ 0 & (他) \end{cases}$$



$W_k \sim \chi_k^2$ に対して,

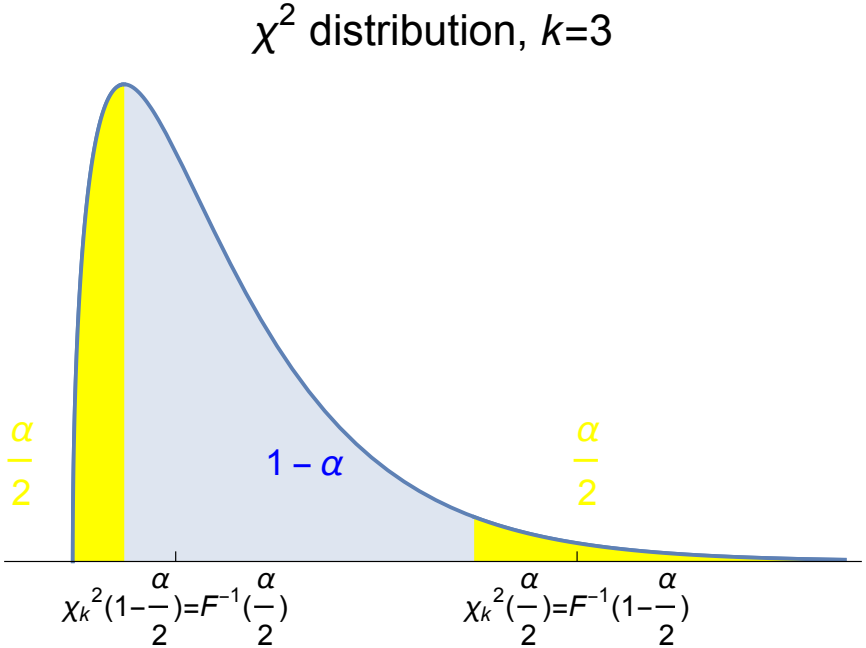
$\mathrm{E}[W_k] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{V}[W_k] = 2k, \mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

カイ二乗分布の確率密度関数と累積分布関数

```
1 rvx2=stats.chi2(df=\text{自由度}k)
```

上側確率 $\alpha = P(W > \chi_k^2(\alpha))$ となる境い目 $\chi_k^2(\alpha) = F^{-1}(1-\alpha)$.

岩薩林 確率・統計 付表 3 付表 3 は付表 1 とフォーマットが違う



$\chi_k^2(\alpha)$ の定義

岩薩林 確率・統計 例題 5.7

$\alpha = P(W > \chi_k^2(\alpha))$.

$\chi_k^2(\alpha) = F^{-1}(1-\alpha)$.

L12-Q4

Quiz(カイ二乗分布の確率と $\chi^2_k(\alpha)$)

標準正規分布にしたがう確率変数 $Z \sim \mathrm{N}(0, 1^2)$ と, 自由度 $k = 1$ のカイ二乗分布にしたがう確率変数 $W \sim \chi^2_1$ を考える. 各分布の累積分布関数の逆関数 Φ^{-1}, F^{-1} を使って答え, さらに Python, 数表を使って数値にしよう.

1. 確率 $P(Z > x_0) = 0.025$ となる $x_0 = z(0.025)$ を求めよう.
2. 確率 $P(Z > x_0) = 1 - 0.025$ となる $x_0 = z(1 - 0.025)$ を求めよう.
3. 確率 $P(W > w_0) = 0.05$ となる $w_0 = \chi^2_1(0.05)$ を求めよう.
4. 確率 $P(W > w_0) = 1 - 0.05$ となる $w_0 = \chi^2_1(1 - 0.05)$ を求めよう.

ここまで来たよ

11 母集団と標本・母平均値/母分散/母比率の点推定

12 母比率の区間推定
- 母平均値の区間推定 (正規母集団, 母分散既知)
- 母比率の区間推定
- 不偏標本分散のちらばりとカイ二乗分布
- 母分散の区間推定

不偏標本分散のしたがう分布

不偏標本分散のしたがう分布 岩薙林 確率・統計 定理 5.6

正規分布 $\mathrm{N}(\mu, \sigma^2)$ からサイズ n の標本 $X_1, X_2, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$, iid を取り出すとき, 不偏標本分散

$$S^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2]$$

から定めた

$$W = (n-1) \times \frac{S^2}{\sigma^2}$$

は, 自由度 $k = n - 1$ の カイ二乗分布 χ^2_{n-1} にしたがう.

比 $\frac{\text{不偏標本分散}}{\text{母分散}}$ は 1 に近いところに分布するが, 実は, 確率変数 $\frac{W}{n-1}$.
$(W \sim \chi^2_{n-1})$

証明じゃないけど説明

独立な $X_i \sim \mathrm{N}(\mu, \sigma^2)$ $(i = 1, \ldots, n)$ に対して,

$$n \times \frac{1}{n}\left[\left(\frac{X_1 - \mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \mu}{\sigma}\right)^2\right]$$

は自由度 n のカイ二乗分布 χ^2_n にしたがう.
不偏標本分散 S^2 に対して,

$$W = (n-1) \times \frac{S^2}{\sigma^2} = (n-1) \times \frac{1}{n-1}\left[\left(\frac{X_1 - \overline{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \overline{X}}{\sigma}\right)^2\right]$$

は自由度は $n-1$ のカイ二乗分布 χ^2_{n-1} にしたがう.
$-\mu$ でなく $-\overline{X}$ であるため自由度 $n-1$.

母分散・母平均値の区間推定

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　L13(2023-07-03 Mon)

最終更新: Time-stamp: "2024-07-15 Mon 11:52 JST hig"

今日の目標

- 母分散を区間推定できる　岩薙林 確率・統計 §7.2
- t 分布を説明できる　岩薙林 確率・統計 §5.5
- 母平均値を区間推定できる　岩薙林 確率・統計 §7.1

L12-Q1

Quiz 解答: 正規分布の上側確率

1. $1 - \Phi(u)$.
2. $z(\alpha) = \Phi^{-1}(1-\alpha)$.
3. Z の確率密度関数が偶関数であることからいえる.
4. 上より, $P(Z < -u) = P(u < Z) = \alpha/2$ となる u を考えて, $z(\frac{\alpha}{2}) = \Phi^{-1}(1-\frac{\alpha}{2})$.
5. $\Phi^{-1}(1-\frac{0.05}{2}) = \Phi^{-1}(0.975) = \texttt{.ppf}(0.975) = 1.96$.

L12-Q2

Quiz 解答: 母平均値の区間推定 (母分散既知)

1. 標本平均値は $\overline{X} = 50$. よって, 信頼係数 0.95 の信頼区間は

$$50 - \Phi^{-1}(1 - \frac{\alpha}{2}) \times \sqrt{\frac{3^2}{4}} < \mu < 50 - \Phi^{-1}(\frac{\alpha}{2}) \times \sqrt{\frac{3^2}{4}}.$$

$$50 - 1.96 \times \sqrt{\frac{3^2}{4}} < \mu < 50 + 1.96 \times \sqrt{\frac{3^2}{4}}.$$

   すなわち, $47.06 < \mu < 52.94$.

2. 同様に,

$$50 - 2.58 \times \sqrt{\frac{3^2}{4}} < \mu < 50 + 2.58 \times \sqrt{\frac{3^2}{4}}.$$

   すなわち, $46.13 < \mu < 53.87$.

L12-Q3

Quiz 解答: 母比率の区間推定

A 候補に投票したを $X = 1$, しなかったを $X = 0$ とする.

1. 標本比率は $\hat{p} = \frac{35}{50} = 0.7$. 母比率 p を 0.7 と推定する.

2. 母比率 p の信頼係数 $1-\alpha=0.95$ の信頼区間は,
$\Phi^{-1}(1-\frac{0.05}{2})=z(\frac{0.05}{2})=1.96$ より

$$\frac{7}{10}-1.96\times\sqrt{\frac{\frac{7}{10}\cdot(1-\frac{7}{10})}{50}}<p<\frac{7}{10}+1.96\times\sqrt{\frac{\frac{7}{10}\cdot(1-\frac{7}{10})}{50}}$$
$$0.7-0.13<p<0.7+0.13$$
$$0.57<p<0.83$$

$0.5<0.57$ なので，信頼係数 0.95 では当選ってことですね (放送用語「当選確実」で, 多くの選挙区で判定したとき, 後で社長があやまらなきゃいけない確率は 0.05).

3. 母比率 p の信頼係数 $1-\alpha=0.99$ の信頼区間は, $z(\frac{0.01}{2})=2.58$ より

$$\frac{7}{10}-2.58\times\sqrt{\frac{\frac{7}{10}\cdot(1-\frac{7}{10})}{50}}<p<\frac{7}{10}+2.58\times\sqrt{\frac{\frac{7}{10}\cdot(1-\frac{7}{10})}{50}}$$
$$0.7-0.17<p<0.7+0.17$$
$$0.53<p<0.87$$

カイ二乗分布 岩薙林 確率・統計 p.123

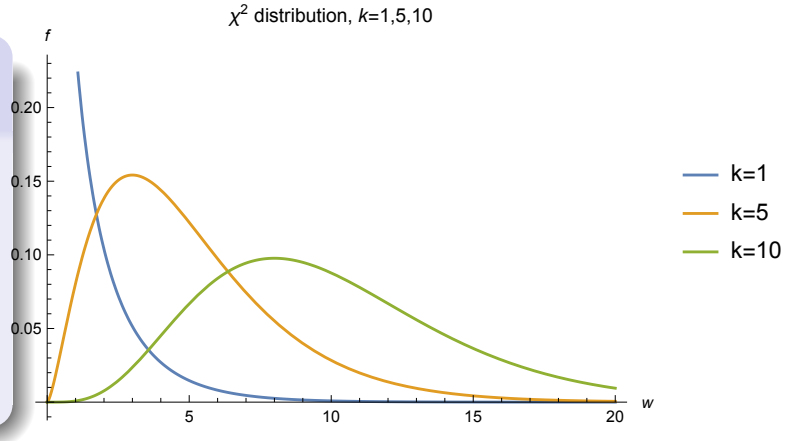
カイ二乗分布

$Z_1, \ldots, Z_k \sim \mathrm{N}(0, 1^2)$, iid のとき, 確率変数 $W = Z_1^2 + \cdots + Z_k^2$ は, 自由度 k のカイ二乗分布 χ_k^2 にしたがう.

言語	小	大	読み
英語	x	X	エクス
ギリシャ語	χ	X	カイ

χ_k^2 の確率密度関数

$$f_k(y) = \begin{cases} C_k \times y^{\frac{k}{2}-1} \mathrm{e}^{-\frac{1}{2}y} & (y \geq 0) \\ 0 & (他) \end{cases}$$



$W_k \sim \chi_k^2$ に対して,

$\mathrm{E}[W_k] = \mathrm{E}[Z_1^2 + \cdots + Z_k^2] = k, \mathrm{V}[W_k] = 2k, \mathrm{E}[(W_k)^\ell] =$ 簡単じゃない.

カイ二乗分布の確率密度関数と累積分布関数

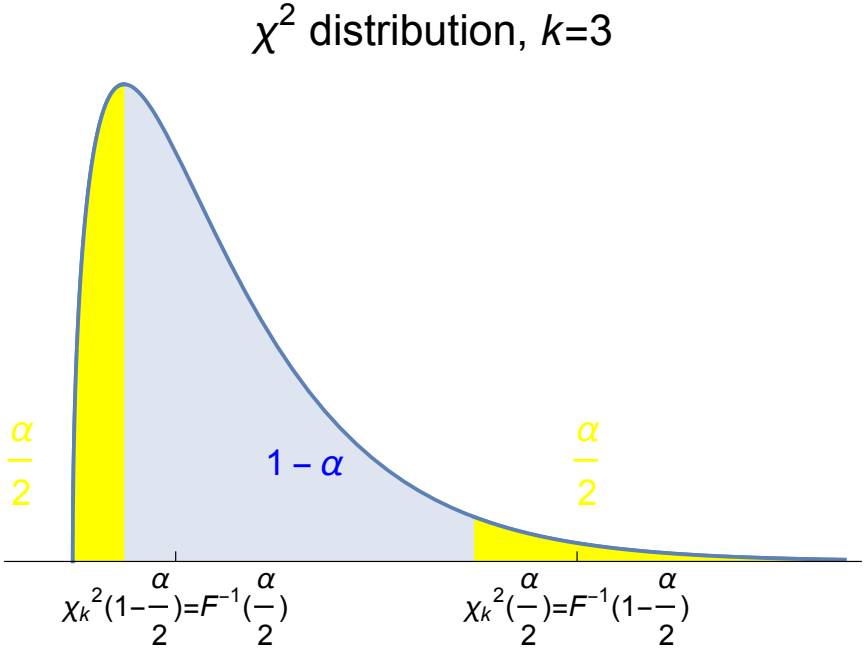
カイ二乗分布

```
1 rvw=stats.chi2(df=k) # k は自由度
```

上側確率 $\alpha = P(W > \chi_k^2(\alpha))$ となる境い目 $\chi_k^2(\alpha) = F^{-1}(1-\alpha)$.

岩薩林 確率・統計 例題 5.7

岩薩林 確率・統計 付表 3 付表 3 は付表 1 とフォーマットが違う



L12-Q4

Quiz(カイ二乗分布の確率と $\chi^2_k(\alpha)$)

標準正規分布にしたがう確率変数 $Z \sim \mathrm{N}(0, 1^2)$ と, 自由度 $k = 1$ のカイ二乗分布にしたがう確率変数 $W \sim \chi^2_1$ を考える. 各分布の累積分布関数の逆関数 Φ^{-1}, F^{-1} を使って答え, さらに Python, 数表を使って数値にしよう.

1. 確率 $P(Z > x_0) = 0.025$ となる $x_0 = z(0.025)$ を求めよう.
2. 確率 $P(Z > x_0) = 1 - 0.025$ となる $x_0 = z(1 - 0.025)$ を求めよう.
3. 確率 $P(W > w_0) = 0.05$ となる $w_0 = \chi^2_1(0.05)$ を求めよう.
4. 確率 $P(W > w_0) = 1 - 0.05$ となる $w_0 = \chi^2_1(1 - 0.05)$ を求めよう.

ここまで来たよ

12 母比率の区間推定

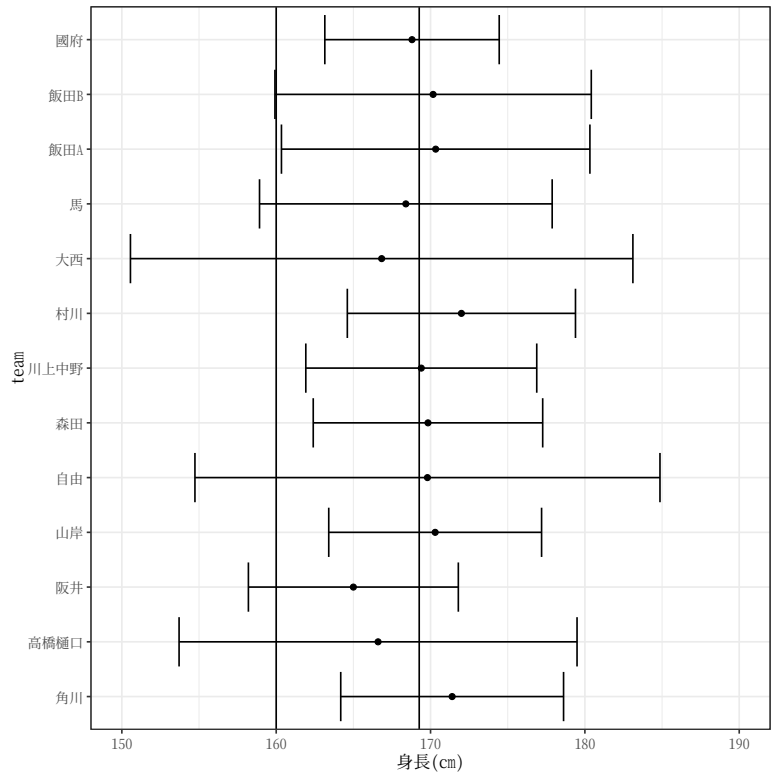
13 母分散・母平均値の区間推定
- 母分散の区間推定
- 母平均値の区間推定 (正規母集団, 母分散未知)

標本抽出 → 区間推定の結果

母集団=回答者全体, 母集団サイズ=75
身長の母平均値=169.3cm, 身長の母分散=72.9cm^2, 滋賀県高校の母比率=0.31.
標本平均値 $\overline{X}$, 不偏標本分散 S^2, 標本比率 p, 標本サイズ N

team	$\overline{X}$(cm)	S^2(cm^2)	p	N
國府	168.80	20.700	0.20000	5
飯田 B	170.17	95.367	0.16667	6
飯田 A	170.33	90.667	0.33333	6
馬	168.40	58.300	0.40000	5
大西	166.83	240.567	0.50000	6
村川	172.00	49.600	0.50000	6
川上中野	169.40	36.300	0.20000	5
森田	169.83	50.167	0.33333	6
自由	169.80	147.200	0.20000	5
山岸	170.30	92.900	0.40000	10
阪井	165.00	30.000	0.20000	5
高橋樋口	166.60	107.800	0.20000	5
角川	171.40	33.800	0.20000	5

信頼係数 $1-\alpha = 0.95$ 信頼区間



不偏標本分散のしたがう分布

不偏標本分散のしたがう分布 岩薙林 確率・統計 定理 5.6

正規分布 $\mathrm{N}(\mu, \sigma^2)$ からサイズ n の標本 $X_1, X_2, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$, iid を取り出すとき, 不偏標本分散

$$S^2 = \frac{1}{n-1}[(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2]$$

から定めた

$$W = (n-1) \times \frac{S^2}{\sigma^2}$$

は, 自由度 $k = n - 1$ の カイ二乗分布 χ^2_{n-1} にしたがう.

比 $\frac{\text{不偏標本分散}}{\text{母分散}}$ は 1 に近いところに分布するが, 実は, 確率変数 $\frac{W}{n-1}$.
$(W \sim \chi^2_{n-1})$

証明じゃないけど説明

独立な $X_i \sim \mathrm{N}(\mu, \sigma^2)$ $(i = 1, \ldots, n)$ に対して,

$$n \times \frac{1}{n}\left[\left(\frac{X_1 - \mu}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \mu}{\sigma}\right)^2\right]$$

は自由度 n のカイ二乗分布 χ^2_n にしたがう.
不偏標本分散 S^2 に対して,

$$W = (n-1) \times \frac{S^2}{\sigma^2} = (n-1) \times \frac{1}{n-1}\left[\left(\frac{X_1 - \overline{X}}{\sigma}\right)^2 + \cdots + \left(\frac{X_n - \overline{X}}{\sigma}\right)^2\right]$$

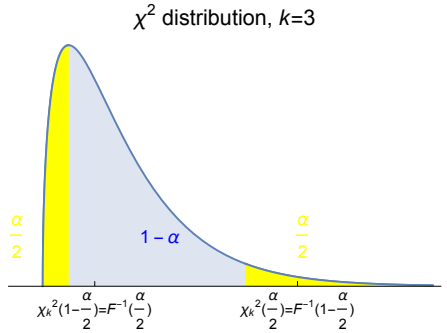
は自由度は $n-1$ のカイ二乗分布 χ^2_{n-1} にしたがう.
$-\mu$ でなく $-\overline{X}$ であるため自由度 $n-1$.

母分散の区間推定

$$P\left(F_{\mathrm{w}}^{-1}(\tfrac{\alpha}{2}) < (n-1)\times\frac{S^2}{\sigma^2} < F_{\mathrm{w}}^{-1}(1-\tfrac{\alpha}{2})\right) = 1-\alpha$$

$$P\left(\chi^2_{n-1}(1-\tfrac{\alpha}{2}) < (n-1)\times\frac{S^2}{\sigma^2} < \chi^2_{n-1}(\tfrac{\alpha}{2})\right) = 1-\alpha$$

不等式を σ^2 について解いて次の結果を得る.



$\chi_k^2(\alpha)$ の定義

岩薩林 確率・統計 例題 5.7

$\alpha = P(W > \chi_k^2(\alpha)).$

$\chi_k^2(\alpha) = F_{\mathrm{w}}^{-1}(1-\alpha).$

命題 (母分散の信頼区間 岩薩林 確率・統計 定理 7.3**)**

標本の不偏標本分散が S^2 のとき, 母分散 σ^2 の信頼係数 $1-\alpha$ の信頼区間は

$$\frac{n-1}{F_{\mathrm{w}}^{-1}(1-\frac{\alpha}{2})}\times S^2 \quad <\sigma^2< \quad \frac{n-1}{F_{\mathrm{w}}^{-1}(\frac{\alpha}{2})}\times S^2.$$

$$\frac{n-1}{\chi^2_{n-1}(\frac{\alpha}{2})}\times S^2 \quad <\sigma^2< \quad \frac{n-1}{\chi^2_{n-1}(1-\frac{\alpha}{2})}\times S^2.$$

だいたい S^2 だけど,「かける」補正係数 $(n-1)/W \simeq 1$, $W \sim \chi^2_{n-1}$.

L13-Q1

Quiz(母分散の区間推定)

あるファーストフードチェーンのポテトフライ S の重さは正規分布に従うという.
お店で 9 個のポテトフライ S サイズを買って重さを量り, サイズ 9 の標本とした.
このとき標本平均値は $80\mathrm{g}$, 不偏標本分散は $72\mathrm{g}^2$ だった.
母分散を信頼係数 $1-\alpha=0.95$ で区間推定しよう.

岩薩林 確率・統計 例題 7.3(p163), 問題 4(p.164), 練習問題 7.1(4)(p.173)

ここまで来たよ

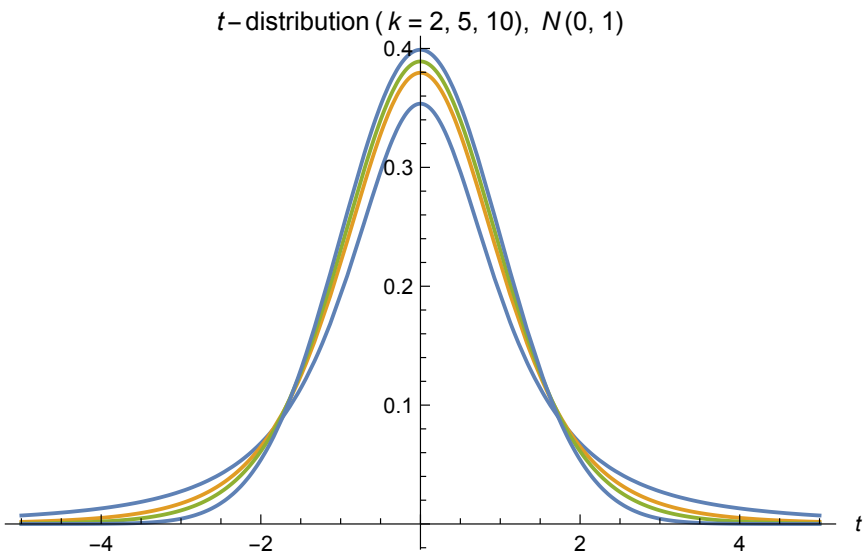
12 母比率の区間推定

13 母分散・母平均値の区間推定
- 母分散の区間推定
- 母平均値の区間推定 (正規母集団, 母分散未知)

t 分布

定義 (t 分布 岩薩林 確率・統計 §5.5)

$Z \sim \mathrm{N}(0, 1^2)$, $W \sim \chi^2_k$, Z と W が独立,
のとき連続型確率変数 $T = \dfrac{Z}{\sqrt{W/k}}$ のしたがう分布を自由度 k の (スチューデントの, またはゴセットの)t 分布 t_k という.



自由度 k が小さいとき, $\mathrm{N}(0, 1^2)$ より低く広い.
自由度 $k \to +\infty$ で $\mathrm{N}(0, 1^2)$ に一致する.
確率密度関数

$$f_k(x) = A_k \cdot \left(1 + \frac{1}{k}x^2\right)^{-\frac{k+1}{2}}.$$

t 分布の確率密度関数と累積分布関数

岩薩林 確率・統計 付表 2

t 分布

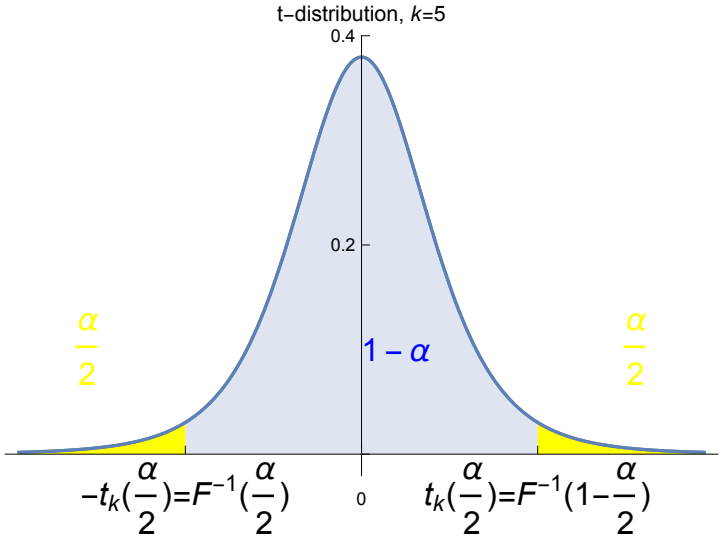
```
1  rvt=stats.t(df=k)  # kは自由度
```

上側確率 $\alpha = P(T > t_k(\alpha))$ となる $t_k(\alpha)$ の値.

t 分布の確率密度関数は偶関数なので, $t_k(1-\alpha) = -t_k(\alpha)$.

岩薩林 確率・統計 付表 2 付表 2 は付表 1 とフォーマットが違う

$\mathrm{t}_k \to \mathrm{N}(0, 1^2), t_k(0.025) \to 1.960, t_k(0.005) \to 2.576\ (k \to +\infty)$.



L13-Q2

Quiz(t 分布の確率と $t_k(\alpha)$)

標準正規分布にしたがう確率変数 $Z \sim \mathrm{N}(0, 1^2)$ と, 自由度 $k = 40$ の t 分布にしたがう確率変数 $T \sim \mathrm{t}_{40}$ を考える. 各分布の累積分布関数の逆関数 F^{-1} を使って答え, さらに Python, Excel, 数表を使って数値にしよう.

1. 確率 $P(Z > z) = 0.025$ となる z を求めよう.
2. 確率 $P(Z > z) = 1 - 0.025$ となる z を求めよう.
3. 確率 $P(T > w) = 0.025$ となる w を求めよう.
4. 確率 $P(T > w) = 1 - 0.025$ となる w を求めよう.

母平均値の区間推定 (正規母集団, 母分散未知) 岩薩林 確率・統計 §7.1

母分散既知の区間推定 確率統計 I(2023)L12 では, 標本平均値を標準化した. $Z = \frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}$.
ふつう, μ がわからないときは σ^2 もわかってない.
σ^2 のかわりに不偏標本分散 S^2 (確率変数) を使った, 標準化もどき

$T = \frac{\overline{X}-\mu}{\sqrt{S^2/n}} = \frac{\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}$ でやっちゃいたい. 分子 $\sim \mathrm{N}(0,1^2)$, 分母
$= \sqrt{W/(n-1)}, W \sim \chi^2_{n-1}$ より, $T \sim \mathrm{t}_{n-1}$.

命題 (T の分布 岩薩林 確率・統計 定理 5-7(p.130))

母集団 $\mathrm{N}(\mu, \sigma^2)$ から, サイズ n の標本 $X_1, \cdots, X_n$ を取り出したとき, 標本平均値 $\overline{X}$, 不偏標本分散 S^2 に対して, 次の T は自由度 $n-1$ の Student の t 分布にしたがう.

$$T = \frac{\overline{X}-\mu}{\sqrt{S^2/n}}.$$

母集団が厳密に正規分布にしたがわなくても近似的に正しいことが多い.

$$P\left(F_{\mathrm{t}}^{-1}(\tfrac{\alpha}{2}) < \frac{\overline{X}-\mu}{\sqrt{S^2/n}} < F_{\mathrm{t}}^{-1}(1-\tfrac{\alpha}{2})\right) = 1-\alpha$$

$$P\left(-t_k(\tfrac{\alpha}{2}) < \frac{\overline{X}-\mu}{\sqrt{S^2/n}} < t_k(\tfrac{\alpha}{2})\right) = 1-\alpha$$

命題 (母平均値の信頼区間 (母分散未知) 岩薩林 確率・統計 定理 7.1(7.3))

(母分散未知の) 正規分布 $\mathrm{N}(\mu, ?^2)$ にしたがう母集団から, サイズ n の標本を得たとき, 母平均値 μ の 信頼係数 $1-\alpha$ の信頼区間は, 自由度 $n-1$ の t 分布の累積分布関数 F で,

$$\overline{X} - F_{\mathrm{t}}^{-1}(1-\tfrac{\alpha}{2}) \times \sqrt{S^2/n} < \mu < \overline{X} - F_{\mathrm{t}}^{-1}(\tfrac{\alpha}{2}) \times \sqrt{S^2/n}.$$

$$\overline{X} - t_{n-1}(\tfrac{\alpha}{2}) \times \sqrt{S^2/n} < \mu < \overline{X} + t_{n-1}(\tfrac{\alpha}{2}) \times \sqrt{S^2/n}.$$

母分散既知 確率統計 I(2023)L12 と比べて, $\sigma^2 \rightsquigarrow S^2$ $z(\frac{\alpha}{2}) \rightsquigarrow t_{n-1}(\frac{\alpha}{2})$

L13-Q3

Quiz(母平均値の区間推定 (母分散未知))

あるドーナツ製造マシンが製造するドーナツの重さ $X\mathrm{g}$ は, 正規分布にしたがう確率変数である
製造された 4 個のドーナツの重さを測定したところ, 次のようだった.
$51\mathrm{g}, 52\mathrm{g}, 47\mathrm{g}, 50\mathrm{g}$.

1. 母平均値 $\mu = \mathrm{E}[X]$ を, 信頼係数 $1-\alpha = 0.95$ で区間推定しよう. まず, $t_k(\alpha)$ または F_{t}^{-1} で書き, 小数に直そう.
2. 母平均値 $\mu = \mathrm{E}[X]$ を, 信頼係数 $1-\alpha = 0.99$ で区間推定しよう.

岩薩林 確率・統計 例題 7.1(p.158), 問題 1(p.158) 第 7 章練習問題 1(2)

区間推定のまとめ 岩薩林 確率・統計 p.xiv

- 母比率 (母集団がベルヌーイ分布) 岩薩林 確率・統計 定理 7.5(p.170)
  - 入力 信頼係数 α, 標本サイズ n, 該当するデータの個数 k
  - 係数 $z(\alpha/2)$ を求める分布: 標準正規分布 岩薩林 確率・統計 付表 1 下
- 母分散 (母集団が正規分布) 岩薩林 確率・統計 定理 7.3(p.163)
  - 入力 信頼係数 α, 標本サイズ n, 不偏標本分散 S^2
  - 係数 $\chi^2_{n-1}(\alpha/2)$ を求める分布 自由度 $n-1$ のカイ二乗分布
    岩薩林 確率・統計 付表 3
- 母平均値 (母集団が正規分布) 岩薩林 確率・統計 定理 7.1(p.157)
  - 入力 信頼係数 α, 標本サイズ n, 不偏標本分散 S^2, 標本平均値 $\overline{X}$
  - 係数 $t_{n-1}(\alpha/2)$ を求める分布 自由度 $n-1$ の t 分布 岩薩林 確率・統計 付表 2
- 「母分散がわかってるときの母平均値」という実用上あまりないケース 岩薩林 確率・統計 定理 6.1(p.145) 岩薩林 確率・統計 付表 1 下

統計的仮説検定・母平均値の両側検定

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　L14(2023-07-10 Mon)

最終更新: Time-stamp: "2023-07-10 Mon 07:47 JST hig"

今日の目標

- 最尤推定，線形モデルの推定が説明できる
- 統計的仮説検定を説明できる
- 母平均値の両側 t 検定ができる　岩薩林 確率・統計 §7.1

L12-Q3

Quiz 解答: カイ二乗分布の確率と $\chi^2_k(\alpha)$

1. $F_{\mathrm{Z}}^{-1}Z(1-0.025) = z(0.025) = 1.960.$
2. 標準正規分布の確率密度関数は偶関数なので,
   $F_{\mathrm{Z}}^{-1}(0.025) = z(1-0.025) = -z(0.025) = -1.960.$
3. $F_{\mathrm{W}}^{-1}(1-0.05) = \chi^2_1(0.05) = 3.841.$ 別解. $0.05 = P(W > w_0) = P(Z^2 > w_0) = P(Z < -\sqrt{w_0}$ or $Z > +\sqrt{w_0}) = 2 \times P(Z > \sqrt{w_0})$. よって, $\sqrt{w_0} = 1.960.$
4. $F_{\mathrm{W}}^{-1}(0.05) = \chi^2_1(1-0.05) = 0.00393.$

L13-Q1

Quiz 解答: 母分散の区間推定

標本サイズは $n = 9$, 自由度は $9-1$, 母分散 σ^2 の信頼係数 $1-\alpha = 0.95$ の信頼区間は,

$$\frac{n-1}{F_{\mathrm{W}}^{-1}(1-\frac{\alpha}{2})} \times S^2 < \sigma^2 < \frac{n-1}{F_{\mathrm{W}}^{-1}(\frac{\alpha}{2})} \times S^2$$

$$\frac{n-1}{\chi^2_{n-1}(\frac{\alpha}{2})} \times S^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1}(1-\frac{\alpha}{2})} \times S^2$$

$$\frac{8}{17.53} \times 72 < \sigma^2 < \frac{8}{2.180} \times 72$$

$$32.85 < \sigma^2 < 264.2$$

L13-Q2

Quiz 解答:t 分布の確率と $t_k(\alpha)$

1. $F_{\mathrm{Z}}^{-1}(1-0.025) = \mathsf{ppf}(1-0.025) = z(0.025) = 1.960.$
2. 標準正規分布の確率密度関数は偶関数なので,
   $F_{\mathrm{Z}}^{-1}(0.025) = \mathsf{ppf}(0.025) = z(1-0.025) = -z(0.025) = -1.960.$
3. $F_{\mathrm{t}}^{-1}(1-0.025) = \mathsf{ppf}(1-0.025) = t_{40}(0.025) = 2.021.$
4. t 分布の確率密度関数は偶関数なので, $F^{-1}F_{\mathrm{Z}}^{-1}(0.025)(0.025) = \mathsf{ppf}(1-0.025) = t_{40}(1-0.025) = -t_{40}(0.025) = -2.021.$

L13-Q3

Quiz 解答: 母平均値の区間推定 (母分散未知)

1. 重さの標本平均値は $\overline{X} = 50\mathrm{g}$. 不偏標本分散は $S^2 = \frac{1}{4-1} \cdot 14\mathrm{g}^2$. 信頼係数 0.95 の信頼区間は

$$
\begin{aligned}
50 - F_{\mathrm{t}}^{-1}(1-0.05/2) \times \sqrt{\tfrac{14/3}{4}} &< \mu < 50 - F_{\mathrm{t}}^{-1}(0.05/2) \times \sqrt{\tfrac{14/3}{4}} \\
50 - t_3(0.05/2) \times \sqrt{\tfrac{14/3}{4}} &< \mu < 50 + t_3(0.05/2) \times \sqrt{\tfrac{14/3}{4}} \\
50 - 3.182 \times \sqrt{\tfrac{14/3}{4}} &< \mu < 50 + 3.182 \times \sqrt{\tfrac{14/3}{4}} \\
46.563 &< \mu < 53.437
\end{aligned}
$$

2. 同様に,

$$
\begin{aligned}
50 - F_{\mathrm{t}}^{-1}(1-0.01/2) \times \sqrt{\tfrac{14/3}{4}} &< \mu < 50 - F_{\mathrm{t}}^{-1}(0.01/2) \times \sqrt{\tfrac{14/3}{4}} \\
50 - t_3(0.01/2) \times \sqrt{\tfrac{14/3}{4}} &< \mu < 50 + t_3(0.01/2) \times \sqrt{\tfrac{14/3}{4}} \\
50 - 5.841 \times \sqrt{\tfrac{14/3}{4}} &< \mu < 50 + 5.841 \times \sqrt{\tfrac{14/3}{4}} \\
43.691 &< \mu < 56.309
\end{aligned}
$$

ここまで来たよ

13 母分散・母平均値の区間推定
- 最尤推定
- 線形モデルとしての回帰分析

14 統計的仮説検定・母平均値の両側検定
- 統計的仮説検定の考え方
- 正規分布にしたがう母集団の母平均値の両側 t 検定

線形モデル (統計モデルのある一族)

あるドーナツ製造機の作るドーナツの重さ Y は次のモデルに従う.

$$Y = \beta_0 + \epsilon, \qquad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Y, ϵ: 連続型確率定数, $\sigma > 0, \beta_0$: 定数=パラメタ=母数

ϵ: 誤差, ノイズ. 小文字だけど確率変数.
(隠された) パラメタ母数 μ, σ^2 を, n 個のドーナツの重さのデータから推定したい.

$$\begin{aligned} \mathrm{E}[Y] &= \beta_0 + \mathrm{E}[\epsilon] = \beta_0, \\ \mathrm{V}[Y] &= \qquad \mathrm{V}[\epsilon] = \sigma^2. \end{aligned}$$

正規分布と限定した以外は, ここしばらくやってた, 母平均値, 母分散の推定の言い換えに過ぎない.
だけど, 多数のパラメタを含む一般的なモデルにも使える考え方をする.

尤度 likelihood

$\epsilon = Y - \beta_0 \sim \mathrm{N}(0, \sigma^2)$ より, ドーナツの重さ y を得る確率密度は,

$$f(y|\beta_0, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y-\beta_0)^2}{2\sigma^2}}$$

サイズ n の標本が $y_1, \cdots, y_n$ である確率密度は, 独立同分布なので積で,

$$\begin{aligned} f(y_1, y_2, \cdots, y_n|\beta_0, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y_i-\beta_0)^2}{2\sigma^2}} \\ &= \left(2\pi\sigma^2\right)^{-n/2} \mathrm{e}^{-\sum_{i=1}^{n} \frac{(y_i-\beta_0)^2}{2\sigma^2}} \end{aligned}$$

n 次元正規分布

多変量解析及び演習

この f を, y_i の確率密度関数と思わず, が測定済データ $y_1, \ldots, y_n$ が定数, β_0, σ が変数と思ったとき, 尤度 (ゆうど) 関数 $L(\beta_0, \sigma)$ という.

$$L(\beta_0, \sigma) = f(y_1, y_2, \cdots, y_n|\beta_0, \sigma)$$

最尤推定 岩薩林 確率・統計 なし

最尤推定

β_0, σ の推定値として $L(\beta_0, \sigma)$ が最大になる値を選ぶ

2 変数関数の最大値 $\rightsquigarrow$ 偏微分 微積分 II

$$0 = \frac{\partial L}{\partial \beta_0}(\beta_0, \sigma) = \frac{\partial L}{\partial \sigma}(\beta_0, \sigma)$$

ここでは最初の等式だけ解く. 合成微分.

$$0 = (\text{定数})^{-n/2} \times \sum_{i=1}^{n} \frac{1}{\sigma^2}(y_i - \beta_0) \times \mathrm{e}^{\text{同じ}}$$

$$0 = \sum_{i=1}^{n} (y_i - \beta_0)$$

$$\beta_0 = \frac{1}{n} \sum_{i=1}^{n} y_i$$

ここでは最初の式だけ.
合成微分.
β_0(母平均値) の最尤推定 (MLE=maximum likelihood estimation) 値 $\hat{\beta}_0$ は標本平均値じゃん

ここまで来たよ

13 母分散・母平均値の区間推定
- 最尤推定
- 線形モデルとしての回帰分析

14 統計的仮説検定・母平均値の両側検定
- 統計的仮説検定の考え方
- 正規分布にしたがう母集団の母平均値の両側 t 検定

(確率変数でない) 変数 x に依存する確率変数 Y

このドーナツ製造機の作るドーナツの重さ Y は, 温度 x によるらしい.
次の線形回帰モデルを仮定する.

$$Y = \beta_0 + \beta_1 \cdot x + \epsilon, \qquad \epsilon \sim \mathrm{N}(0, \sigma^2)$$

Y, ϵ: 連続型確率定数, β_0, β_1: 回帰係数, $\sigma > 0$: 定数=パラメタ=母数

Y: 目的変数 (従属変数) ここでは確率変数
x: 説明変数 (独立変数) ここでは確率変数でない

ノイズ・誤差 $\epsilon = Y - \beta_0 + \beta_1 \cdot x \sim \mathrm{N}(0, \sigma^2)$.
$\epsilon = Y - \beta_0 - \beta_1 x \sim \mathrm{N}(0, \sigma^2)$ より, ドーナツの重さ y を得る確率密度は,

$$f(y|x, \beta_0, \beta_1, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y-\beta_0-\beta_1 \cdot x)^2}{2\sigma^2}}$$

(正確に) 指定した温度 x_i $(i = 1, \ldots, n)$ で製造したときの重さが y_i $(i = 1, \ldots, n)$ である確率密度は, 独立分布なので積.

$$
\begin{aligned}
f(y_1, y_2, \cdots, y_n | x_1, \ldots, x_n, \beta_0, \beta_1, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y_i - \beta_0 - \beta_1 \cdot x)^2}{2\sigma^2}} \\
&= \left(2\pi\sigma^2\right)^{-n/2} \mathrm{e}^{-\sum_{i=1}^{n} \frac{(y_i - \beta_0 - \beta_1 \cdot x_i)^2}{2\sigma^2}} \\
&= L(\beta_0, \beta_1, \sigma)
\end{aligned}
$$

最尤推定

測定済のデータ x_i, y_i $i = 1, \cdots, n$ を定数と思ったときの, 3 変数関数
$L(\beta_0, \beta_1, \sigma) = f(y_1, y_2, \cdots, y_n | x_1, \ldots, x_n, \beta_0, \beta_1, \sigma)$ の最大値は? $\rightsquigarrow$ 偏微分 微積分 II

$$0 = \frac{\partial L}{\partial \beta_0}(\beta_0, \beta_1, \sigma) = \frac{\partial L}{\partial \beta_1}(\beta_0, \beta_1, \sigma) = \frac{\partial L}{\partial \sigma}(\beta_0, \beta_1, \sigma)$$

ここでは最初の 2 つの等式だけ解く.

$$0 = \frac{\partial L}{\partial \beta_0} = (\text{定数}) \sum_i \frac{1}{\sigma^2}(y_i - \beta_0 - \beta_1 x_i) \times \mathrm{e}^{\text{同じ}}$$

$$0 = \frac{\partial L}{\partial \beta_0} = (\text{定数}) \sum_i \frac{1}{\sigma^2} x_i (y_i - \beta_0 - \beta_1 x_i) \times \mathrm{e}^{\text{同じ}}$$

しょせん, β_0, β_1 の連立 1 次方程式

正規方程式

岩薩林 確率・統計 §9.2

$$n\beta_0 + (\sum_i x_i)\beta_1 = \sum_i y_i$$

$$(\sum_i x_i)\beta_0 + (\sum_i x_i^2)\beta_1 = \sum_i x_i y_i$$

加減法 $\rightsquigarrow$

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}}, \qquad y = \frac{s_{xy}}{s_{xx}}x + \overline{y} - \frac{s_{xy}}{s_{xx}}\overline{x}$$

$$\hat{\beta}_0 = \overline{y} - \frac{s_{xy}}{s_{xx}}\overline{x} \qquad y - \overline{y} = \frac{s_{xy}}{s_{xx}}(x - \overline{x})$$

ここで,

$$\overline{x} = \frac{1}{n}\sum_i x_i \quad \text{平均値っぽい形}$$

$$\overline{y} = \frac{1}{n}\sum_i y_i \quad \text{平均値っぽい形}$$

$$s_{xy} = \frac{1}{n}\sum_i x_i y_i - \overline{x}\cdot\overline{y} \overset{\text{岩薩林 確率・統計 定理 1.5}}{=} \frac{1}{n}\sum_i (x_i - \overline{x})(y_i - \overline{y}) \quad \text{共分散っぽ}$$

$$s_{xx} = \frac{1}{n}\sum_i x_i^2 - \overline{x}^2 \overset{\text{岩薩林 確率・統計 定理 1.2}}{=} \frac{1}{n}\sum_i (x_i - \overline{x})^2 \quad \text{分散っぽい形}$$

$Y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x + \epsilon$ ($\hat{\beta}_0, \hat{\beta}_1$: 回帰係数) に代入.

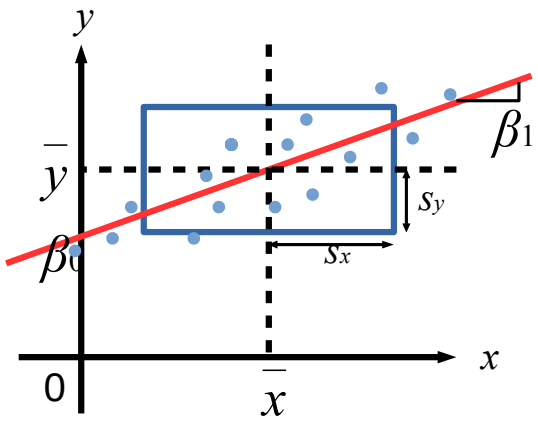
回帰直線

岩薩林 確率・統計 §9.1

推定結果 $\hat{\beta}_0, \hat{\beta}_1$ を係数とする xy 平面の直線

データ分析

$$y - \overline{y} = \frac{s_{xy}}{s_{xx}}(x - \overline{x})$$



回帰係数, 予測値の信頼区間

確率統計 II,III

ここまで来たよ

13 母分散・母平均値の区間推定
- 最尤推定
- 線形モデルとしての回帰分析

14 統計的仮説検定・母平均値の両側検定
- 統計的仮説検定の考え方
- 正規分布にしたがう母集団の母平均値の両側 t 検定

推定 (estimation) と検定 (test)

点推定　μ は値 xx と推定　岩薩林 確率・統計 §6.1

区間推定　μ は値 yy と値 zz の間と推定 (信頼係数 $1-\alpha$ で)　岩薩林 確率・統計 §6.2

仮説検定　μ は値 xx と差がある と, (時々) 断言 (有意水準 α で) =見逃し多いけど発色したら正しい ($1-\alpha$ で) 血痕試験紙

岩薩林 確率・統計 §6.3

あるドーナツ製造器は, 重さ X(確率変数) の母平均値が 55g であるように調整済みだという. しかし, 5 個買ってみたら, 違う感じ. これ, 本当に母平均値 55 g なの?(っていうか 55 g でないと言いたい).

ある学習法を使ってるある生徒の, 毎日のテストでの 1 か月の平均点は 63 点. 自分が別の学習法で教えた 5 日間の平均点は… 自分の方法は優れていると言いたい.

検定はだいたいこんな考え方

岩薩林 確率・統計 §6.3

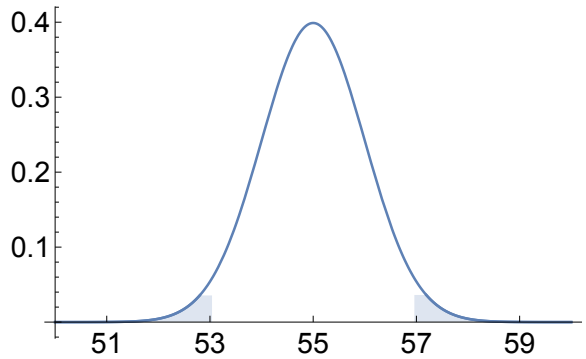
確率変数 X は, 正規分布 $\mathrm{N}(55, 2^2)$ にしたがうという. $\sigma^2 = 2^2$ は確かだとわかってるけど, $\mu = 55$ (帰無仮説) が本当なのか疑っている. というより，本当でないと言いたい
サイズ 4 の標本を抽出したところ,

$$54, 57, 57, 60$$

だった. $\to$ 標本サイズ $N = 4$, 標本平均値 $\overline{X} = \frac{1}{4}(X_1 + \cdots + X_4) = 57$.

$\overline{X} \sim \mathrm{N}(55, 2^2/4)$.

確率 α(=有意水準) でしか起きないようなまれなことが起きたら (検定統計量 $\overline{X}$ が棄却域にはいったなら)
おかしいと判定する (帰無仮説を棄却する)



帰無仮説と対立仮説

- H_0: 帰無仮説 (null hypothesis) = 背理法の仮定 =「真の母平均値 μ は $55\mathrm{g}$ に等しい」
- H_1: 対立仮説 (alternative hypothesis) = 示したい命題 =「真の母平均値 μ は $55\mathrm{g}$ でない」

検定 (test)=統計的仮説検定 (statistical hypothesis test)
心理学, 教育学, 社会科学などでは標本サイズを大きくできないことが多い. 小さくても Yes/No の結論を出す, 科学業界で合意された方法.

帰無仮説を棄却する reject
検定統計量の実現値が境目を越えて大きすぎたり小さすぎたりしたら (棄却域にはいったら) 帰無仮説 (=背理法の仮説) が偽, 対立仮説が真と結論する (試験紙が発色した)
帰無仮説を棄却しない accept
実験失敗 (背理法使おうとしたけど矛盾導けなかった). 何も言えない (発色しなかった)

ここまで来たよ

13 母分散・母平均値の区間推定
- 最尤推定
- 線形モデルとしての回帰分析

14 統計的仮説検定・母平均値の両側検定
- 統計的仮説検定の考え方
- 正規分布にしたがう母集団の母平均値の両側 t 検定

L14-Q1

Quiz(母平均値の両側検定 (母分散未知)=両側 t 検定)

あるドーナツ製造マシンが次々に製造するクロワッサンドーナツの重さ Xg は, 正規分布にしたがうことがわかっている. 母平均値は 55g だと言われていたが疑っている. きょう 5 個製造したところ, 下のようだった.

$50\text{g}, 50\text{g}, 51\text{g}, 46\text{g}, 48\text{g}.$

ドーナツ製造マシンが次々に製造するクロワッサンドーナツの重さ Xg の母平均値が 55g と異なるかどうか, 有意水準 $\alpha = 0.05$ で統計的仮説検定を行おう.

一般の統計的仮説検定の，レポートや論文での書き方

母集団を決める．母集団の分布タイプを仮定する．使う検定を決める/決まる (将来は自作できる).

1. 「有意水準 $\alpha = \cdots$ で」「…検定を行う」(2,3 を名前で予告する)
2. 「帰無仮説を…とする」「対立仮説を…とする」
3. 「帰無仮説のもとでナントカ検定統計量 Y は …分布にしたがう」
4. 「この標本に対してナントカ検定統計量の実現値は $y = \cdots$ である」
5. (棄却域の境い目の値を計算しておく)
6. 「‘y 不等号 (境い目)’ より帰無仮説を棄却する/棄却できない」「よって母ナントカは…である/とはいえない」

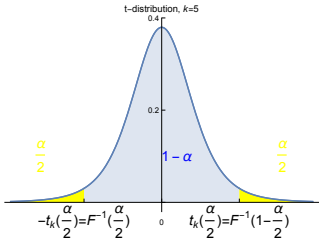
正規分布にしたがう母集団の母平均値の両側 t 検定

母平均値の両側 t 検定 岩薩林 確率・統計 定理 7.2(p.159)

前提 母集団が正規分布 $\mathrm{N}(\mu, \text{何か})$ にしたがう.

1. 有意水準 α で母平均値の両側 t 検定を行う.
2. 帰無仮説を母平均値 $\mu = \mu_0$, 対立仮説を $\mu \neq \mu_0$ とする.
3. 帰無仮説のもとで, 検定統計量 $T = \dfrac{\overline{X} - \mu_0}{\sqrt{S^2/n}}$ は自由度 $n-1$ の t 分布にしたがう.
4. T の実現値 t を標本平均値 $\overline{X}$, 不偏標本分散 S^2, 標本サイズ n から計算すると $t = \cdots$.
5. 棄却域は $|t| > t_{n-1}(\frac{\alpha}{2})$ である.
6. (結論) 帰無仮説を棄却する/できない, 母平均値は $\mu \neq \mu_0$ と結論する/できない.

棄却域は $(-\infty, -u) \cup (u, +\infty)$. $P(T < -u \text{または} u < T) = \alpha$ であるように, $u = t_{n-1}(\frac{\alpha}{2})$ と決めた.



t 分布では $F^{-1}(1-\frac{\alpha}{2}) = -F^{-1}(\frac{\alpha}{2}) = t_k(\frac{\alpha}{2})$.

L14-Q1

Quiz 解答: 母平均値の両側検定 (母分散未知)=両側 t 検定

1. 有意水準 0.05 で, 正規分布の母平均値に対する両側 t 検定を行う.
2. 帰無仮説を「ドーナツの重さの母平均値 μ が $\mu_0 = 55\mathrm{g}$ に等しい」すなわち「$\mu = \mu_0$」とする. 対立仮説を「$\mu \neq \mu_0$」とする.
3. サイズ $n = 5$ の標本の標本平均値を $\bar{X}$, 不偏標本分散を S^2 とするとき, 検定統計量
$$T = \frac{\bar{X} - 55}{\sqrt{S^2/5}}$$
は, 帰無仮説のもとで, 自由度 $5-1$ の t 分布に従う.
4. この標本に対する検定統計量の実現値は,
$t = \frac{\overline{X}-\mu_0}{\sqrt{\frac{S^2}{n}}} = \frac{49-55}{\sqrt{\frac{1}{5}\frac{16}{5-1}}} = -3\sqrt{5} = -6.708.$
5. 棄却域は $|t| > t_4(0.05/2)$. rvt=stats.t(df=4)　rvz.ppf(1-0.05/2) または t 分布表より $|t| > 2.776$.

6. $|-6.708| > 2.776$ であり, 実現値 t は棄却域に含まれるので, 帰無仮説を棄却する. ドーナツの重さの母平均値は $55\mathrm{g}$ と異なる, と結論する.
(注: このことを,「有意」significant という言葉で表現する人もいる. 結果は有意である, 母平均値 μ は $55\mathrm{g}$ と有意に異なる, 母平均値 μ と 55 の間には有意差がある, 有意な標本である, など)

重さは負にならないし, 正規分布にしたがうというのはおかしな前提だが, ここは練習ってことで. 世の中には変な状況下で強引に t 検定を使う人が多くいるが, 数理の人はおかしさを認識できるように.

岩薩林 確率・統計 例題 7.2(p.159), 問題 2(p.160)

L14-Q2

Quiz(正規分布の母平均値に関する t 検定)

あるコンビニには, ドーナツ販売開始前には, 9:00–10:00 に平均 196 人の客が来店していた. ドーナツ販売開始後の 4 日間, 来店客数は次の通りだった. $204, 208, 188, 200$
来店者数は正規分布にしたがうと考える. ドーナツ販売開始後に来店客数の母平均値は変化したか? 有意水準 0.05 で考える.

2023-07-17 月祝は授業なし，練習問題締切あり．
ふつうじゃない Trial L14 非参照テストではなく, 2023-07-24 月 20:00 までにアップロードするレポート.

- 練習問題 L14-01, L14-02 に合格すると, 個人別レポート問題の取得や提出が可能になる.
- 趣旨: 検定の長い文を (参照ありで) 書けることを確かめる必要. 非参照や短時間のテストで確認することが難しい.

定期試験期間中の任意参加授業内小テスト用途:Moodle 成績評価方法参照.
2023-07-31 月 09:15-10:15, 参照なし. 7-002. 事前の出欠連絡不要．理由を問わず, 欠席したときの追試や配慮はなし. Trial L01-08 の出題計画をあわせたもの．要学生証, 個人別座席指定.
すべて持込不可, ただし, 岩薩林 確率・統計 pp. xii,xiii の公式集を印刷して当日配布
成績評価考慮対象になるかもしれない欠席の届は 2023-07-24 月 20:00 までに提出．Trial L14, L15 の採点結果の公表は，2023-07-24 より後，2023-07-31 直前になってしまうかも.

母比率・母分散の検定・片側検定

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　L15(2023-07-24 Mon)

最終更新: Time-stamp: "2023-07-24 Mon 07:03 JST hig"

今日の目標

- 母比率の片側検定ができる　岩薩林 確率・統計 §7.3
- p 値を使った検定ができる
- 第 1 種, 第 2 種の過誤, 有意水準と検出力を説明できる

L14-Q1
Quiz 解答: 母平均値の両側検定 (母分散未知)=両側 t 検定

1. 有意水準 0.05 で, 正規分布の母平均値に対する両側 t 検定を行う.
2. 帰無仮説を「ドーナツの重さの母平均値 μ が $\mu_0 = 55\mathrm{g}$ に等しい」すなわち「$\mu = \mu_0$」とする. 対立仮説を「$\mu \neq \mu_0$」とする.
3. サイズ $n = 5$ の標本の標本平均値を $\bar{X}$, 不偏標本分散を S^2 とするとき, 検定統計量
$$T = \frac{\bar{X} - 55}{\sqrt{S^2/5}}$$
は, 帰無仮説のもとで, 自由度 $5-1$ の t 分布に従う.
4. この標本に対する検定統計量の実現値は,
$t = \frac{\overline{X}-\mu_0}{\sqrt{\frac{S^2}{n}}} = \frac{49-55}{\sqrt{\frac{1}{5}\frac{16}{5-1}}} = -3\sqrt{5} = -6.708.$
5. 棄却域は $|t| > t_4(0.05/2)$. rvt=stats.t(df=4)　rvz.ppf(1-0.05/2) または t 分布表より $|t| > 2.776$.

6. $|-6.708| > 2.776$ であり, 実現値 t は棄却域に含まれるので, 帰無仮説を棄却する. ドーナツの重さの母平均値は $55\mathrm{g}$ と異なる, と結論する.
   (注: このことを,「有意」significant という言葉で表現する人もいる. 結果は有意である, 母平均値 μ は $55\mathrm{g}$ と有意に異なる, 母平均値 μ と 55 の間には有意差がある, 有意な標本である, など)

L14-Q2
Quiz 解答: 正規分布の母平均値に関する t 検定

1. 有意水準 0.05 で, 正規分布の母平均値に対する両側 t 検定を行う.
2. 帰無仮説を「ドーナツ販売開始後の, 来店客数の母平均値 μ は 196 に等しい」, すなわち $\mu = 196$ とする. 対立仮説を $\mu \neq 196$ とする.

3. サイズ n の標本の標本平均値を $\bar{X}$, 不偏標本分散を S^2 とすると, 検定統計量
$$T = \frac{\bar{X} - 196}{\sqrt{S^2/4}}$$
は, 帰無仮説のもとで, 自由度 $n-1$ の t 分布に従う.
4. この標本の実現値は $\bar{X} = 200, S^2 = \frac{224}{4-1} = 74.7$. よって, $t = \frac{200-196}{\sqrt{\frac{1}{4}\frac{224}{3}}} = 0.92582$.
5. t 分布表より, 棄却域は $|t| > t_3(0.05/2)$ すなわち, $|t| > 3.182$.
6. $|0.92582| < 3.182$ であり, この標本の実現値は棄却域に含まれないので, 帰無仮説は棄却できない. 来店客数の母平均値が変化したとは結論できない.
(注: 結果は有意でなかった, 母平均値 μ と 196g の間には有意差がない, など).

ここまで来たよ

14 統計的仮説検定

15 母比率・母分散の検定・片側検定
- 片側検定
- 母比率の検定=二項検定 (の正規近似)
- p 値=有意確率
- 仮説検定での第 1 種・第 2 種の過誤と混同行列

母平均値・母分散・母比率の検定

	両側	片側
母平均値	**両側 t 検定** 確率統計 I(2023)L14 岩薩林 確率・統計 問題 2(p.160) 前回の授業	**片側 t 検定** 岩薩林 確率・統計 例題 7.2(p.159)
母分散	**両側カイ二乗検定** 岩薩林 確率・統計 例題 7.4(p.165)	**片側カイ二乗検定** 岩薩林 確率・統計 問題 5(p.166) 岩薩林 確率・統計 §7 練習問題 1(3)
母比率	**両側二項検定** 岩薩林 確率・統計 練習問題 2(p.173)	**片側二項検定** 岩薩林 確率・統計 例題 7.7(p.172) 岩薩林 確率・統計 例題 7(p.172) 岩薩林 確率・統計 練習問題 3(p.173) 今回の授業 確率統計 I(2023)L15

例: 分散が大きすぎる心配

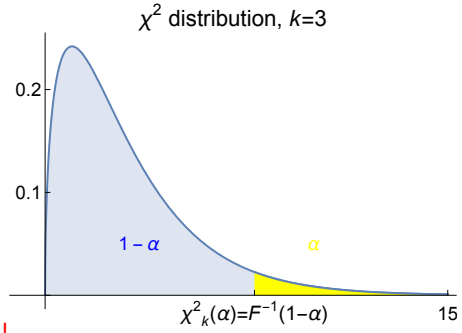
L15-Q1

Quiz(母分散の片側カイ二乗検定)

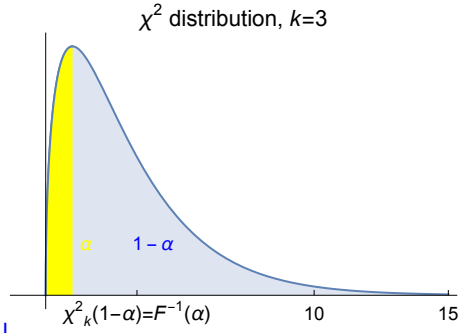
あるファーストフードチェーンでは，ポテトフライ S の重さを一定にするために，その母分散が $4\mathrm{g}^2$ であることが定められているという．
トレーニング中のアルバイトの人に, ポテトフライ S サイズを 9 個作ってもらったところ, 重さは下のようだった (単位は g).

$$76, 76, 76, 76, 80, 84, 84, 84, 84.$$

このアルバイトの作るポテトフライ S の重さの母分散 σ^2 は, $2^2\mathrm{g}^2$ より大きいか? 重さが正規分布にしたがうと仮定し, 有意水準 $\alpha = 0.05$ で, 母分散のカイ二乗検定で判定しよう.



上側片側



下側片側

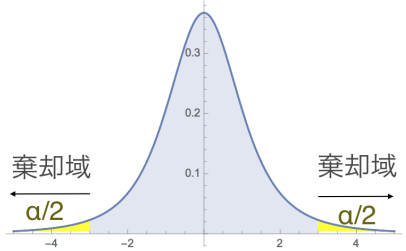
岩薩林 確率・統計 定理 7.4 | 岩薩林 確率・統計 練習問題 7-1(3)(p.173)

片側検定と両側検定 岩薩林 確率・統計 p.150

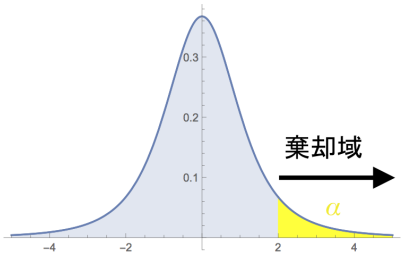
- 片側・両側検定の帰無仮説 $\mu = \mu_0$.
- 片側 t 検定の対立仮説 $\mu > \mu_0 (\mu < \mu_0)$
- 両側 t 検定の対立仮説 $\mu \neq \mu_0$

片側検定では, どちらか片側だけに確率=面積 α の棄却域ができる. 実現値がここにはいったら帰無仮説を棄却.
両側検定のときは, 両側に確率=面積 $\alpha/2$ ずつの棄却域ができる. 実現値がいずれかにはいったら帰無仮説を棄却.

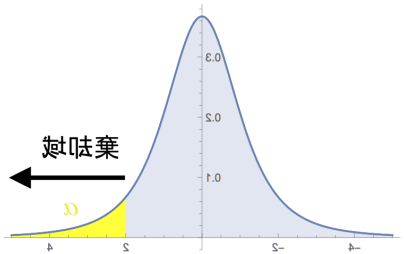
両側



上側片側



下側片側



ここまで来たよ

14 統計的仮説検定

15 母比率・母分散の検定・片側検定
- 片側検定
- 母比率の検定=二項検定 (の正規近似)
- p 値=有意確率
- 仮説検定での第 1 種・第 2 種の過誤と混同行列

母比率の検定 I

p と書いてた母比率, 二項分布の $X \sim \mathrm{B}(p, n)$ の p を, 今日は r と書きます

例 (母比率の検定のアイデア)

このマジック用コインは表の出る確率 (母比率)$r = 1/5$ だと言われているが, もっと表がでちゃう気がする.
50 回投げた (サイズ $n = 50$ のサンプル) ところ, 表 $X = 15$ 回表が出た.
こんな「稀な」ことが起きるってことは, そうに違いない!

$\rightsquigarrow$ これ, 統計的仮説検定として定式化できる考え方
前提 表が出る回数 $X \sim \mathrm{B}(r, n)$.
帰無仮説 $r = \frac{1}{5}$, 対立仮説 $r > \frac{1}{5}$.

母比率の検定 II

帰無仮説 (背理法の仮定) のもとで, $X = 15$ はとても稀 (矛盾) と言いたい. 確率統計 I(2023)L14

$$P(X = 15) = \frac{50!}{15!(50-15)!}(\frac{1}{5})^{15}(1 - \frac{1}{5})^{50-15} = 0.0299 \quad \text{稀じゃん}$$

ちょっと待て. 特定の回数が出ることじたい稀. 上の確率は
$1/$場合の数 $= 1/51 = 0.0196$ より大きい.
もっともらしい $50 \cdot \frac{1}{5}$ 回

$$P(X = 10) = \frac{50!}{10!(50-10)!}(\frac{1}{5})^{10}(1 - \frac{1}{5})^{50-10} = 0.1398$$

だって稀じゃん?
これ $(X = 15)$ 以上に極端なことが起きる確率を考えよう.

$$P(15 \leq X \leq 50) = \sum_{k=15}^{50} P(X = k)$$

母比率の検定 III

36 個も加えるの? …これを計算機などで直接計算してしまう手もある正確確率検定

正規近似による計算 確率統計 I(2023)L10

$\mathrm{E}[X] = 50 \cdot \frac{1}{5} = 10, \mathrm{V}[X] = 50 \cdot \frac{1}{5} \cdot \frac{4}{5} = 8.$
中心極限定理より, '近似的に' $X \sim \mathrm{N}(10, 8)$. $Z = \frac{X-10}{\sqrt{8}} \sim \mathrm{N}(0, 1^2)$.

$$
\begin{aligned}
P(15 \leq X \leq 50) &= P(\tfrac{15-10}{\sqrt{8}} \leq Z < +\infty) \\
&= F_{\mathrm{Z}}(\infty) - F_{\mathrm{Z}}(\tfrac{15-10}{\sqrt{8}}) \qquad \text{確率統計 I(2022)L05} \\
&= 1 - \mathsf{cdf}(\tfrac{15-10}{\sqrt{8}}) = 1 - 0.96145 = 0.0385.
\end{aligned}
$$

母比率の検定 IV

$P(X \geq 15) = 0.0385$ (これがあとで出てくる p 値) が稀と思うかは主観.
けんかにならないように, 計算し始める前に基準を決めておく. それが,
有意水準 $\alpha = 0.05$ or 0.01.

上で出てきた, Z は, 分子分母を $n = 50$ で割ると,

$$\frac{15 - 10}{\sqrt{8}} = \frac{\frac{15}{50} - \frac{10}{50}}{\sqrt{\frac{1}{50}\frac{1}{5}\frac{4}{5}}} = \frac{\hat{r} - r_0}{\sqrt{\frac{1}{n} \cdot r_0 \cdot (1 - r_0)}}$$

という, 母比率の区間推定の導出に似た形になる.
ただし, 平方根の中にあるのは帰無仮説の母比率 $r_0 = \frac{1}{5}$, 標本比率は $\hat{r}$.

再掲: 一般の統計的仮説検定の，レポートや論文での書き方

確率統計 I(2023)L14

母集団を決める．母集団の分布タイプを仮定する．使う検定を決める/決まる (将来は自作できる).

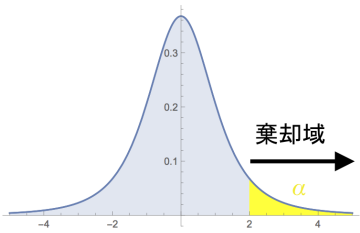
1. 「有意水準 $\alpha = \cdots$ で」「…検定を行う」(2,3 を名前で予告する)
2. 「帰無仮説を…とする」「対立仮説を…とする」
3. 「帰無仮説のもとでナントカ検定統計量 X は …分布にしたがう」
4. 「この標本に対してナントカ検定統計量の実現値は $y = \cdots$ である」
5. (棄却域の境い目の値を計算しておく)
6. 「'y 不等号 (境い目)' より帰無仮説を棄却する/棄却できない」「よって母ナントカは…である/とはいえない」

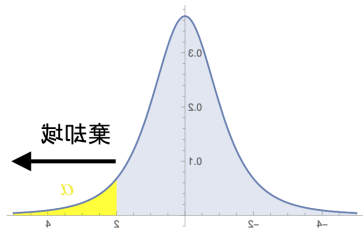
母比率の片側 (二項) 検定 (の正規近似)

岩薙林 確率・統計 定理 7.6(p.171)

前提 $X \sim \mathrm{B}(r, n)$.

1. 有意水準 α で正規近似による母比率の片側 (二項) 検定を行う.
2. 帰無仮説を母比率 $r = r_0$, 対立仮説を母比率 $r > r_0$ (or $r < r_0$) とする.
3. 帰無仮説のもとで検定統計量 $Z = \dfrac{X - nr_0}{\sqrt{nr_0(1-r_0)}} = \dfrac{\hat{r} - r_0}{\sqrt{r_0(1-r_0)/n}}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう (n 大のとき).
   $\hat{r} = \dfrac{X}{n}$: 標本比率, r_0: (帰無仮説の) 母比率.
4. この標本に対して検定統計量の実現値は…
5. 棄却域は $z > z(\alpha)$ (or $z < -z(\alpha)$).
6. 「(…) より帰無仮説を棄却する/できない. よって母比率 $r > r_0$ (or $r < r_0$) と結論する/できない.

上側片側 

下側片側 

L15-Q2

Quiz(母比率の片側検定)

あるスピードくじ (「あたり」と「はずれ」だけがある) は, あたりの母比率 r は $\frac{1}{10}$ に等しいと言われている. しかし, 実際の r はこれより大きいのではないかと疑っている.
くじを 100 本ひいたところ, 15 本があたりだった.
有意水準 $\alpha = 0.05$ で母比率の検定を行おう.

岩薩林 確率・統計 例題 7.7, 問題 8(p.172), 第 7 章練習問題 2,3

L15-Q3

Quiz(母比率の片側検定)

あるスピードくじ (「あたり」と「はずれ」だけがある) は, あたりの母比率 r は $\frac{1}{10}$ に等しいと言われている. しかし, 実際の r はこれより小さいのではないかと疑っている.
くじを 100 本ひいたところ, 5 本があたりだった.
有意水準 $\alpha = 0.01$ で母比率の検定を行おう.

ここまで来たよ

14 統計的仮説検定

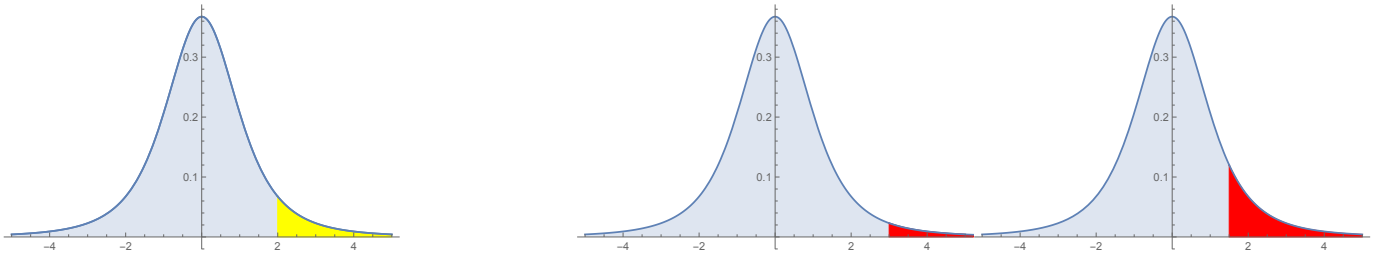
15 母比率・母分散の検定・片側検定
- 片側検定
- 母比率の検定=二項検定 (の正規近似)
- p 値=有意確率
- 仮説検定での第 1 種・第 2 種の過誤と混同行列

p 値=有意確率による棄却する/しない判定 岩薙林 確率・統計 p.152

検定の Step6 では, X や Z が極端かどうか判定している
$X = 15 \Leftrightarrow Z = \frac{15-10}{\sqrt{8}}$ が棄却域 $(\alpha = 0.05)$ に入っているか?
$\Updownarrow$
p 値 $P(X \geq 15) = P(z \geq \frac{15-30}{\sqrt{8}})$ が有意水準 α より小さいか?

標本の p 値 (p-value)=有意確率 岩薙林 確率・統計 p.152

帰無仮説のもとで, 検定統計量がこの標本よりも極端な値をとる確率
$\alpha > p$ のとき帰無仮説を棄却



上側片側検定のとき p 値 $= 1 - F(z)$.
棄却域の考え方での棄却: $z > F^{-1}(1-\alpha)$.　　$F(z)$ は増加関数だから…
p 値の考え方での棄却: $F(z) > 1-\alpha$ すなわち $p = 1 - F(z) < \alpha$.

ここまで来たよ

14 統計的仮説検定

15 母比率・母分散の検定・片側検定
- 片側検定
- 母比率の検定=二項検定 (の正規近似)
- p 値=有意確率
- 仮説検定での第 1 種・第 2 種の過誤と混同行列

母分布についての真実と，検定での棄却の有無

	帰無仮説を棄却, 有意である	帰無仮説を棄却できない, 有意でない
対立仮説が成立 $\mu \neq \mu_0$	確率 $1-\beta$	第 2 種の過誤 確率 β
帰無仮説が成立 $\mu = \mu_0$	第 1 種の過誤 確率 α	確率 $1-\alpha$

α, β はどちらも小さくしたいが, 両立しない.
α はほぼゼロ ($\alpha = 0.01$ or 0.05) に固定して, β をなるべく小さくするように試みる習慣.
見逃し率: β. 検出力:$1-\beta$.
誤検出率, 有意水準, 危険率: α.

検定とインフルエンザ検査

岩薙林 確率・統計 §6.3

	検査で陽性	検査で陰性
病気	○ (TP 真陽性)	×(FN 見逃し, 偽陰性)
病気でない	×(FP 誤検出, 偽陽性)	○ (TN 真陰性)

真/偽=True/False, 陽性/陰性=Postive/Negative

混同行列 (confusion matrix) 上の度数や頻度を表にして，行列と思ったもの．機械学習でも使う (画像の猫判定)

以下はどれも大きい方がいいが, 一つを大きくすると他が小さくなりがち

- 感度=再現率=真陽性率=recall=sensitivity= $\frac{\mathrm{TP}}{\mathrm{TP+FN}} = 1-\beta$
- 特異度=真陰性率=specificity= $\frac{\mathrm{TN}}{\mathrm{TN+FP}} = 1-\alpha$
- 適合率=precision= $\frac{\mathrm{TP}}{\mathrm{TP+FP}}$
- 正解率=精度=accuracy = $\frac{\mathrm{TP+TN}}{\text{すべて}}$
- F 値=recall と precision の調和平均

連絡

- 2023-07-24 月 20:00 まで TrialL14
- 2023-07-31 月 1 任意参加の小テスト 確率統計 I(2023)L14
- 2023-07-31 月 20:00 まで TrialL15=Moodle の Web 練習問題的なもの+アンケート+作文レポート
- 2023-07-24 月 20:00 まで 考慮対象になるかもしれない欠席の届を提出．L15 の分は 2023-07-31 月 20:00
- Trial L14 の採点結果の公表は，2023-07-24 より後，2023-07-31 直前になってしまうかも．
- 3 年前期必修 数理と社会でお会いしましょう

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 L16(2023-07-31 Mon)

最終更新: Time-stamp: "2023-07-24 Mon 07:46 JST hig"

今日の目標

L15-Q1
L15-Q2

Quiz 解答: 母比率の片側検定

1. 有意水準 $\alpha = 0.05$ で正規近似による母比率の片側 (二項) 検定を行う.
2. 帰無仮説を母比率 $r = \frac{1}{10}$, 対立仮説を母比率 $r > \frac{1}{10}$ とする.
3. 帰無仮説のもとで, 100 本中のあたりの回数 $T \sim \mathrm{B}(\frac{1}{10}, 100)$. あたりの標本比率を $\hat{r}$ とすると, $Z = \frac{\hat{r} - \frac{1}{10}}{\sqrt{\frac{1}{100} \cdot \frac{1}{10}(1 - \frac{1}{10})}}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう.
4. この標本に対して $\hat{r} = \frac{15}{100}$ なので, 検定統計量の実現値は $Z = \frac{5}{3} = 1.667$.
5. 棄却域は $z > z(0.05) =$ scipy.stats.norm(loc=0,scale=1).ppf$(1 - 0.05) = 1.645$.

6. $1.667 > 1.645$ なので,z は棄却域に含まれる. 帰無仮説を棄却する. 母比率 $r > \frac{1}{10}$ と結論する.

不等式が逆の場合

6. 帰無仮説を棄却できない. 母比率 $r > \frac{1}{10}$ とは結論できない.」

p 値による場合

5. p 値は $p = P(Z > \frac{5}{3}) = F(+\infty) - F(\frac{5}{3}) =$ $1 -$ scipy.stats.norm(loc=0,scale=1).cdf$(\frac{5}{3}) = 0.0475$.
6. $0.0475 = p < \alpha = 0.05$ なので帰無仮説を棄却する. よって母比率 $r > \frac{1}{10}$ と結論する.

L15-Q3

Quiz 解答: 母比率の片側検定

1. 有意水準 $\alpha = 0.01$ で正規近似による母比率の片側 (二項) 検定を行う.
2. 帰無仮説を母比率 $r = \frac{1}{10}$, 対立仮説を母比率 $r < \frac{1}{10}$ とする.
3. 帰無仮説のもとで, 100 本中のあたりの回数 $T \sim \mathrm{B}(\frac{1}{10}, 100)$. あたりの標本比率を $\hat{r}$ とすると, $Z = \frac{\hat{r} - \frac{1}{10}}{\sqrt{\frac{1}{100} \cdot \frac{1}{10}(1 - \frac{1}{10})}}$ は近似的に標準正規分布 $\mathrm{N}(0, 1^2)$ にしたがう.
4. この標本に対して $\hat{r} = \frac{5}{100}$ なので, 検定統計量の実現値は $Z = -\frac{5}{3} = -1.667$.
5. 棄却域は $z < -z(0.01) = \mathsf{scipy.stats.norm(loc=0,scale=1).ppf}(0.01) = -2.326$.
6. $-2.326 < -1.667$ なので,z は棄却域に含まれない. 帰無仮説を棄却しない. 母比率 $r < \frac{1}{10}$ と結論できない.

不等式が逆の場合

6. 帰無仮説を棄却する. 母比率 $r < \frac{1}{10}$ と結論する.」

p 値による場合

5. p 値は $p = P(Z < -\frac{5}{3}) = F(-\frac{5}{3}) - F(0) =$ scipy.stats.norm(loc=0,scale=1).cdf$(-\frac{5}{3}) = 0.0475$.
6. $0.0475 = p > \alpha = 0.01$ なので帰無仮説を棄却しない. よって母比率 $r < \frac{1}{10}$ と結論しない.